Modern Methods of Map Construction Using Optical Sensors Fusion

Ramil Safin, Tatyana Tsoy, Roman Lavrenov, Ilya Afanasyev
Department of Intelligent Robotics, Kazan Federal University, Kazan, Russia

Evgeni Magid
Department of Intelligent Robotics, Kazan Federal University, Kazan, Russia
Higher School of Economics University, Moscow, Russia
E-mail: safin.ramil@it.kfu.ru

Abstract

Map construction, or mapping, plays an important role in robotic applications. Mapping relies on inherently noisy sensor measurements to construct an accurate representation of a surrounding environment. Generally, individual sensors suffer from performance degradation issues under certain conditions in the environment. Sensor fusion allows to obtain statistically more accurate perception and to cope with performance degradation issues by combining data from multiple sensors of different modalities. This article reviews modern sensor fusion methods for map construction applications based on optical sensors, such as cameras and laser range finders. State-of-the-art mapping solutions built upon different mathematical theories and concepts, such as machine learning, are considered.

Keywords: Sensor Fusion, Mapping, SLAM, Machine Learning, Camera, LiDAR

1. Introduction

Mapping is the process of constructing a map of an environment using robot perception. There exist multiple map representations, such as sparse point clouds, topological maps, and dense voxel grids. Mapping could be difficult due to adverse conditions (e.g., low lightning) and presence of dynamic objects.

Each type of sensor has its own limitations. For example, cameras underperform in low-light environments, while laser scanners cannot provide high-resolution data. To obtain reliable maps, it is required to combine strengths of each sensor to cope with their weaknesses. Sensor fusion, or data fusion, is a technique for combining data from multiple sensors in a way that allows to obtain more reliable and accurate information about the system being measured (Fig. 1). Data fusion is used in many robotics and machine vision applications, such as autonomous navigation and localization of mobile robots[1].

This article reviews modern sensor fusion methods for map construction applications based on optical sensors, such as cameras and laser scanners. State-of-the-art mapping solutions built upon different mathematical theories and concepts, such as machine learning, are considered.

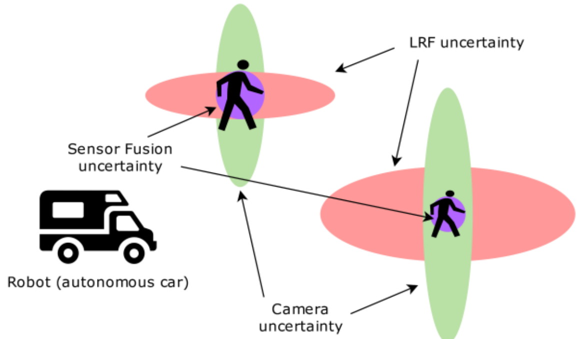

Fig. 1. An illustration of a laser range finder and camera sensor fusion. Uncertainty of the state is reduced due to multiple sources of information.

2. Sensor Fusion

High-level architecture of the sensor fusion system consists of the following components:

1. ***sensors*** that independently measure an observed quantity;
2. mathematical ***models*** that convert the observed quantity into a target value (e.g., calculating a position from distance measurements);
3. an ***inference algorithm*** that calculates the resulting target value by combining data from all sensors.

The inference algorithm combines all available measurements using provided sensor models and calculates an optimal state of a system with respect to specified criteria. The optimality of the solution and the criteria are determined by a cost function (e.g., an absolute error or a mean squared error).

The problem of data fusion can be viewed as combining measurements from multiple sensors to obtain a more optimal assessment (from statistical point of view) of the system's state. There are at least two reasons for data integration: data redundancy and completeness (complementarity). There are three main approaches for data fusion from sensors of various modalities[2]: a high-level (takes place on a decision-making level[3]), a mid-level (uses features of an environment[4]), and a low-level (operates with raw sensor data[5]). Methods can be also categorized into traditional and machine learning approaches[6].

Traditional approaches, such as statistical and probabilistic methods, create a model of a system being measured using various probability distributions. This enables to update an already initialized models when integrating new data. Probabilistic methods are built on Bayesian recursive rules, which enables to calculate and update a probability distribution function of an estimated state given sensor data.

One of the basic Bayesian integration methods is the Probabilistic Occupancy Grid (POG)[7]. The occupancy grid map divides an environment into cells of equal size. Each cell carries information about a probability of its occupancy. The Bayesian rule allows to update the initially constructed map. Computational performance is highly dependent on a map resolution and on the number of degrees of freedom. In practice, 2D maps are often considered due to a lower complexity.

Kalman filter is an analytical solution to the Bayesian approach, provided that a dynamical system is linear, and a state follows a Gaussian distribution. Dynamic systems with non-linear dependencies can be estimated using the Extended Kalman filter (EKF) – the state is linearized at each moment of time by expanding it into Taylor series and calculating the Jacobian matrix. In Unscented Kalman Filter (UKF) the problem is solved by linearization in a certain neighborhood – a group of weighted points (sigma points) is taken in a certain vicinity of a current state. Multi-state constrained Kalman Filter (MSCKF) is adapted to combine IMU data, extracted angles from a laser scanner, and features from a camera[8].

However, Kalman filters (the entire family) are unimodal approaches. Particle filters go further and suggest to model multimodal probability distributions of the state. Methods based on Monte Carlo describe the probability density function in terms of weighted elements of a space of linear and non-linear states. Particle filters are a subclass of the Monte Carlo method that allow calculating an optimal estimate of the state of a nonlinear state space with a Gaussian distribution. Compared to Kalman filters, a complexity of the algorithm grows much faster with a dimension of the state space (an exponential growth).

On the other hand, machine learning methods build a multi-layer network that allows extracting special features from a data stream. One of the most used architectures are convolutional (CNN)[9] and recurrent neural networks (RNN). For example, in[10] the authors propose a CNN-based approach, which combines data from a camera and a laser scanner to improve the object detection algorithm performance.

The problem with the machine learning approach is the quality of data on which a model is trained. The data may not contain all possible real-world scenarios, it may be biased and contain noise. Also, it is noted that machine learning methods are highly susceptible to malicious attacks[11].

Simultaneous localization and camera-based mapping algorithms often use the Bundle Adjustment (BA) approach and its modifications, such as windowed BA[12]. It simultaneously optimizes a pose and a map, can include several data sources by modifying a cost function. Combining visual data with laser scanner data allows obtaining data on a depth of a scene, being reliable under bad weather conditions[13].

Earlier simultaneous localization and mapping (SLAM) algorithms use only one sensor, mostly a camera [14][15] or laser scanner[16]. Currently, SLAM algorithms go through a stage where data is combined from several sensors of different modality. The most popular options are multiple camera configurations and laser scanners.

Deep learning methods are being used in dynamic SLAM problems. In[17], it is proposed to segment dynamic objects in a frame for their further masking and separation from a static scene. This enables to construct a map of an environment without modelling dynamic objects. CNNs are being used to estimate a depth of a scene. The estimated depth can be combined with data from a laser scanner to refine a map of an environment. For example, in[18], a scene depth scene was estimated based on an integration of data from a monocular camera and a laser rangefinder based on CNNs.

In[19], the authors proposed a monocular SLAM algorithm that used a one-dimensional LRF in conjunction with a camera to refine the distance to singular points in the image.

The paper[20] proposed an algorithm for constructing a dense environment map based on combining data from a monocular camera and predicting a scene depth using CNNs. In some works, image features and scanner point clouds are matched using scan matching methods[21], which allows to obtain dense and accurate 3D-maps of an environment. In[22] the author built a voxel environment map to effectively refine poses of key points using data from a laser scanner.

There are mapping algorithms that are based on the idea of associating point clouds from a laser and a camera. For example, in[23], a positioning system was proposed based on combining data from a camera and a laser scanner. The system searches for correlations between point clouds and images. In order to obtain accurate depths of point clouds methods such as segmentation of image areas (planar surfaces, foreground, etc.) are used.

In[24] a method for constructing dense 3D occupancy grid maps was based on sparse data from a laser scanner. To model dynamic objects in the environment, it was proposed to use dynamic occupancy mapping[25]. Recently, there are works of dense surfel mapping systems that can operate on CPU[26].

3. Conclusion

In conclusion, most of the reviewed sensor fusion mapping techniques are based on traditional methods, such as Extended Kalman Filtering and particle filters. In case of SLAM systems, where mapping and localization are done simultaneously, bundle adjustment optimization methods prevail. Most of the time, laser scanners are used to improve the camera depth estimation capabilities. Machine learning solutions enable to learn depth form images but lack accuracy and reliability.

Acknowledgements

The reported study was funded by the Russian Science Foundation (RSF) and the Cabinet of Ministers of the Republic of Tatarstan according to the research project No. 22-21-20033.

References

1. Safin R, Lavrenov R, Martinez-Garcia E A Evaluation of Visual SLAM Methods in USAR Applications Using ROS/Gazebo Simulation[C]//Int. Conf. on Electromechanics and Robotics" Zavalishin's Readings". Springer, 2021: 371-382.
2. Banerjee K et al. Online camera lidar fusion and object detection on hybrid data for autonomous driving[C]//IEEE Int. Vehicles Symposium (IV). IEEE, 2018: 1632–1638.
3. Shahian J B, Tulabandhula T, Cetin S. Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles[J]. Sensors. MDPI, 2019, 19(20): 4357.
4. Li Y et al. Feature level sensor fusion for target detection in dynamic environments[C]//American Control Conf. (ACC). IEEE, 2015: 2433–2438.
5. Yoo J H et al. 3d-cvf: Generating joint camera and lidar features using cross-view spatial feature fusion for 3d object detection[C]//European Conf. on Computer Vision. Springer, 2020: 720–736.
6. Fayyad J et al. Deep learning sensor fusion for autonomous vehicle perception and localization: A review[J]. Sensors. MDPI, 2020, 20(15): 4220.
7. Yue Y et al. Hierarchical probabilistic fusion framework for matching and merging of 3-d occupancy maps[J]. IEEE Sensors Journal. IEEE, 2018, 18(21): 8933–8949.
8. Xu X et al. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR[J]. Remote Sensing. MDPI, 2022, 14(12): 2835.
9. Buyval A, Gavrilenkov M, Magid E. A multithreaded algorithm of UAV visual localization based on a 3D model of environment: implementation with CUDA technology and CNN filtering of minor importance objects[C]//Int. Conf. on Artificial Life and Robotics, 2017: 356–359.
10. Kim J, Kim J, Cho J. An advanced object classification strategy using YOLO through camera and LiDAR sensor fusion[C]//Int. Conf. on Signal Proc. and Comm. Sys. (ICSPCS). IEEE, 2019: 1–5.
11. Pitropakis N et al. A taxonomy and survey of attacks against machine learning[J]. Computer Science Review. Elsevier, 2019, 34: 100199.
12. Zhou L, Kaess M. Windowed bundle adjustment framework for unsupervised learning of monocular depth estimation with u-net extension and clip loss[J]. IEEE Robotics and Automation Letters. IEEE, 2020, 5(2): 3283–3290.

13. Xiao L et al. Hybrid conditional random field based camera-LIDAR fusion for road detection[C]//Inf. Sciences. Elsevier, 2018, 432: 543–558.
14. Abbyasov B et al. Comparative analysis of ROS-based centralized methods for conducting collaborative monocular visual SLAM using a pair of UAVs[C]//Int. Conf. on Climbing and Walking Robots and Support Tech. for Mobile Machines (CLAWAR), 2020: 113–120.
15. Mingachev E et al. Comparison of ros-based monocular visual slam methods: DSO, LDSO, ORB-SLAM2 and DynaSLAM[C]//Int. Conf. on Inter. Collab. Robotics. Springer, 2020: 222–233.
16. Alishev N et al. Network failure detection and autonomous return algorithms for a crawler mobile robot navigation[C]//Int. Conf. on Development in eSystems Eng. (DeSE). IEEE, 2018: 169–174.
17. Xiao L et al. Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. Robotics and Autonomous Systems. Elsevier, 2019, 117: 1–16.
18. Gao H et al. Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment[J]. IEEE Trans. on Industrial Informatics IEEE, 2018, 14(9): 4224–4231.
19. Zhang Z et al. Scale estimation and correction of the monocular simultaneous localization and mapping (SLAM) based on fusion of 1D laser range finder and vision data[J]. Sensors. MDPI, 2018, 18(6): 1948.
20. Laidlow T, Czarnowski J, Leutenegger S. DeepFusion: Real-time dense 3D reconstruction for monocular SLAM using single-view depth and gradient predictions[C]//Int. Conf. on Robotics and Automation (ICRA). IEEE, 2019: 4068–4074.
21. Khattak S et al. Complementary multi–modal sensor fusion for resilient robot pose estimation in subterranean environments[C]//Int. Conf. on UAS (ICUAS). IEEE, 2020: 1024–1029.
22. Wang P et al. Vanishing Point Aided LiDAR-Visual-Inertial Estimator[C]//IEEE Int. Conf. on Robotics and Automation (ICRA). IEEE, 2021: 13120–13126.
23. Graeter J, Wilczynski A, Lauer M. Limo: Lidar-monocular visual odometry[C]//IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). IEEE, 2018: 7872–7879.
24. Gan L et al. Bayesian spatial kernel smoothing for scalable dense semantic mapping[J]. IEEE Robotics and Automation Letters. IEEE, 2020, 5(2): 790–797.
25. Min Y, Kim D U, Choi H L. Kernel-based 3-d dynamic occupancy mapping with particle tracking[C]//IEEE Int. Conf. on Robotics and Automation (ICRA). IEEE: 2021: 5268-5274.
26. Wang K, Gao F, Shen S. Real-time scalable dense surfel mapping[C]//Int. Conf. on Robotics and Automation (ICRA). IEEE, 2019: 6919–6925.

Authors Introduction

Mr. Ramil Safin

He is a Research Assistant and Ph.D. student in Kazan Federal University. His research interests are computer vision, mobile robotics, and SLAM.

Ms. Tatyana Tsoy

In 2012 she graduated from the University of Tsukuba. Since 2018 she has been a PhD student in Robotics at the Institute of Information Technology and Intelligent Systems of Kazan Federal University.

Assistant Professor Roman Lavrenov

He received his PhD from Kazan Federal University in 2020. Currently, he works as an Assistant Professor at Kazan Federal University, Russia. His research interests are path planning, ROS/Gazebo, and mobile robotics.

Assistant Professor Ilya Afanasyev

He received his PhD from Federal State Unitary Enterprise All-Russian Center “Vavilov State Optical Institute" in 2006. Currently, he works as an Assistant Professor at Kazan Federal University, Russia. He authors over 150 publications.

Professor Evgeni Magid

A Professor, a Head of Intelligent Robotics Department and a Head of Laboratory of Intelligent Robotic Systems (LIRS) at Kazan Federal University, Russia. Professor at HSE University, Russia. Senior IEEE member. He authors over 200 publications.