International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-7, May 2021

Autonomous Indoor Navigation for Mobile Robots

Karthik Valliappan C, Vikram R

Abstract*: An autonomous navigation system for a robot is key for it to be self-reliant in any given environment. Precise navigation and localization of robots will minimize the need for guided work areas specifically designed for the utilization of robots. The existing solution for autonomous navigation is very expensive restricting its implementation to satisfy a wide variety of applications for robots. This project aims to develop a low-cost methodology for complete autonomous navigation and localization of the robot. For localization, the robot is equipped with an image sensor that captures reference points in its field of view. When the robot moves, the change in robot position is estimated by calculating the shift in the location of the initially captured reference point. Using the onboard proximity sensors, the robot generates a map of all the accessible areas in its domain which is then used for generating a path to the desired location. The robot uses the generated path to navigate while simultaneously avoiding any obstacles in its path to arrive at the desired location.

***Keywords*: Autonomous, Self-Reliant, Localization, Expensive, Image Sensor, Simultaneously.**

I. INTRODUCTION

A fundamental task in autonomous robots is to move safely in a random environment. Autonomous robot navigation has seen significant progress in the past few years, particularly navigation involving SLAM - Simultaneous Localization and Mapping. Predominant implementation of SLAM is related to range-sensing SLAM which uses the surrounding range of the robot as the feature reference to localize the robot while simultaneously generating a map of the robot surrounding. This method is of SLAM is highly dependent on the accuracy of the range-finding sensors since inaccurate readings will drastically affect the output of the system.

To examine SLAM more cost-effectively, optical features of the robot surroundings can be utilized. Visually guided systems have been in development for a decade. There are predominantly two different ways to implement this system - i) add artificial landmarks to be used by the robots as a reference point - ii) Use the camera to capture natural landmark features as the reference points for the robot. A camera can be attached to the robot to capture the image of the robot's surroundings, which then is processed to obtain reference points in the image. When the robot moves, the Change in the position of the reference point between multiple frames of images will estimate the change in the robot's position.

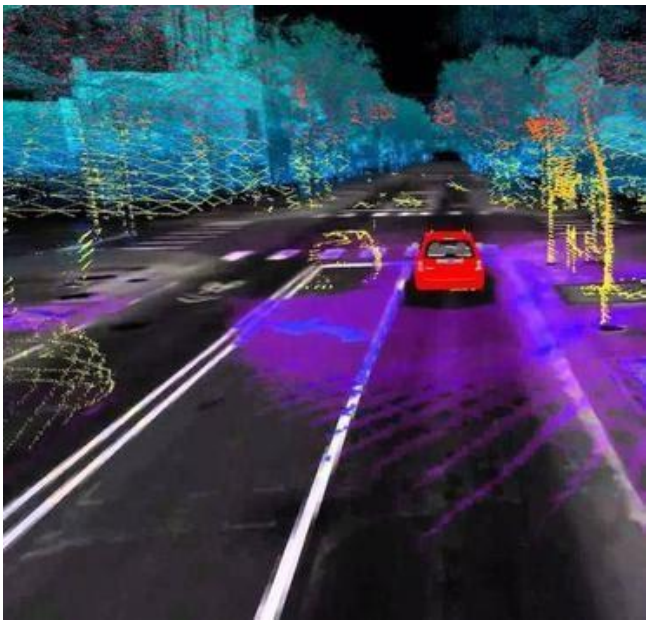

Figure 1.Visualization of Sensor data

The reference point is calculated by a convolutional neural network layer and the shift of position concerning the reference point is calculated by a long-term, short memory neural network which is also a type of recurrent neural network where the previous step output is given as input in the next step. As we use deep learning concepts, we have to first train the dataset by feeding in a different set of input images to recognize the reference point. The training process may or may not be slower depending upon the size of the data input but once the training process is complete, the image sensor recognizes the images very quickly so thus it can be used as a replacement for LIDAR.

II. SLAM

Simultaneous localization and mapping (SLAM) is the most predominant implementation for location awareness and recording of the environment in a map of any autonomous vehicle. Most robots with SLAM integration are primarily dependent upon a laser range finder for localization and mapping the environment. But the laser range finder has to very accurate since SLAM is ill-conditioned. The high accurate laser range finder is very expensive limiting its implementation in a wide variety of robots.

Manuscript received on May 21, 2021.
Revised Manuscript received on May 26, 2021.
Manuscript published on May 30, 2021.
* Correspondence Author
Karthik Valliappan C*, Robotics and automation Department, PSG College of technology, Coimbatore, India. Email: ckarthikv@gmail.com
Vikram R, Robotics and automation Department, PSG College of technology, Coimbatore, India. Email: viki.ares6@gmail.com

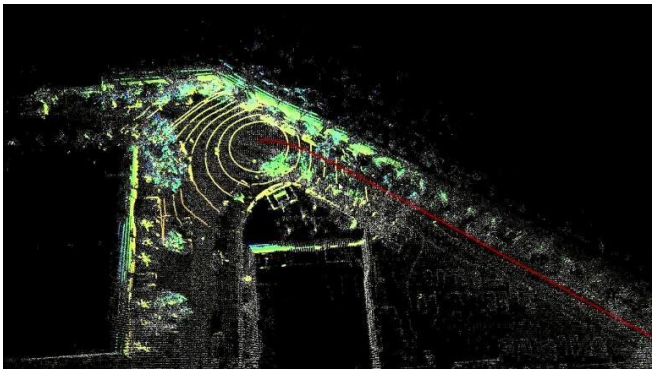

Figure 2.Visualization of LIDAR data points

This paper is based on visual SLAM, where the primary mode of sensing is through a camera, this method is a comparatively more cost-effective method of SLAM implementation in robots. Both the robot's position and the proximity sensor reading are essential for SLAM to run. The different tasks involved to execute SLAM require feature-based image tracking because the feature image is known ahead. The map recovered when the robot is in motion is the combination of maps at different robot locations. The challenge in SLAM is to recover both camera pose and map structure while initially knowing neither.

Figure 3.Multiple object detection

III. PROSPOSED METHODOLOGY

First, the training of CNN and LSTM is done by providing it with images of random objects. When the overall training accuracy is around 90%, the training is stopped and the weights of the neural networks are frozen. After the completion of training, the neural networks are then passed

Onto the Open-cv python library for multiple object detection. The detected objects are sent to the VSLAM Robot operating system package for depth measurement and further

it is combined with the map generated by the image sensor in the G-mapping package included in ROS and now the combined data acts as input to the AMCL algorithm for navigating to a goal point with obstacle avoidance.

A. Convolution Neural Network

A convolution Neural Network (CNN) is a deep learning algorithm in which its inputs, outputs, and hidden layers are arranged in a way that matches the approximate architecture of a human brain. It is mainly used for image recognition and also for object edge detection. CNN has more accuracy when it deals with the MNIST data type. CNN networks use weights as an input and also a varying parameter. Weights are the frequency of transmission of data through the different layers of neural networks. Whenever data is passed onto another layer its weight parameter is frequently updated. CNN consists of 5 main operations which are Subsampling, pooling, RELU operation, Activation function, and finally flatten features.

First, an image is given as an input to a neural network. The computer identifies the input image as a two-dimensional matrix that contains its value in the range 0-255.

Figure 4 Matrix format of a picture

The final result of the matrix multiplication becomes the feature maps.

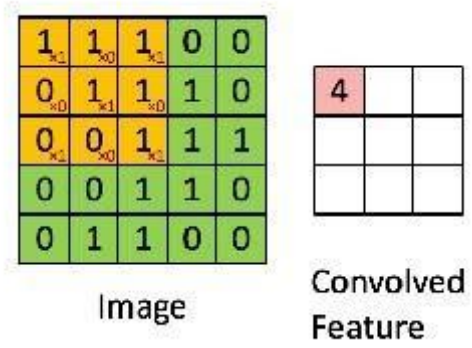

Figure 5.Convolution operation

RELU stands for Rectified Linear Unit. The output is $f(x) = \mathbf{Max(0,x)}$.The RELU operation is used to remove any negative values formed in the feature maps. The RELU operation is important because RELU's goal is to introduce non-linearity in our CNN. Since the real-world data would want our CNN Net to learn from non-negative linear values. There are other nonlinear functions such as tanh or sigmoid that can also be used instead of RELU. Most of the data scientists use RELU since performance-wise RELU is more reliable than the other two.

When RELU operation is done now it sends to the pooling layer where it can consist of three types which are Max pooling, average pooling, and sum pooling.

Max pooling takes the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature map call as sum pooling.

After all these above steps are done the matrix is flattened into a vector and feed into a fully connected layer feed-forward neural network.

Retrieval Number: 100.1/ijitee.G90380510721
DOI: 10.35940/ijitee.G9038.0510721
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

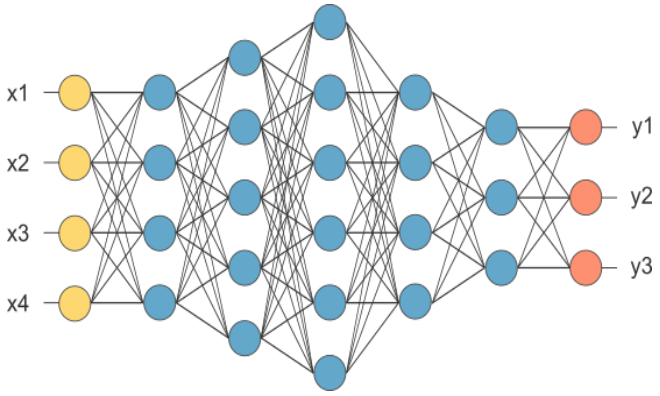

Figure 6.Fully connected neural network

The above steps are part of the training of neural networks. After the training process, the dynamic image is again sent to the neural network to compare and check with the previous models.

And thus object edges are detected and also it is recognized.

B. Recurrent Neural Networks

A recurrent Neural Network is a type of Neural Network where the output from the previous step is given as input to the present step. In traditional neural networking methods, all the inputs and outputs are independent of each other, but in specific cases like when it is required to predict the next word of a sentence, the previous words are necessary and hence there is a need for remembering the previous output data. Thus RNN came into existence, which solved this problem with the help of a Hidden Layer. The most important feature of RNN is the Hidden state, which remembers some information about a sequence. So thus Rnn can be used for object tracking concerning the reference point of a previous position.

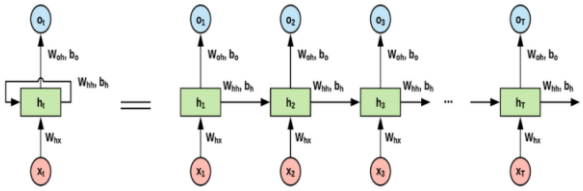

Figure 7.Example of RNN

Training the RNN

1. A single input is provided initially.
2. Then it calculates the next state using a set of current input and the previous state.
3. The current h(t) becomes h(1) for the next time step.
4. There can be any number of time steps according to the problem and join the information from all the previous states.
5. When every time steps are completed the final current state is used to calculate the output.
6. The output is then compared to the actual output for error detection.
7. The error is then back-propagated to the first layer network to update the weights and finally, the network (RNN) is trained.

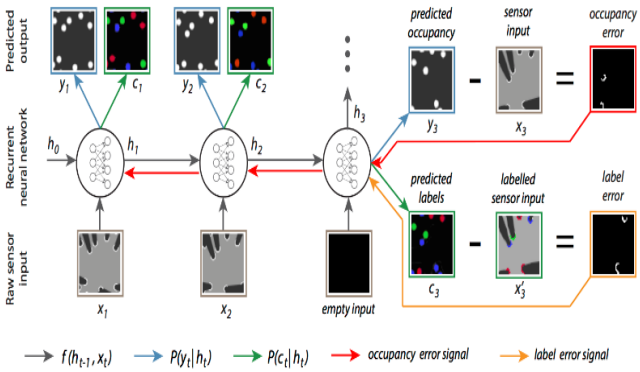

Figure 8.Training process of RNN

C. Long Term Short Memory

Long Term Short Memory (LSTM) is also the type of Re-current neural network but it helps to remove the error that can be back propagated through layers. LSTMs, unlike other networks they contain data outside the normal flow of the recurrent network in a gated cell. These gated cell act like a digital switch and makes decisions about what to store, and when to allow read, write and erase. However, these gates are analogous as it is implemented with element-wise multiplication by sigmoid function, which is all in the range of 0-1 which are different from the digital storage of computers. Analog has more advantage over digital where it can be easily differentiable which is eventually suitable for backpropagation.

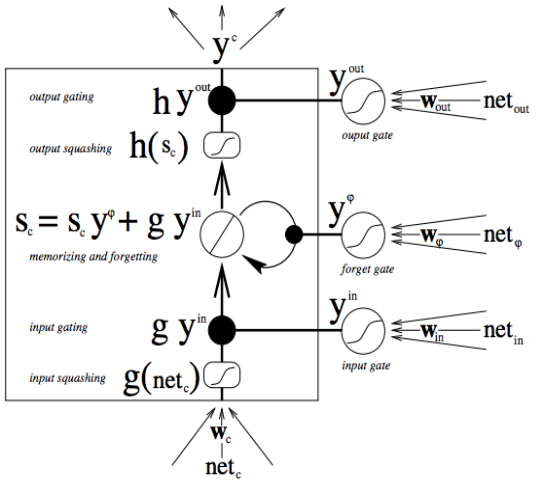

Figure 9.LSTM neural network representation

These gates act according to the signals from what they receive, and similar to the neural network's nodes, they block or pass on information based on its data strength and import, which they filter with their pair of weights. LSTM also removes the error which happens in RNN which are vanishing and exploding.

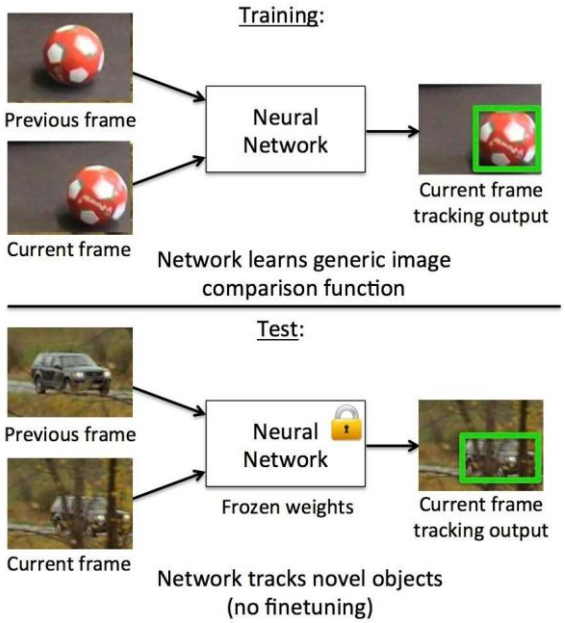

Figure 10.Object tracking

The Convolutional neural network (CNN) layer in the neural network is used for image classification and the Long term Short term memory(LSTM) layer is used for reference point tracking, Both of these layers are combined to find the distance between the object and the robot so that the robot can navigate to a goal point with obstacle avoidance.

IV. LOCALISATION AND MAPPING

For localization, a monocular tracking system – Mono SLAM sanction the reconstruction of every frame in a special plane of the set of reference points:

Eq. (1) $n=Pi$

Where $1 \leq i \leq n$

Eq. (2) $Pi=(xi,yi,zi)^T$

Figure 11. Object contour detection

It enables to track the configuration of the camera a long time $x_{v,k}$, where k is an index corresponding to time/camera frame and x_v is a representation of the 3D configuration of the camera, i.e., an element of the Special Euclidean group. The vision algorithms reconstruct the scene up to an unknown scale factor, this indicates that the true 3D points Q_i are related to P_i through

Eq. (3) $Qi=dPi+v$

Where d is the global scale factor for the whole scene, and v is the reconstruction error noise. We aim surroundings d adopting Bayesian inference, based on object detections given by a generic object detector.

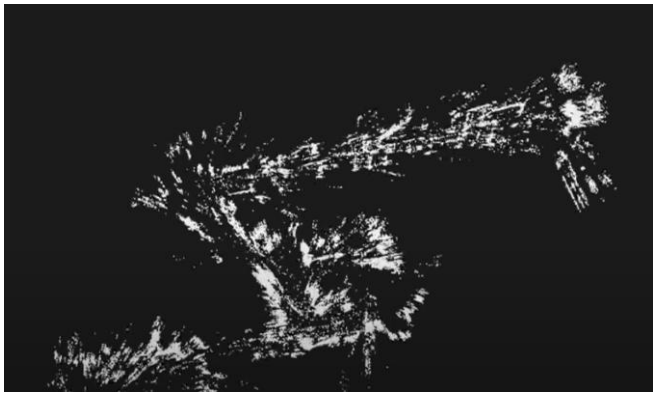

Figure 12.Output data representation of VSLAM

V. RESULT AND DISCUSSION

For the deep learning framework, we have used Tensor-flow software which helps us in building the neural network, and after training the CNN and LSTM neural network for 12 hours the training accuracy came around 96%. In the mapping part, we achieved an accuracy of 78% on average by tuning the Adaptive Monte Carlo localization package which is present in the robot operating system.

Table 1. Accuracy of image detection

S.no	Type of network	Image detection accuracy (Percentage)	Time taken for detection (Seconds)
1	CNN	95	4
2	LSTM	97	5

VI. CONCLUSION

In this research, we have seen how the navigation of a robot is done using a deep learning algorithm by replacing the expensive range finders sensor with an image sensor. The data which is acquired from the image sensor is processed by the neural network and is sent to the SLAM software where the SLAM algorithm generates a map which is eventually used for the navigation of the robot while simultaneously avoiding any obstacles .However, the efficiency of this method will be low in environment which has low lighting therefore this method can be used for robot navigation in well-lit indoor environments.

REFERENCES

1. h. huang, d. sun, r. wang, c. zhu, and b. liu, "ship target detection based on improved yolo network," mathematical problems in engineering, vol. 2020, p. 6402149, 2020/08/17 2020.
2. t. h. le, "applying artificial neural networks for face recognition," advances in artificial neural systems, vol. 2011, p. 673016, 2011/11/03 2011.
3. k. zhang, z. zhang, z. li, and y. qiao, "joint face detection and alignment using multitask cascaded convolutional networks," ieee signal processing letters, vol. 23, no. 10, pp. 1499-1503, 2016

4. jie zhou, ying cao, xuguang wang, peng li, and wei xu. deep recurrent models with fast-forward connections for neural machine translation. corr, abs/1606.04199, 2016.
5. a. graves and j. schmidhuber. framewise phoneme classification with bidirectional lstm networks. in proc. int. joint conf. on neural networks ijcnn 2005, 2005.
6. klaus greff, rupesh kumar srivastava, jan koutník, bas r. steunebrink, and jürgen schmidhuber. lstm: a search space odyssey. corr, abs/1503.04069, 2015.
7. O. Abdel-hamid, L. Deng and D. Yu, Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition, pp. 3366-3370, August 2013.
8. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going Deeper with Convolutions, 2014.
9. A. Buyval, I. Afanasyev, E. Magid. "Comparative analysis of ROS-based monocular slam methods for indoor navigation", in Proc. SPIE.
10. M. Sokolov, O. Bulichev and I. Afanasyev, "Analysis of ROS-based Visual and Lidar Odometry for a Teleoperated Crawler-type Robot in indoor environment", in Proc. Int. Conf. on Informatics in Control, Automation and Robotics (ICINCO), Madrid, Spain, 2017.
11. Magid, E., Tsubouchi, T.: Static balance for rescue robot navigation: discretizing rotationalmotion within random step environment. In: International Conference on Simulation, Model-ing, and Programming for Autonomous Robots, pp. 423–435. Springer, Berlin (2010) .
12. J. Engel, J. Stuckler and D. Cremers, "Large-scale direct slam with stereo cameras", Proc. IEEE/RSJ Intelligent Robots and Systems (IROS), pp. 1935-1942, 2015.
13. A.J. Davison, I.D. Reid, N.D. Molton and O. Stasse, "MonoSLAM: Real-time single camera SLAM", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1052-1067, 2007.
14. 14.W. Hess, D. Kohler, H. Rapp and D. Andor, "Real-Time Loop Closure in 2D LIDAR SLAM", *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 1271-1278, 2016
15. A. Buyval, I. Afanasyev and E. Magid, "Comparative analysis of ROS-based monocular slam methods for indoor navigation", *Proc. SPIE 10341 of Int. Conf. on Machine Vision (ICMV)*, 2016.

AUTHORS PROFILE

Karthik Valliapan C, is a student pursing his under graduate program in PSG College of technology in the field of Robotics and automation. His area of research focuses on Robot navigation, Vision systems and process flow automation. He has also done various projects which includes Robot collaborative material handling system which is an inventory management robot and current working on the research work of bipedal robot navigation.

Vikram R, is a student pursing his under graduate program in PSG College of technology in the field of Robotics and automation. His area of research focuses on deep learning, mobile robotics and CNC machines. He has done projects which includes 3 AXIS laser CNC machine and currently working on the project titled "Quadrupped robots".