Adaptive Multimodal Localisation Techniques for Mobile Robots in Unstructured Environments

A Review

Niall O' Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Adolfo Velasco-Hernández, Daniel Riordan, Joseph Walsh
IMaR Technology Gateway
Institute of Technology Tralee
Tralee, Ireland
niall.omahony@research.ittralee.ie

***Abstract*— Mobile robots can be integrated as an entity in the new paradigm of the Internet of Things (IoT) and can be instrumental in extending sensing and manipulation capabilities to remote environments where the installation of sensor networks is unfeasible. Many anticipated applications of autonomous mobile robots require for them to navigate in diverse complex environments without support from exterior infrastructures. To perform this on-board navigation, the robot must make use of the available sensor technologies and fuse the most reliable data respective to the present environment in an adaptive manner. This paper will review recent efforts to develop onboard navigation systems which can seamlessly transition between outdoor and indoor environments and different terrains seamlessly. The methodologies surveyed include visual SLAM, Odometry and Place Recognition. An overview of the state-of-the-art is provided with a focus on approaches which are adaptive to dynamic sensor uncertainty, dynamic objects and dynamic scenes. In addition, the paper also provides an analysis of the most common sensor modalities and the factors affecting sensor uncertainty for the same.**

Keywords—Localisation Technologies; Deep Learning; Sensor Fusion; Mobile Robotics;

I. INTRODUCTION

Mobile robots can be integrated as an entity in the new paradigm of the Internet of Things (IoT) and can be instrumental in extending sensing and manipulation capabilities to remote environments where the installation of sensor networks is unfeasible. Much work has been done in the field of indoor localisation making use of infrastructure exterior to the navigating robot such as wireless networks [1], beacons [2], Ultra-wideband (UWB) technology [3], Visible light communication (VLC) [4], infrastructure-to-vehicle (I2V) communications and vehicle-to-vehicle (V2V) communications [5]. In outdoor scenarios, localisation is relatively straightforward and can be achieved with high accuracy with the use of Global Positioning Systems (GPS) or to the standalone cellular systems [6]. However, developing a navigation system which can seamlessly navigate between outdoor and indoor environments is far more challenging. It requires the use of a sophisticated perception system consisting of sensors of multiple modalities that perform well in certain conditions and an ability to decide which sensors contribute the most valuable information in the present circumstances.

This paper will be split into three sections. Section I will give a brief overview of the main sensor modalities pertinent to the state-of-the-art sensor-fusion based robot localisation techniques of today. This list will try to capture the reasons why data from each sensor modalities might be subject to uncertainty in certain circumstances. Section II will discuss the main challenges still present for robot localisation techniques and will survey some recent attempts to solve these challenges in the domains of visual SLAM, Odometry and Deep Learning. Section III will provide an analysis of techniques which may be used to cope with dynamics such as sensor noise due to influences discussed in the first section as well as new and moving obstacles in the robot's surroundings, scenes which appear to be different between viewings and transitioning between places which present drastically different sensor challenges. Finally, we draw conclusions and insights from our observations of the field of mobile robot localisation.

II. MULTIMODAL SENSOR SYSTEMS

If a robot is to navigate in any environment, it must not require any specialised infrastructure to be set up exterior to it and must rely on its onboard sensors. The recent availability of sensors of various modalities has enabled research into combining sensory data into DL models or through complementary sensor fusion techniques to achieve automated scene understanding.

There are two main types of sensors used within an autonomous vehicle: Exteroceptive sensors are used for perceiving the robots surrounding environment while Proprioceptive sensors are used to measure parameters internal to the robot system such as motor speed, wheel position and joint angles [7]. The following section will give a brief overview of the quality of data provided and the conditions required for such quality by some common sensors in multimodal perception systems.

A. Exteroceptive sensors

1) Electro-magnetic Radiation-based Systems

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie)

978-1-5386-4980-0/19/$31.00 ©2019 IEEE

Vision Systems are a natural choice as the primary sensor for autonomous mobile robots because they provide a great deal of information about the surrounding environment while working in both indoor and outdoor scenes. Vision systems can be in the form of monocular cameras, 3D depth sensing cameras or LiDAR. The field of computer vision has seen a huge amount of activity in recent years as the 'unreasonable effectiveness' of Deep Learning [8] and Convolutional Neural Networks has resulted in cameras becoming the primary sensor in many autonomous platforms

Many of the Deep Learning techniques to date have predominantly focused on the use of monocular cameras due to a vast amount of data available for training. There is a significant change on the way in computer vision and robotics as 3D vision systems radically improve in terms of performance and cost. 3D cameras are being used increasingly as a distance measurement between objects and the camera is provided for every pixel which greatly simplifies image segmentation tasks. There are a few different types of 3D Vision systems including Time of Flight (ToF), Stereo and Structured light. Each have different capabilities in terms of accuracy resolution, lighting conditions and processing requirements [9].

The information from 3D LiDAR is also very powerful for Simultaneous Location and Mapping (SLAM) tasks [10], however, the cost of these sensors prohibits their use in many mobile robot applications. The imminent advent of low-cost LiDAR will also enable a plethora of applications in mobile robotics [9]. Radar systems work in a similar manner compared to LiDAR but offer a different field of view, use radio waves rather than light and are also becoming increasingly more available.

With any vision system, an obvious limitation is field of view (FOV). Limited FOV can result in multiple cameras being required and an increase in the cost of the system. Omnidirectional cameras, also known as spherical or panoramic cameras, can be very advantageous to this regard, however, the use of data from these cameras in localisation algorithms has only begun to be explored recently [11], [12].

2) Sound-based Systems

We assign this category to sensors which measure sound waves which in contrast to electromagnetic radiation, must travel through a medium. Microphones and ultrasonic sensors fall into this category. the sensing capabilities of these sensors depend on the temperature, humidity, and environmental conditions of the medium of sound propagation. To accommodate for changes in temperature, many sensors are limited to close range proximity sensing applications and some utilize algorithms that adjust readings based on ambient temperature [10].

3) Communications-based systems

As mentioned previously, there are many examples of localisation systems which make use specialised infrastructure. However, they are not relevant to navigation in unstructured environments. GPS-aided navigation is one exception to this, however, as that infrastructure is in place and available in open outdoor environments. The accuracy of localisation can vary from a few meters for single GPS units to a few centimetres for differential or RTK (Real-Time Kinematics) systems [13]. Accuracy is severely affected by the presence of buildings and overhead obstacles which means that the weight to apply to GPS information must actively be called into question in systems which transition between indoor and outdoor scenes.

Communication between robots is also a possibility in unstructured environments. Cooperative Localisation (CL) is a technique used to improve localisation accuracy in multi-robot systems. In a study to investigate when CL is worthwhile, and how CL performance is affected under various conditions, it was found that accuracy has a substantial effect on performance, a communication rate that is too fast can be detrimental, and heterogeneous systems are better candidates for cooperative localisation than homogeneous systems [14].

B. Proprioceptive sensors

4) IMU (Inertial Measurement Unit)

A 6-axis IMU consists of three gyroscopes and three accelerometers which provide data on the rotational and linear motion of the platform with respect to the orthogonal X, Y, and Z axes. 9-axis IMUs also contain a magnetometer which provides information on orientation relative to the earth's magnetic field This magnetic field information can be used to compensate for drift and allows the absolute change in position and orientation to be tracked more accurately [15].

5) Rotary Encoders

Rotary encoders are often used to provide odometer data on vehicles. As with all sensors, the performance of an encoder is only as good as its signal which is subject to electrical interference from adjacent systems and sensor faults [7].

IMU and odometry data is used in dead reckoning techniques which can be quite effective in improving localisation performance [16]–[18]. However, on its own dead reckoning is subject to drift which may arise due to wheel slip or any measurement errors which accumulate over time.

III. LOCALISATION TECHNOLOGIES

Robot Localisation presents several challenges on which this section will now discuss.

Extracting accurate inferences from raw sensor data is challenging within the noisy and complex environments where these systems are deployed. Estimating sensor measurement noise is an essential factor when producing uncertainty models for state-of-the-art robotic positioning systems [19].

Secondly, the problem of place recognition from any number of viewpoints and over any timescale is subject to many challenges as each scene might appear drastically different for a single sensor modality due to variations in lighting, seasonal and weather conditions, new objects in a scene and occluded and partially incomplete representations of a scene.

The third challenge is that of the huge memory demands present if a map is to be kept which contains within it a representation of all the multi-sensor information required for localisation.

Machine learning is one of the most promising approaches for overcoming these challenges and techniques developed within this rapidly evolving area of machine learning are now state-of-the-art for many inference tasks in mobile robot localisation. This section will survey how machine learning has been applied to tasks involved in the navigation of a mobile robot primarily taking advantage of the data from vision sensors but also supported by sensors such as IMUs and LiDAR.

A. Visual SLAM

Visual SLAM (vSLAM) methods use vision system as the primary sensor for the registration of landmarks in a scene. VSLAM has the advantages of photogrammetry (rich visual data, low-cost, lightweight and low power consumption) without the associated heavy computational workload involved in post-processing. The vSLAM problem consists of steps such as environment sensing, data matching, motion estimation, as well as location update and registration of new landmarks [20].

Conventional probabilistic solutions to vSLAM follow the general workflow of; feature extraction and matching, refinement of matching errors, loop closure detection and global map optimization. The use of range imaging systems provides information on both the visual appearance and distance from the camera of the object which increases the robustness of real-time mapping [21].

Reasonable results have been achieved using these procedures in conjunction with a variety of different sensors including a combination of LIDAR, RGB-D camera, IMU and sonar [22], a multi-camera system [23], IMU and a monocular camera[1] [24] and stereovision [21]. However, challenges with these approaches exist such as limited sensing range of range cameras and accumulated memory usage on onboard computers [21].

An analytic solution (opposed to conventional probabilistic reasoning) to vSLAM has been demonstrated using a Time of Flight (Tof) range camera although it has yet to be tested on large regions [20]. A novel approach for mobile robot visual localization based on supervised learning using topological representations for the environment is proposed by [24]. The work demonstrated an increase in efficiency and reliability compared to classical localisation system providing high accuracy (99.94%) and low computational time (47.3µ s and 0.165 s for classification and extraction respectively).[25]

B. Odometry

A topic closely related to vSLAM is Visual Odometry or Visual-Inertial Navigation (VIN) where the motion of a robot is estimated through feature extraction in camera images. Recent advances in visual-inertial navigation on mobile robots are enabling unprecedented performance in pose estimation in GPS-denied environments using just IMUs and monocular cameras. A comprehensive review of the subject is provided by [26]. More recent works include a technique that improves localization accuracy in the presence of the effects of sensor noise or uncertainty arising from self-similar textures, variations in lighting, moving objects, and motion blur [27] and a self-improving technique which uses self-supervised dataset of point correspondences for retraining and achieves state-of-the-art performance [28].

Kalman Filters (KF) are generally for fusing odometry data with other data. The KF algorithm and its variants are a form of recursive optimal estimations, which efficiently utilize information in the time domain to reduce system errors. KF has become a standard approach for reducing errors in a least squares sense and is widely applied in navigation and positioning fields [29].

C. Place Recognition with Deep Learning

Systems which leverage the efficiency and accuracy achievable by CNNs have been proposed for multiple facets of robot localisation. For example, systems which can estimate the position of a robot's joints and body while allowing the camera to be free to move relative to the robot [30], [31] allow greater flexibility in sensor configuration. Another interesting example is the use of CNNs to predict dense depth maps and fuse them with depth measurements obtained from direct monocular SLAM. The system uses the depth predictions to improve performance in areas where monocular SLAM approaches tend to fail, e.g. along low-textured regions, and to yield semantically coherent scene reconstructions [32].

Visual place recognition is a challenging problem due to the vast range of ways in which the appearance of real-world places can vary. 3D cameras and LiDAR can help with this task as a scene that looks entirely different between seasons may have changed very little dimensionally. Deep Learning has also been used to tackle this problem through algorithms for semantic scene understanding [15].

Many approaches present solutions to improving generalisation and minimising the memory requirement of visual localisation which can be classified into two main categories; feature-based and image retrieval-based. Feature-based techniques regress the vision data for each scene to some sort of descriptor to represent the scene on a topological map. These descriptors may be local features [28], bag-of-words, global descriptors[29] or a combination as reviewed by [33]. Retrieval-based or image-based localization methods trade-off accuracy for scalability, by modelling the scene as an image database, and visual localization as an image retrieval problem. Many approaches employ image pre-processing or translation to bring all images to a visually similar and condition-invariant representation. However, image translation methods are complex to train, require a full retraining for every condition and the current state-of-the-art avoids this computational complexity by eliminating such pre-processing steps by pre-computing a global descriptor for every image in the database [34], [35].

IV. ADAPTIVE LOCALISATION

An intelligent mobile robot is required that could travel autonomously in various static and dynamic environments. Several techniques have been applied by various researchers for mobile robot navigation and obstacle avoidance. The section will discuss intelligent navigation techniques, which are capable of navigating a mobile robot autonomously in dynamic as well as static environments.

A. Dynamic Sensor Noise and Uncertainty

System identification is the process of designing a mathematical model of a dynamic system through the analysis of measured input and output signals of the system. Optimal estimation uses much of the same formulae as system identification. However, rather than identifying the relationship between the input signals and the system response, optimisation schemes estimate the states of a system [36]. When used in a multi-sensor fusion, optimisation techniques such as KF apply a tunable filter to the multimodal data to reduce the impact of measurement errors on the state estimate and estimate not directly observable system states.

As previously discussed, the traditional KF has proved to be extremely efficient over the years but unfortunately, it needs accurate statistical information about the robot's kinematics to be accurate. It has been demonstrated in some trials that it is possible to set up experiments to calculate these system characteristics [37]. Alternative methods such as Strong Tracking Filtering (STF) [38] offer solutions which do not require such priors.

With the increased diversity of perception platforms due to the increased number of sensor modalities available for data fusion, the risk of hardware and software faults increases in terms of sensor failures, actuators malfunctions, and processing failures. To overcome these issues and detect any faults, a fault tolerant control strategy needs to be developed to ensure more reliable performance outcomes with respect to autonomous systems. Fault tolerant control (FTC) combines diagnosis with control methods in order to handle faults in a systemic way [36].

B. Dynamic Objects

Robust, real-time detection of obstacles such as pedestrians in a robot's path is a critical requirement from a safety standpoint. The recent success achieved in classification, localization and detection tasks by CNNs has led to the adoption of this methodology for pedestrian detection [39]. The recent history of object detection and the super-human accuracy achieved by successive state of the art algorithms is reviewed by [40]. In the domain of autonomous driving, accurate 3D localization and pose estimation of objects beyond 2D boxes are desired. Current state-of-the-art 3D object localisation methods have been developed with the use of use of 3D vision systems including stereovision [41], RGBD cameras [42] and 3d LiDAR [43]. [25]

Obstacle tracking plays an integral part of a robots navigation laws whether it occurs in mapping algorithms, path tracking systems, higher level decision making, iterative learning, planning algorithm implementations or end-to-end deep learning approaches [44].

Object tracking for local collision avoidance involves associating detections corresponding to the same object between successive frames over time and allows estimation of an object's direction and velocity of movement relative to the vision system. Tracking is another vital requirement of autonomous vehicles for predicting the path of moving obstacles in order to make more intelligent decisions regarding their own trajectory. For example, it can be used for person-following as reviewed by [45]. Tracking can be executed with model predictive and sliding mode control [33] or with the use of a combination of data association methods such as Nearest Neighbour for associating detections and Kalman filters for estimating direction and velocity [39]. [25]

What does a robot do if its path through an environment is obstructed completely? Global obstacle avoidance involves the use of mapping algorithms and reasons about free-space, obstacles and the topology of the environment, guided by common sense rules and heuristics for navigation [44].

C. Dynamic Scenes

Hard cases in vSLAM may arise due to significant changes in a setting due to missing or moved structure or the scene may be visually different due to illumination changes, seasonal changes or changing weather conditions. It has been demonstrated by [46] that a feature-based vSLAM solution can effectively meet these challenging requirements.

A feature-based system is also presented by [47] for hybrid map-building and localization, which suits operating environments with unstructuredness and moderate dynamics. The work can accommodate disturbed acquisitions due to sensor vibration and poor stitching.

The previous paragraphs have described cases where a specific scene may change between viewings. Intelligent navigation systems must also be able to cope with different types of scenes, i.e. it must be generalisable to a diverse range of scenes as the type of environment may change drastically as the robot moves from one area to the next. Take for example a robot moving from an outdoor to an indoor scene. When outdoors, the robot may have been able to leverage precise GPS information for localisation and now that that information has a great deal of uncertainty, it must rely on other sensors such as visual information. A similar case may exist in the opposite direction. A vision-based system may have a feature-rich scene when indoors, but once outdoors an open field will not provide enough features for landmark establishment

Many projects have looked at enabling a robot to seamlessly transition between outdoor and indoor environments. Classic approaches include Markov localization, global positioning systems, KF, and fuzzy-logic [29]. Most solutions to this problem use a data quality metric to anticipate and mitigate the degradation in performance of the localisation by discarding the most affected data [48]. KF is still very popular, however, it requires priors specific to the particular robot system. More recent approaches include STF [38] which applies a fading factor to the initial gain matrix and Bayesian Optimisation (BO) [49] which applies a Gaussian distribution as a prior. In each of these cases, STF has demonstrated improvements over KF in GPS localisation and BO has been successfully applied to the task of learning a model of terrain traversability while guiding the robot through more traversable areas [49]. An alternative approach to these optimisation based techniques is a condition-based deep learning architecture presented by [35]. Their solution is an image retrieval-based localisation technique which computes a descriptor for a given image in a way that depends on the capturing conditions. This condition-specific sub-networks based on a siamese architecture was trained using image annotations with the 3D pose for the camera [35].

V. Conclusion

It is evident that the field of multimodal sensor systems for mobile robot navigation has received much attention in recent years and as a result, the performance of perception systems has improved substantially. This paper has included a survey of the most recent works in the implementation of multisensory fusion in localisation techniques for mobile robots in unstructured environments.

The applications examined include vSLAM, odometry and place recognition. Many of the algorithms surveyed in this paper use techniques that take advantage of machine learning as is consistent with many fields of computer science. This paper has also reviewed some techniques which perform localisation in dynamic environments. Each technique tackles the problem of sensor uncertainty due to sensing conditions, difficult scenes and dynamic environments in different ways and present trade-offs in terms of accuracy, scalability and computation/memory requirements. Many projects use optimisation-based solutions to mitigate sensor uncertainty. Other feature-based and image retrieval-based techniques actively detect dynamic objects or dynamic sensing conditions. This field is a very active area of research and will be subject to further change in the future as the increased deployment of sensors such as 3D vision systems and affordable LiDAR will lead to the generation of new datasets, and the increased level of information from these sensors should benefit the performance of algorithms.

It is evident that problems of dynamic sensor uncertainty and the maintenance of scalable, memory efficient maps are very challenging. These problems are key to realising long-term autonomy in mobile robots.

References

[1] P. Addesso, L. Bruno, and R. Restaino, "Adaptive localization techniques in WiFi environments," in *IEEE 5th International Symposium on Wireless Pervasive Computing 2010*, 2010, pp. 289–294.

[2] "System and method for automating beacon location map generation using sensor fusion and simultaneous localization and mapping," May 2017.

[3] B. McLoughlin *et al.*, "Uncertainty Characterisation of Mobile Robot Localisation Techniques using Optical Surveying Grade Instruments," *Sensors*, vol. 18, no. 7, p. 2274, Jul. 2018.

[4] Y. S. Erogluy, I. Guvency, N. Palay, and M. Yukselz, "AOA-based localization and tracking in multi-element VLC systems," in *2015 IEEE 16th Annual Wireless and Microwave Technology Conference (WAMICON)*, 2015, pp. 1–5.

[5] B. Ko, H. Lee, and S. H. Son, "GPS-Less Localization System in Vehicular Networks Using Dedicated Short Range Communication," in *2016 IEEE 22nd International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 2016, pp. 106–106.

[6] H. Du, C. Zhang, Q. Ye, W. Xu, P. L. Kibenge, and K. Yao, "A hybrid outdoor localization scheme with high-position accuracy and low-power consumption," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 4, Dec. 2018.

[7] S. Campbell *et al.*, "Sensor Technology in Autonomous Vehicles - A Review," in *29th Irish Signals and Systems Conference*, 2018.

[8] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," Jan. 2018.

[9] N. O' Mahony *et al.*, "Computer Vision for 3D Perception," in *Intelligent Systems and Applications*, Springer, Cham, 2018, pp. 788–804.

[10] J. Z. Varghese, M. S. E. E. Candidate, and R. G. Boone, "Overview of Autonomous Vehicle Sensors and Systems."

[11] L. Ran *et al.*, "Convolutional Neural Network-Based Robot Navigation Using Uncalibrated Spherical Images," *Sensors*, vol. 17, no. 6, p. 1341, Jun. 2017.

[12] L. Payá, A. Gil, and O. Reinoso, "A State-of-the-Art Review on Mapping and Localization of Mobile Robots Using Omnidirectional Vision Sensors," *Journal of Sensors*, vol. 2017, pp. 1–20, Apr. 2017.

[13] A. Asvadi, "Multi-Sensor Object Detection for Autonomous Driving," University of Coimbra, 2018.

[14] N. Sullivan, S. Grainger, and B. Cazzolato, "Analysis of cooperative localisation performance under varying sensor qualities and communication rates," *Robotics and Autonomous Systems*, vol. 110, pp. 73–84, Dec. 2018.

[15] R. Zhi, "A Drift Eliminated Attitude & Position Estimation Algorithm In 3D."

[16] M. Alatise, G. Hancke, M. B. Alatise, and G. P. Hancke, "Pose Estimation of a Mobile Robot Based on Fusion of IMU Data and Vision Data Using an Extended Kalman Filter," *Sensors*, vol. 17, no. 10, p. 2164, Sep. 2017.

[17] Y.-H. Cheng, Q.-H. Meng, Y.-J. Liu, M. Zeng, L. Xue, and S.-G. Ma, "Fusing sound and dead reckoning for multi-robot cooperative localization," in *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, 2016, pp. 1474–1478.

[18] S. Wang, Z. Deng, G. Yin, S. Wang, Z. Deng, and G. Yin, "An Accurate GPS-IMU/DR Data Fusion Method for Driverless Car Based on a Set of Predictive Models and Grid Constraints," *Sensors*, vol. 16, no. 3, p. 280, Feb. 2016.

[19] L. Carlone and S. Karaman, "Attention and Anticipation in Fast Visual-Inertial Navigation."

[20] X. Hai-Xia, Z. Wei, and Z. Jiang, "3D visual SLAM with a Time-of-Flight camera," in *2015 IEEE Workshop on Signal Processing Systems (SiPS)*, 2015, pp. 1–6.

[21] Z. Shang and Z. Shen, "Real-time 3D Reconstruction on Construction Site using Visual SLAM and UAV," Dec. 2017.

[22] R. Kannan Megalingam, C. Ravi Teja, S. Sreekanth, and A. Raj, "ROS based Autonomous Indoor Navigation Simulation Using SLAM Algorithm," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 7, pp. 199–205, 2018.

[23] C. Häne *et al.*, "3D Visual Perception for Self-Driving Cars using a Multi-Camera System: Calibration, Mapping, Localization, and Obstacle Detection," Aug. 2017.

[24] V. Murali, H.-P. Chiu, S. Samarasekera, Rakesh, and Kumar, "Utilizing Semantic Visual Landmarks for Precise Vehicle Navigation," Jan. 2018.

[25] N. O' Mahony *et al.*, "Deep Learning for Visual Navigation of Unmanned Ground Vehicles; A review," in *29th Irish Signals and Systems Conference*, 2018.

[26] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-Manifold Preintegration Theory for Fast and Accurate Visual-Inertial Navigation," *arXiv preprint arXiv:1512.02363v1*, no. arXiv:1512.02363v1, 2015.

[27] V. Peretroukhin, L. Clement, M. Giamou, and J. Kelly, "PROBE: Predictive Robust Estimation for Visual-Inertial Navigation," Aug. 2017.

[28] J. Thoma, D. Pani Paudel, A. Chhatkuli, T. Probst, and L. Van Gool, "Image-based Navigation using Visual Features and Map," 2018.

[29] H. Xiong, J. Tang, H. Xu, W. Zhang, and Z. Du, "A Robust Single GPS Navigation and Positioning Algorithm Based on Strong Tracking Filtering," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 290–298, Jan. 2018.

[30] J. Miseikis, I. Brijacak, S. Yahyanejad, K. Glette, O. J. Elle, and J. Torresen, "Multi-Objective Convolutional Neural Networks for Robot Localisation and 3D Position Estimation in 2D Camera Images," Apr. 2018.

[31] J. Miseikis *et al.*, "Robot Localisation and 3D Position Estimation Using a Free-Moving Camera and Cascaded Convolutional Neural Networks," Jan. 2018.

[32] K. Tateno, F. Tombari, I. Laina, and N. Navab, "CNN-SLAM: Real-Time Dense Monocular SLAM with Learned Depth Prediction," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6565–6574.

[33] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization methods: A survey," *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, Feb. 2015.

[34] M. Angelina, U. Gim, and H. Lee, "PointNetVLAD: Deep Point Cloud Based Retrieval for Large-Scale Place Recognition."

[35] H. Germain, G. Bourmaud, and V. Lepetit, "Efficient Condition-based Representations for Long-Term Visual Localization."

[36] M. El-Gindy, A. Mohamed, J. Ren, H. Lang, and A. N. Ouda, "Literature survey for autonomous vehicles: Sensor fusion, computer vision, system identification and fault tolerance," *Article in International Journal of Automation and Control*, vol. 12, no. 4, pp. 555–581, 2018.

[37] S. M. C. Porto, C. Arcidiacono, A. Giummarra, U. Anguzza, and G. Cascone, "Localisation and identification performances of a real-time location system based on ultra wide band technology for monitoring and tracking dairy cow behaviour in a semi-open free-stall barn," *Computers and Electronics in Agriculture*, 2014.

[38] H. Xiong, J. Tang, H. Xu, W. Zhang, and Z. Du, "A Robust Single GPS Navigation and Positioning Algorithm Based on Strong Tracking Filtering," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 290–298, Jan. 2018.

[39] D. Ribeiro, A. Mateus, P. Miraldo, and J. C. Nascimento, "A Real-Time Deep Learning Pedestrian Detector for Robot Navigation," Jul. 2016.

[40] A. Garcia-Garcia, S. Orts-Escolano, S. O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation," *arXiv preprint arXiv:1704.06857v1*, no. arXiv:1704.06857v1, 2017.

[41] C. C. T. Mendes, V. Fremont, and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3174–3179.

[42] J. Thörnberg, "Combining RGB and Depth Images for Robust Object Detection using Convolutional Neural Networks," KTH Royal Institute of Technology, 2015.

[43] B. Li, T. Zhang, and T. Xia, "Vehicle Detection from 3D Lidar Using Fully Convolutional Network," *arXiv preprint arXiv: 1608.07916*, Aug. 2016.

[44] M. Hoy, A. S. Matveev, and A. V Savkin, "Algorithms for collision-free navigation of mobile robots in complex cluttered environments: a survey," *Robotica*, vol. 33, pp. 463–497, 2015.

[45] M. J. Islam, J. Hong, and J. Sattar, "Person Following by Autonomous Robots: A Categorical Overview," *arXiv preprint arXiv:1803.08202v1*, 2018.

[46] C. Enright, "Visual SLAM and Localization – The Hard Cases," *Electronic Imaging*, vol. 2018, no. 17, pp. 281-1-281–5, Jan. 2018.

[47] S. Rady, "Vision-Based Hybrid Map-Building and Robot Localization in Unstructured and Moderately Dynamic Environments," Springer, Cham, 2016, pp. 231–250.

[48] J. Li, H. Cheng, H. Guo, and S. Qiu, "Survey on Artificial Intelligence for Vehicles," *Automotive Innovation*, pp. 1–13, Mar. 2018.

[49] R. Oliveira, L. Ott, V. Guizilini, and F. Ramos, "Bayesian Optimisation for Safe Navigation under Localisation Uncertainty," Sep. 2017.