Journal of Engineering Sciences

Vol 13, Issue 04, APRIL /2022

ISSN NO:0377-9254

Obstacle-Avoidance Algorithm Using Deep Learning Based on RGBD Images and Robot Orientation

Vipin Sagar Nagelli, BTech, CMR Institute of Technology, Department of Mechanical Engineering, vipinsagar2807@gmail.com

Kondamu Sai Teja, BTech, CMR Institute of Technology, Department of Mechanical Engineering, kondamusaiteja@gmail.com

ABSTRACT: Inspired by the advantages of the hierarchical feature extraction of deep learning, this work investigates the development of a Convolutional Neural Network (CNN) algorithm to solve the problem of the mobile robot obstacle avoidance in an indoor environment. The algorithm takes raw images and robot orientation as input and generates control commands as network output. Control commands include gostraight-forward, turn-full-left, turn-half-left, turn-full-right, and turn-half-right. A dataset compiled using depth images (RGBD) and robot orientation data obtained by an Inertial Measurement Unit (IMU). In addition, the performance of the algorithm in terms of training options, hyperparameters, and output precision is evaluated and recommendations are provided accordingly. The final results show that the accuracy can be improved by including the robot orientation in the dataset, increasing the size of data, and tuning the network's hyperparameters. The CNN algorithm has shown great potential to get high path classification accuracy for obstacle avoidance for mobile robots.

Keywords- *mobile robot; deep learning; obstacle avoidance*

1. INTRODUCTION

Navigation is one of the most important tasks of mobile robots, where staying operational while avoiding collisions and maintaining safety standard are priorities in mobile robots [1]. To develop an autonomous mobile robot, we need to build a system that can grasp environments, react to unexpected events, and plan dynamically in order to achieve the mission. Thus, the objective of the robot's motion planning and control is to find collision-free paths between two positions in static or dynamic environments. In this context, control has different levels; namely, (1) motor control (low level) and (2) behavior or mission control (high level). The latter includes many complicated tasks, such as obstacle avoidance and goal seeking [2]. Despite numerous breakthroughs in the field of autonomous robotics in the last two decades, robot navigation is still an area of active research due to the uncertainties involved with unknown real-life environments. These uncertainties could be attributed to any of the following: (i) no prior knowledge of the environment, (ii) perceptually acquired information is usually unreliable, (iii) unpredictable complex dynamical environment, and (iv) effect of control actions is not completely reliable [3, 4]. Classical navigation approaches relied on geometric models like constructing local cost-maps, which are considered as low-level intelligence without any perception process [5, 6]. The perception of the environment relies on various sensors information.

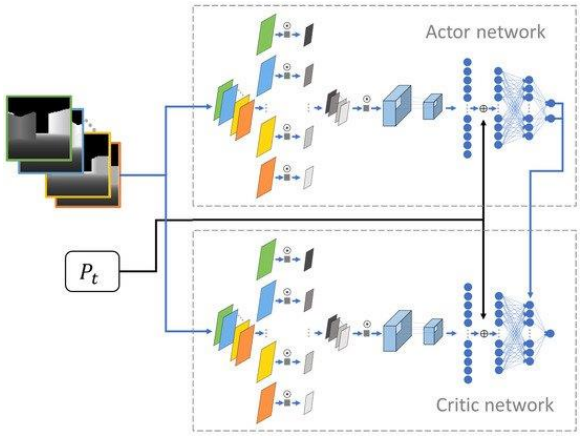

Fig.1: Full architecture of the network including the actor and critic parts

With the continuous development of science and technology, various mobile robots have been widely used in different fields, such as life services, industrial production, education, entertainment and military area, etc. The mobile robot technology

include control theory, mechanical design, computer technology. The ability of mobile robots to navigate and avoid obstacles is an important indicator of the robot's intelligence. Autonomous navigation and obstacle avoidance for mobile robot need to equip with some range sensors and depend on complex algorithms[1]. Some typical range sensors such laser sensors, ultrasonic sensors, and visual sensors, while these sensors have their own limitations. For example, lasers are more expensive and traditional algorithms based vision are relatively complex.

2. LITERATURE REVIEW

2.1 Motion control for mobile robot obstacle avoidance and navigation: a fuzzy logic-based approach. Systems Analysis Modelling Simulation:

One of the ultimate goals of mobile robotics research is to build robots that can safely carry out missions in hazardous and populated environments. Most of today's commercial mobile devices scale poorly along this dimension. Their motion planning relies on accurate, static models of the environments, and therefore they often fail their mission if humans or other unpredictable obstacles block their path. To build autonomous mobile robots one has to build systems that can perceive their environments, react to unforeseen circumstances, and plan dynamically in order to achieve their mission. Thus, the objective of the motion planning and control problem is to find collision-free trajectories, in static or dynamic environments containing some obstacles, between a start and a goal configuration. It has attracted much research in recent years. In this context the term control has a broad meaning that includes many different controls, such as low-level motor control, and behaviour control, where behaviour represents many complicated tasks, like obstacle avoidance and goal seeking. This article describes an intelligent motion planning and navigation system for omnidirectional mobile robots based on fuzzy logic. Owing to its simplicity and hence its short response time, the fuzzy navigator is especially suitable for on-line applications with strong real-time requirements. On-line planning is an on-going activity. The planner receives a continuous flow of information about occurring events and generates new commands in response to the incoming events, while previously planned motions are being executed. The fuzzy-rule-base of the proposed system combines the repelling influence, which is related to the distance and the angle between the robot and nearby obstacles, with the attracting influence produced by the distance and the angular difference between the actual direction and position of the robot and the final configuration, to generate actuating commands for the mobile platform. It can be considered as an on-line local navigation method for omnidirectional mobile robots for the generation of instantaneous collision-free motions. This reactive system is especially suitable for real-time applications. The use of fuzzy logic leads to a transparent system which can be tuned by hand or by a set of learning rules. Furthermore, this approach allows obstacle avoidance and navigation in dynamic environments. The functioning of the fuzzy motion planner with respect to omnidirectional mobile robots and results of simulated experiments are presented.

2.2 Computational principles of mobile robotics:

This is a textbook for advanced undergraduate and graduate students in the field of mobile robotics. Emphassising computation and algorithms, the authors address a range of strategies for enabling robots to perform tasks that involve motion and behavior. The book is divided into three major sections: locomotion, sensing and reasoning. It concentrates on wheeled and legged mobile robots, but discusses a variety of other propulsion systems. Kinematic models are developed for many of the more common locomotive strategies. It presents algorithms for both visual and nonvisual sensor technologies, including sonar, vision and laser scanners. In the section on reasoning, the authors offer a thorough examination of planning and the issues related to spatial representation. They emphasize the problems of navigation, pose estimation, and autonomous exploration. The book is a comprehensive treatment of the field, offering a discussion of state-of-the art methods with illustrations of key technologies.

2.3 The uses of fuzzy logic in autonomous robot navigation:

The development of techniques for autonomous navigation in real-world environments constitutes one of the major trends in the current research on robotics. An important problem in autonomous navigation is the need to cope with the large amount of uncertainty that is inherent of natural environments. Fuzzy logic has features that make it an adequate tool to address this problem. In this paper, we review some of the possible uses of fuzzy logic in the field of autonomous navigation. We focus on four issues: how to design robust behavior-producing modules; how to coordinate the activity of several such modules; how to use data from the sensors; and how to integrate high-level reasoning and low-level execution. For each issue, we review

some of the proposals in the literature, and discuss the pros and cons of fuzzy logic solutions.

2.4 LSD-SLAM: Large-scale direct monocular SLAM:

We propose a direct (feature-less) monocular SLAM algorithm which, in contrast to current state-of-the-art regarding direct methods, allows to build large-scale, consistent maps of the environment. Along with highly accurate pose estimation based on direct image alignment, the 3D environment is reconstructed in real-time as pose-graph of keyframes with associated semi-dense depth maps. These are obtained by filtering over a large number of pixelwise small-baseline stereo comparisons. The explicitly scale-drift aware formulation allows the approach to operate on challenging sequences including large variations in scene scale. Major enablers are two key novelties: (1) a novel direct tracking method which operates on sim(3), thereby explicitly detecting scale-drift, and (2) an elegant probabilistic solution to include the effect of noisy depth values into tracking. The resulting direct monocular SLAM system runs in real-time on a CPU.

2.5 ORB-SLAM: a versatile and accurate monocular SLAM system:

This paper presents ORB-SLAM, a feature-based monocular simultaneous localization and mapping (SLAM) system that operates in real time, in small and large indoor and outdoor environments. The system is robust to severe motion clutter, allows wide baseline loop closing and relocalization, and includes full automatic initialization. Building on excellent algorithms of recent years, we designed from scratch a novel system that uses the same features for all SLAM tasks: tracking, mapping, relocalization, and loop closing. A survival of the fittest strategy that selects the points and keyframes of the reconstruction leads to excellent robustness and generates a compact and trackable map that only grows if the scene content changes, allowing lifelong operation. We present an exhaustive evaluation in 27 sequences from the most popular datasets. ORB-SLAM achieves unprecedented performance with respect to other state-of-the-art monocular SLAM approaches. For the benefit of the community, we make the source code public.

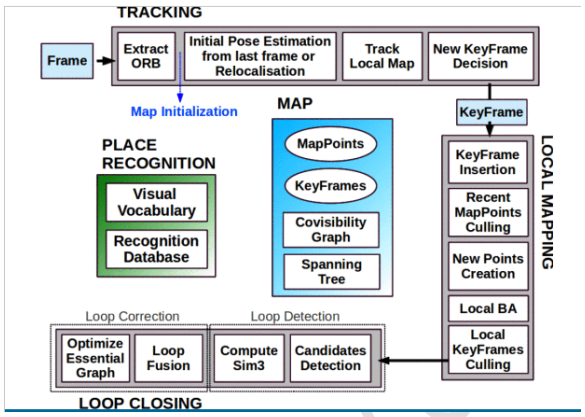

Fig.2: ORB-SLAM system overview, showing all the steps performed by the tracking, local mapping, and loop closing threads.

2.6 Object recognition from local scale-invariant features:

An object recognition system has been developed that uses a new class of local image features. The features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. These features share similar properties with neurons in inferior temporal cortex that are used for object recognition in primate vision. Features are efficiently detected through a staged filtering approach that identifies stable points in scale space. Image keys are created that allow for local geometric deformations by representing blurred image gradients in multiple orientation planes and at multiple scales. The keys are used as input to a nearest neighbor indexing method that identifies candidate object matches. Final verification of each match is achieved by finding a low residual least squares solution for the unknown model parameters. Experimental results show that robust object recognition can be achieved in cluttered partially occluded images with a computation time of under 2 seconds.

2.7: Surf: Speeded up robust features:

In this paper, we present a novel scale- and rotation-invariant interest point detector and descriptor, coined SURF (Speeded Up Robust Features). It approximates or even outperforms previously proposed schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster. This is achieved by relying on integral images for image convolutions; by building on the strengths of the leading existing detectors and descriptors (in casu, using a Hessian

matrix-based measure for the detector, and a distribution-based descriptor); and by simplifying these methods to the essential. This leads to a combination of novel detection, description, and matching steps. The paper presents experimental results on a standard evaluation set, as well as on imagery obtained in the context of a real-life object recognition application. Both show SURF's strong performance.

2.8: BRIEF: Binary Robust Independent Elementary Features:

We propose to use binary strings as an efficient feature point descriptor, which we call BRIEF. We show that it is highly discriminative even when using relatively few bits and can be computed using simple intensity difference tests. Furthermore, the descriptor similarity can be evaluated using the Hamming distance, which is very efficient to compute, instead of the L2 norm as is usually done. As a result, BRIEF is very fast both to build and to match. We compare it against SURF and U-SURF on standard benchmarks and show that it yields a similar or better recognition performance, while running in a fraction of the time required by either.

3. IMPLEMENTATION

Classical perception methods extract information from the raw sensor readings based on artificially designed complex features (e.g., Scale Invariant Feature Transform (SIFT) [7, 8], Speed Up Robust Feature (SURF) [9], Binary Robust Independent Elementary Features (BRIEF) [10], and Oriented FAST and Rotated BRIEF (ORB) [11]). Most of those methods are designed to adapt to generic environments. However, these methods are prone to errors when they encounter unstructured dynamic environments, wide illumination differences, and different terrain types with geometrically similar structures. This work is concerned with the problem of obstacle avoidance in indoor environments for vision-based mobile robots. The main goal is to develop a deep learning algorithm for obstacle avoidance using raw images of robot environment as input and generate control commands as network output, to solve the problem of real-time robot navigation. In order to achieve the main objective, a new dataset is compiled using depth images (RGBD) and robot orientation data obtained by an Inertial Measurement Unit (IMU). RGBD is an image channel in which each pixel relates to a distance between the image plane and the corresponding object in the RGB image. On the other hand, IMU is used to estimate the orientation of the robot while navigating through obstacles. The baseline control algorithm chosen in this work is the Convolutional Neural Network (CNN) that should predict control commands from the images received in real time from the camera. In addition, the performance of the algorithm in terms of training time and output precision is evaluated and recommendations is provided accordingly.

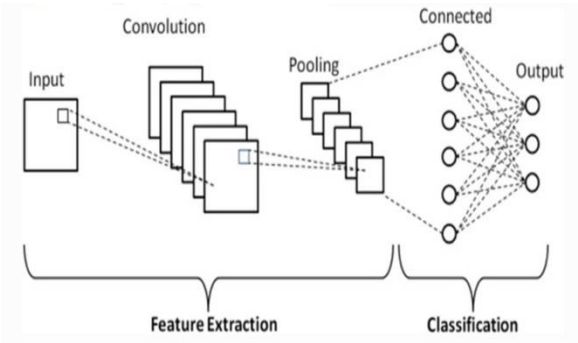

Fig.3: Proposed architecture

MODULES:

1. Data Collection: Collect sufficient data samples and legitimate software samples.

2. Data Preprocessing: Data Augmented techniques will be used for better performance

3. Train and Test Modelling: Split the data into train and test data Train will be used for training the model and Test data to check the performance.

4. Modelling: VGG16 and CNN model build and model is saved

5. Predict: Select an single image and do basic image processing and predict using VGG16 model.

4. ALGORITHMS

CNN is also computationally efficient. It uses special convolution and pooling operations and performs parameter sharing. This enables CNN models to run on any device, making them universally attractive. All in all this sounds like pure magic. We are dealing with a very powerful and efficient model which performs automatic feature extraction to achieve superhuman accuracy (yes CNN models now do image classification better than humans). Hopefully this article will help us uncover the secrets of this remarkable technique.

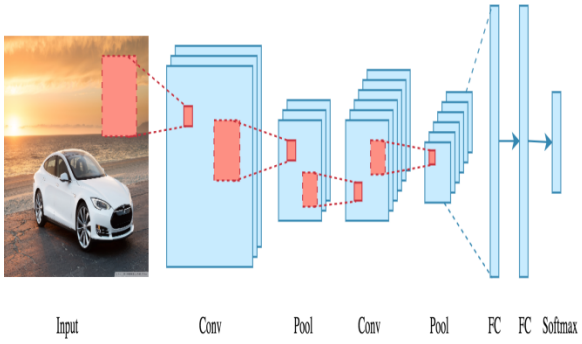

Fig.4: CNN architecture

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU's.

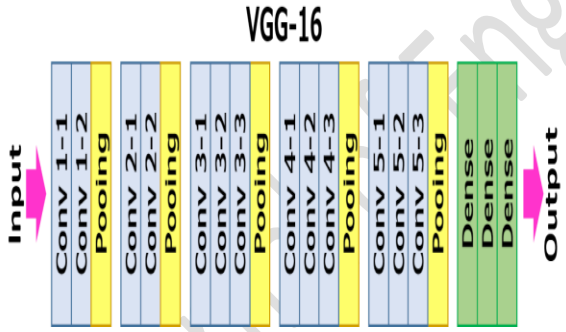

Fig.5: VGG-16 architecture

5. IMAGE CLASSIFICATION USING CNN

This work combines the perception stage and the control stage with a single deep network. The network structure fuses CNN with the decision-making process. The adopted CNN algorithm contains three convolutional stages; each stage has three layers "convolution + activation + pooling", followed by one fully connected layer. The CNN structure is used to detect and understand visual features and the fully connected layers are for decision-making. The first convolution layer uses 32 5×5 filters, followed by a ReLu layer and a 2×2 pooling layer with stride 2. The second stage has a similar structure to the first stage. For the third stage, 64 5×5 filters are used, with no change of the ReLu layer and pooling layer. This results in 64 feature maps of size 20×15. The output control commands consist of five states: one for going straight forward, two for turning left, and two for turning right. The final decision is calculated by applying soft-max classifier using the outcome of the last layer as coefficiencies over the five possible control commands. The great success of the CNN is attributed to the excellent performance of trained hierarchical structure in data representations [12].

The objective of the training process is to learn the network's weights. The weights of CNN are typically calculated by back-propagation and gradient descent. Training data and loss function are the two main elements needed to train any artificial neural network. The training data is composed of images and the corresponding labels, while the loss function measures the inaccuracy of predictions.

5. EXPERIMENTAL RESULTS

Before starting the training process, a data-preparation step has to be done. During this step processes such as deduplication, removing blank images, and image resizing should be carried out. First, the collected data was extracted from the rosbag files format to CSV files format using Python code. For each rosbag file, the code exports each topic to individual CSV file, which contains two main variables; sample time and variable names. The goal of extracting the data from rosbag file into CSV format is only to find the timestamp of each entry so that it matches the RGBD images with its corresponding IMU data. However, this code does not extract the images itself; therefore, another Python code was used in order to extract the RGBD images.

After extracting the datasets from the rosbag files, the following preprocessing steps were performed:

1. Removing all duplicated images from each scenario. The total number of RGBD images after this process is 4,347.

2. Resizing the images resulted from the previous step. The Kinect camera generates RGBD images with a size of 640×480. The input size is downsampled to a quarter of the original size, such that the new resolution becomes 160×120. Even at such small resolution, it is possible to see the path

and distinguish its features. With such size, the operation will largely reduce the computational cost without affecting the accuracy.

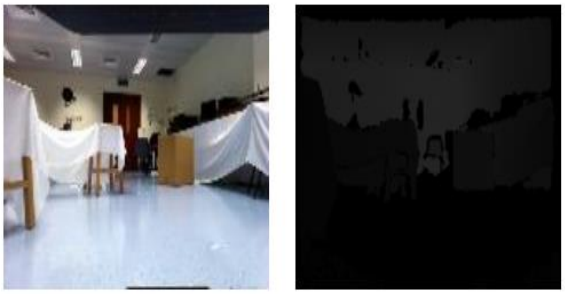

Fig.6: Example of RGB image.

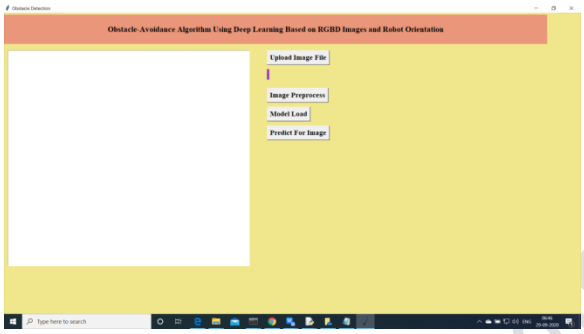

Fig.7: Home screen

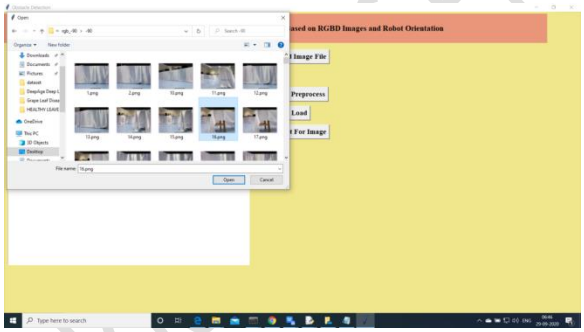

Fig.8: Uploading screen

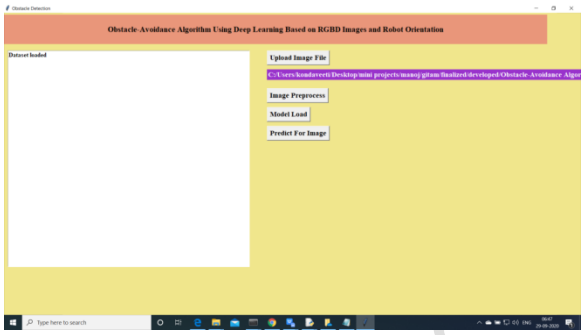

Fig.9: Dataset loaded

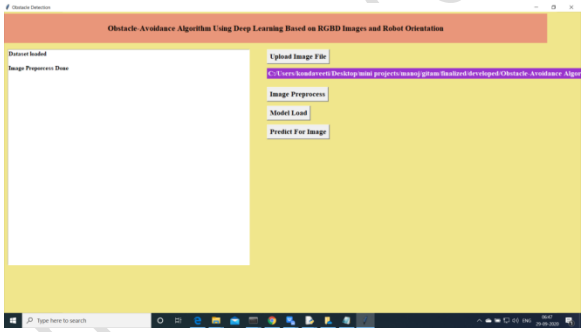

Fig.10: model trained using image processing

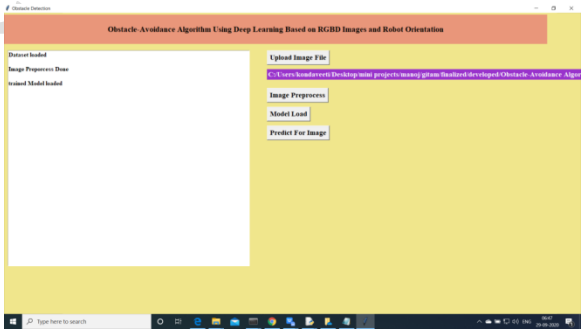

Fig.11: Load the model

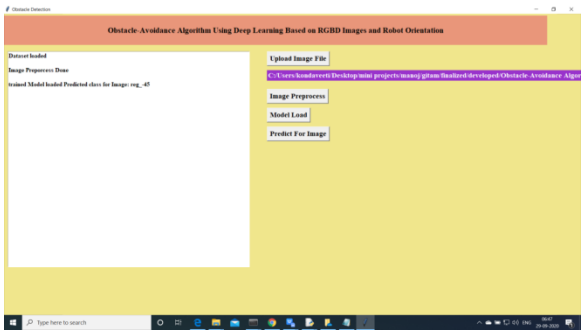

Fig.12: Predict using model

5. CONCLUSION

An obstacle avoidance approach for a mobile robot in indoor environments based on a convolution neural network (CNN) has been developed. A dataset was compiled that contains depth images and robot orientation using a mobile robot platform. The RGBD images obtained by the Kinect camera and the angles acquired by the IMU were recorded. Based on angles, the data was sampled and discretized into five path labels. To the best of our knowledge, this is the first RGBD images dataset that is labeled with IMU data that are used for CNN structures. The results show that our system could manage the obstacle avoidance of mobile robots with an accuracy of approximately 85%. Although it is not completely successful at predictions, it shows a good classification accuracy with greater success than prior works, where the accuracy did not exceed 81.7%. Moreover, the proposed system in this work has quite a low chance to generate a totally opposite decision. Therefore, we believe that the CNN trained by our dataset can result in high classification accuracy in avoiding obstacles in real-time.

6. FUTURE SCOPE

In the next work, we will attempt to do it in a more complex setting, include dynamic and non-structure environment. We will solver the tasks of robot navigation based CNN model and provide a significant contribution for the development of intelligence mobile robotic navigation.

REFERENCES

[1] R.Siegwart, I.R.N., Introduction to Autonomous Mobile Robot. Massachusetts Institute of Technology Press, Cambridge, USA, 2004.

[2] Zavlangas, P.G. and S.G. Tzafestas, Motion control for mobile robot obstacle avoidance and navigation: a fuzzy logic-based approach. Systems Analysis Modelling Simulation, 2003. 43(12): p. 1625-1637.

[3] Gregory, D. and J. Michael, Computational principles of mobile robotics. Cambridge University, 2000.

[4] Saffiotti, A., The uses of fuzzy logic in autonomous robot navigation. Soft Computing-A Fusion of Foundations, Methodologies and Applications, 1997. 1(4): p. 180-197.

[5] Engel, J., T. Schöps, and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. in European Conference on Computer Vision. 2014. Springer.

[6] Mur-Artal, R., J.M.M. Montiel, and J.D. Tardos, ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Transactions on Robotics, 2015. 31(5): p. 1147-1163.

[7] Guzel, M.S. and R. Bicker, Vision based obstacle avoidance techniques, in Recent advances in mobile robotics. 2011, InTech.

[8] Lowe, D.G. Object recognition from local scale-invariant features. in Computer vision, 1999. The proceedings of the seventh IEEE international conference on. 1999. Ieee.

[9] Bay, H., T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. in European conference on computer vision. 2006. Springer.

[10] Calonder, M., et al. Brief: Binary robust independent elementary features. in European conference on computer vision. 2010. Springer.

[11] Rublee, E., et al. ORB: An efficient alternative to SIFT or SURF. in Computer Vision (ICCV), 2011 IEEE international conference on. 2011. IEEE.

[12] Tai, L., et al. PCA-aided fully convolutional networks for semantic segmentation of multi-channel fMRI. in Advanced Robotics (ICAR), 2017 18th International Conference on. 2017. IEEE.

[13] Beale, M.H., M.T. Hagan, and H.B. Demuth, Neural Network Toolbox™ User's Guide. The Mathworks Inc, 1992.

[14] Murphy, K.P., Machine learning: a probabilistic perspective. 2012.

[15] Tai, L., S. Li, and M. Liu. A deep-network solution towards modelless obstacle avoidance. in Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on. 2016. IEEE