International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Indoor Navigation of Unmanned Grounded Vehicle using CNN

Arindam Jain, Ayush Singh, Deepanshu Bansal, Madan Mohan Tripathi

Abstract—This paper presents a hardware and software architecture for an indoor navigation of unmanned ground vehicles. It discusses the complete process of taking input from the camera to steering the vehicle in a desired direction. Images taken from a single front-facing camera are taken as input. We have prepared our own dataset of the indoor environment in order to generate data for training the network. For training, the images are mapped with steering directions, those are, left, right, forward or reverse. The pre-trained convolutional neural network(CNN) model then predicts the direction to steer in. The model then gives this output direction to the microprocessor, which in turn controls the motors to transverse in that direction. With minimum amount of training data and time taken for training, very accurate results were obtained, both in the simulation as well as actual hardware testing. After training, the model itself learned to stay within the boundary of the corridor and identify any immediate obstruction which might come up. The system operates at a speed of 2 fps. For training as well as making real time predictions, MacBook Air was used.

Keywords—Unmanned Ground Vehicle, Self-Driving Vehicle, Convolutional Neural Network, kernel.

I.INTRODUCTION

One of the best prospects for Machine Learning related applications in the near future is self-driving cars. Large amounts of data can be processed nowadays which can be used for autonomous navigation purposes. But, it is imperative that the data must be labelled and contextually rich in order to be used. From a broad viewpoint of perception and controls, the technologies needed to properly solve self-driving can theoretically be applied to other fascinating activities such as video and planning action recognition.

The main sensor used by human drivers is their eye i.e. the vision itself. Similarly, cameras can be used as vision sensors in self driving cars. An economically viable solution for self-driving cars while continuing to expand the Artificial Intelligence frontier is focused on vision i.e. the camera itself. Cameras are more beneficial than other self-driving technology because even without the drivers input, traffic road signs can be read and even a variety of colors can be sensed allowing cars to navigate roads. Cameras are also a cost effective choice compared to its alternatives such as LIDAR.

The main objective of Convolutional Neural Network(CNN) is to map the characteristics of input image data to an output variable. With drastic improvement in computational power and large amounts of data available nowadays, they have proven to be tremendously useful. CNN has been long used for digital image processing and pattern recognition. This work is motivated by two primary objectives: one to develop a self-driving vehicle in minimum possible budget for hardware components. Nowadays, a lot of research is done towards the development of self-driving vehicles. In doing so, it is seen that it is imperative to use expensive equipment like LIDAR, GPU's, GPS trackers etc.. This paper discusses an approach in which only a single front facing camera is used. Secondly, the CNN model used is optimized in such a manner that it takes minimal time to train. Moreover, gives exceptionally accurate results with less amount of training data. The preliminary work for this project was initiated one year ago when a radio control(RC) car was made to self-drive in a custom made track. Similar CNN models are used in both these projects. A set of images mapped with a steering angle as well as a forward velocity are used as training data. The image was formed by combining images from two cameras which were placed adjacent to each other. The RC car was finally able to identify the boundaries and also steer itself successfully in tracks. The main application of the project is to navigate an unmanned vehicle in the corridors of the university. However, the autonomous vehicle can be used for a wide variety of other applications as well, after the dataset for any new environment is recorded and trained. It then takes minimal time for the vehicle to learn about the new environment, due to less time taken for training the CNN model. For instance, the vehicle can be used in mining operations where the path boundaries are defined. Another application can be its use in industries and warehouses for automating material movement in manufacturing units. The rest of the paper is structured as follows : section II describes the past research work done in this area, section III describes the hardware architecture used, section IV discusses the data training and collection, section V acquaints the reader with the Neural Network model used in the project, simulation and comparison of models is discussed in section VI.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Arindam Jain*, Department of Electrical Engineering, Delhi Technological University, New Delhi, India.
Ayush Singh*, Department of Electrical Engineering, Delhi TechnologicalUniversity, New Delhi, India.
Deepanshu Bansal*, Department of Electrical Engineering, Delhi Technological University, New Delhi, India.
Prof. Madan Mohan Tripathi*, Department of Electrical Engineering, Delhi Technological University, New Delhi, India.

N Finally, results and conclusions are displayed in section VII and VIII respectively.

II. RELATED WORK

Various aspects of autonomy are being explored using convolutional neural networks(CNN) such as in the case of [1]. In the research paper [4], gmapping algorithm using ROS is used to localize and navigate. Gmapping algorithm uses laser scan data from the LIDAR sensor to make the map. Developments in indoor and outdoor navigation of mobile robots during the last 20 years and different aspects of robot autonomy were studied in [2].Navigation using sensor fusion of LIDAR, camera and in case of sensor failure has been discussed in [3]. In [7] reinforcement learning techniques were used to teach mobile robots to navigate the corridor using a laser range finder. Mobile robot Navigation using 2D SLAM algorithms and single LIDAR sensor have been discussed in [8]. An algorithm for exploring and navigating in the unknown indoor environment has been explored in [10].

In [6] depth perception algorithm is used to perform 3D hand tracking and to navigate the robot by tracking different hand gestures performed by two hands. Map building using probabilistic grid models and the use of occupancy grid maps from mapping and navigation is examined in [5]. An RGB color and 3D depth imaging data was used for navigation and target tracking and a fuzzy logic controller is implemented for control of the robot in [11]. A fast pixel-based algorithm is developed in [12] for use in vision-guided mobile robot navigation using fast extraction of lines. Through grouping pixels of identical gradient orientation, it builds on an algorithm for extracting lines.

A navigation system with dynamic obstacle avoidance utilizing RGB-D camera as well as 2D LIDAR using ROS 2D navigation stack is discussed in [13]. In some cases Generative adversarial networks(GAN) have been used for helping in the autonomous navigation tasks [14].

Compared to previous works, we use a single camera as sensor for navigating the vehicle and present an economic and computationally less expensive approach.

1. Electric vehicle used for navigation purposes in corridor of Electrical Engineering Department.

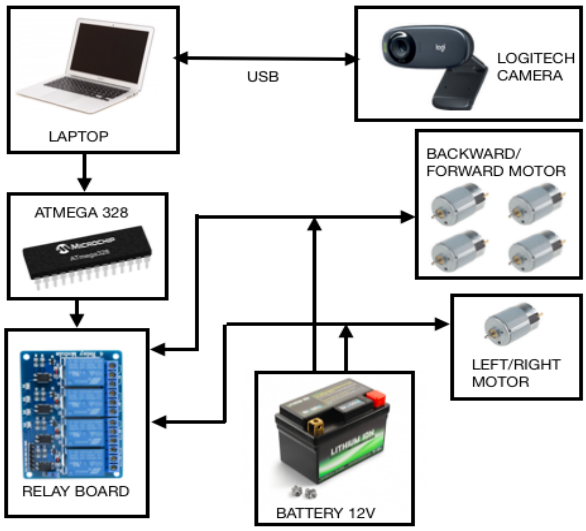
III.HARDWARE

The car used for the purpose of autonomous navigation is a four wheel drive vehicle and is shown in figure 1, the carwas first converted into Electric vehicle. For doing so we used five DC motors, one motor on each wheel and one for turning the car in the desired direction. We used a Li-ion battery of 12V 12000mah, the high capacity ensures higher run time of the vehicle. The microcontroller used is ATMega328 for controlling the wheels. A relay board having four relays is used along with microcontroller, four pins of the microcontroller are connected to four chips on the relay board, these four pins corresponds to forward, reverse, left and right directions respectively.

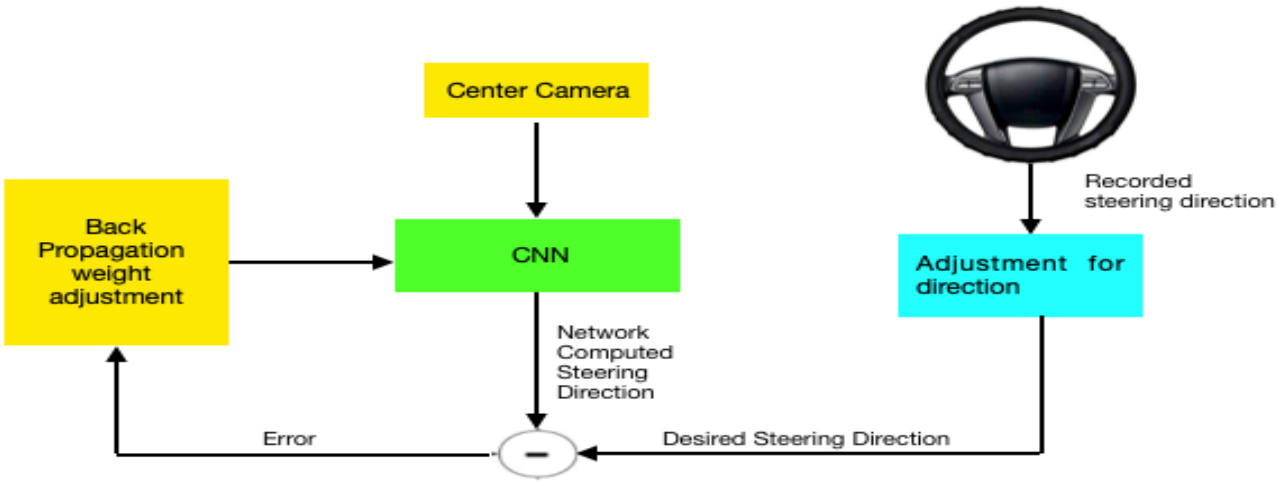
Table-I: Hardware Components Used

COMPONENT	RATINGS/SPECIFICATION
MOTOR	PMDC 12V, 1,2A
MICROCONTROLLER	ATMEGA 12V,1.2A
RELAY BOARD	4 CHANNEL 5V MODULE
BATTERY	LI-ION 12V, 12000 MAH
CAMERA	LOGITECH C310

Two of the relays are used for controlling left and right direction. In the un-energized state, the common pins of two relays are connected to the motor wires and the NC of both the relays is connected to the negative of the battery. In order to move left or right one of the relay is energized resulting in connection of one motor wire to the positive of the battery. Similarly in case of forward or reverse direction, the only difference being is that four motors are connected in series and are controlled in the same manner. The laptop act as the main processing unit for real time data from camera, training of data and implementation of algorithms in python. It processes the real time image taken by the camera and plans the instantaneous motion of the vehicle. It does this using the pretrained model explained in section V.



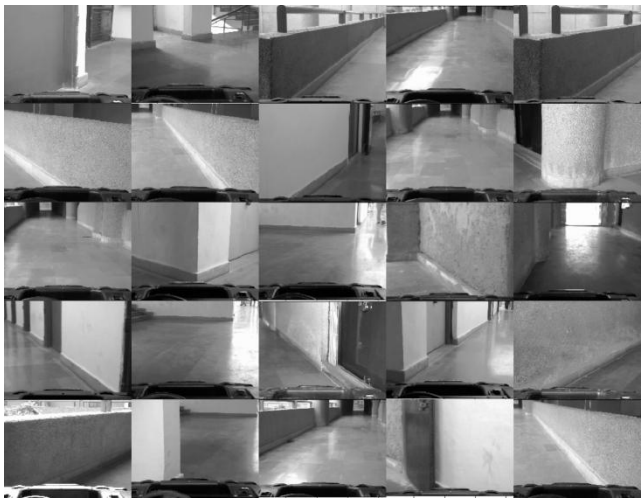
2. Hardware Architecture.



3. Training for neural network.The command was given by human input mapped with the center camera image at corresponding time in real time. After training, the network can predict the steering direction from a single center camera's video images.

IV.DATA COLLECTION

Training data were obtained by driving the vehicle under a diverse set of lighting on a number of corridors with significant variance in gradient of the road. Most of the corridor data was collected in the Electrical Department of Delhi Technological University, some of the data was also collected from college surroundings. Data(Fig .4) was collected in sunny, clear and moist weather during day and night. Due to different illumination levels throughout the day, different shadow areas were observed in the dataset. Also, reflection from the corridor surface caused a glaring effect.



4. Picture showing dataset collected in different corridors under various environment conditions i.e. sunny day and night. The camera was mounted on top for data collection.

Data was acquired using either our phone or Logitech Camera C310 mount on Electric Vehicle Car. Full attentiveness was maintained while driving the vehicle using Laptop. In order to collect the data the vehicle was controlled using Laptop, separate code for driving the vehicle using laptop keys was written. Till November 28, 2019, we had collected about 12 hours of driving data.

The following is the data collection process. The camera captures the image which is then cropped to get thebest possible feature encoded images. Now these three dimensional images i.e. coloured images are converted into single dimension i.e. grayscale images. This is now converted into a numpy array. Then the train image is paired with train label (human input). Train labels are as follows: forward, reverse, left and right. Finally, all image data and labels that are paired are stored in a npz format.

V.NETWORK ARCHITECTURE

The weights of our network were trained in order to minimize the mean squared error between label command paired with input image, which was given by the human driver and the steering command output by the network. We used Adam optimizer which is an adaptive learning rate optimization algorithm which is best suited for our particular application. The network consists of 11 layers, including 5 convolutional layers, 1 dropout layer, 1 flattening layer and 4 fully connected layers. Our network architecture is shown in Figure 5. The input to the model of training is only the region of interest(ROI) which is the filtered image of the actual captured image by the camera to get high feature extraction.

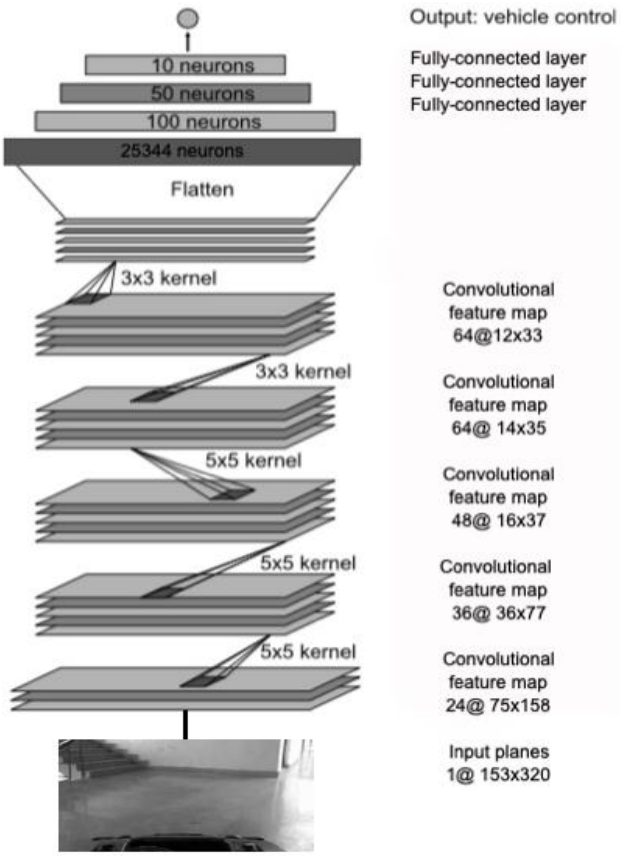
The convolutionary layers were configured for extraction of features and were empirically selected through a number of experiments with varied layer configurations. We have used strided convolutions with a 2×2 stride and a 5x5 kernel in the first three convolution layers and a non-strided convolution in the last two convolution layers with a 3x3 kernel scale.

The five convolutional layers were followed by four fully connected layers leading to the value of output command i.e. forward, reverse, left or right. The output of the fully connected layers is designed to function as a steering controller. This is an end to end learning approach to enable electrical grounded vehicles to navigate autonomously in the given environment.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 75, 158, 24)	624
conv2d_2 (Conv2D)	(None, 36, 77, 36)	21636
conv2d_3 (Conv2D)	(None, 16, 37, 48)	43248
conv2d_4 (Conv2D)	(None, 14, 35, 64)	27712
conv2d_5 (Conv2D)	(None, 12, 33, 64)	36928
dropout_1 (Dropout)	(None, 12, 33, 64)	0
flatten_1 (Flatten)	(None, 25344)	0
dense_1 (Dense)	(None, 100)	2534500
dense_2 (Dense)	(None, 50)	5050
dense_3 (Dense)	(None, 10)	510
dense_4 (Dense)	(None, 4)	44

Total params: 2,670,252
Trainable params: 2,670,252
Non-trainable params: 0

5. CNN Architecture, detail analysis with each layer. Left side explains the respective layer, the center represents the corresponding shape with dimensions which explains the transformation of image in useful features and right most side tells us about the parameters involved at each stage.



6. CNN Architecture. There are about 2 million parameters and 30 million connections. It shows how each image is transferred into useful features finally giving single output that is predicted direction.

VI. SIMULATION AND COMPARISON OF MODELS

Before running the vehicle in the actual area, we simulated the performance of various algorithms and sought out the algorithm which gave very good accuracy with minimal processing time taken.

The simulator is provided with a video of track recorded from a single camera on a human-driven vehicle. Regular snapshots from the video are taken and given as input to the machine learning model. The output of the model which is amongst forward, reverse, left or right. The input image is itself mapped to the direction in which the human transversed the vehicle. The accuracy is calculated by the following formula :

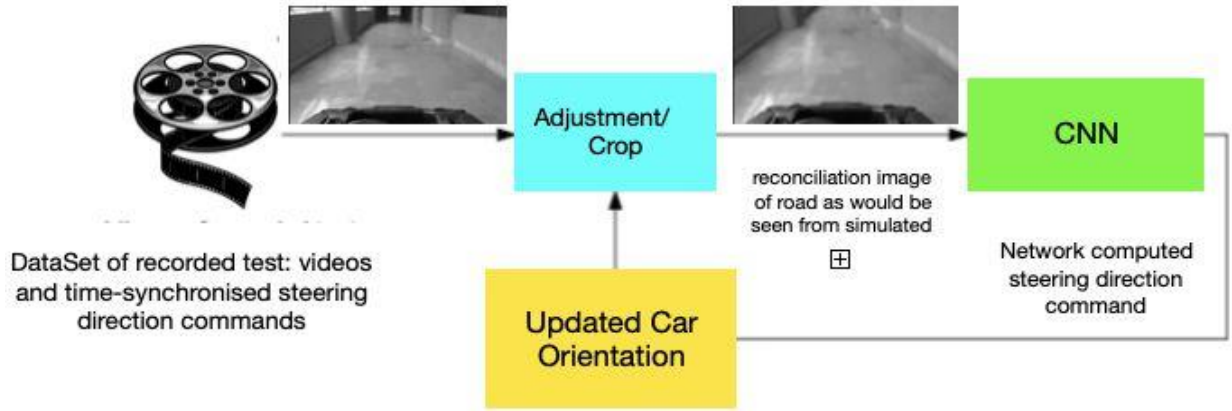
$$\text{Accuracy}(\%) = \frac{\text{No. of correct predictions}}{\text{Total no. of predictions}} \times 100$$

Where, correct predictions is that prediction in which output of the model matches with the command given by the human-driver.

Block diagram is shown in figure 7. The simulator first of all crops the top of the image to get the region of interest. Then the given image is passed through various models. The car position is updated based upon output direction command from the given model.

As shown intable 2, various machine learning models have been simulated and their accuracy displayed. This assessment was conducted using a different dataset (test data set) than the one used for training. Data set was divided into two parts: 80% training dataset and 20% testing dataset.

We compared Support Vector Machine(SVM), K-Nearest Neighbors(KNN), LSTM .In case of KNN, classification was slow and it requires more memory space. The problem that we faced while using LSTM was gradient vanishing and exploding problems as well as it was not able to process long sequences. The main disadvantage in Support Vector Machine(SVM) is that it is sensitive to outliers and the cost of competition is high. Moreover, for every dataset different kernel functions are required. SVM was still able to perform well in many cases, it was able to predict the direction command with 90% + accuracy. However, CNN works best in case of images as it can handle multiple query processing and more important is that it can share pre-trained weights, which is the basis of transfer learning. In the end we kept two choose SVM and CNN for testing. SVM was able to produce satisfactory results but the processing time to build the hyperplane requires more time and each time we have to tune the hyper-parameters as well as kernel for new data set which was quite a tedious task. In that case CNN was the best option as its accuracy was high as well as it requires processing time was comparatively less than other networks. RESNET provided the best accuracy but the major drawback of using it is that it takes weeks to train. As one of the main objectives of our project is to use a computationally cheap model, henceforth we did not use RESNET 50 as our model. Moreover, the accuracy obtained in the case of CNN was sufficient to run the vehicle safely.



7. Block Diagram of drive Simulator, used to select the best model for real time driving. The region of interest is taken and given to various model to check the best accuracy for selecting the model.

Model	Labeling Approach	Author	Accuracy
KNN	Self-proclaimed	P. Cunningham (2007)	78.6%
LSTM	Self-proclaimed	Lsu Chi (2017)	86.3%
SVM	Self-proclaimed	J.A.K Suykens (1999)	92.2%
CNN	Self-proclaimed	W. Shi (2016)	96.8%
RESNET 50	Self-proclaimed	Maqueda (2018)	98.6%

Table II: Comparison of different model tested on real time simulation. On the left side we have models that are passed into the simulation and on the right side we have accuracy for the particular model tested through simulation.

VII.EVALUATION

After the CNN model was able to provide good results in simulation, we started with hardware testing. The same model is used to predict the direction command. In it, the electric vehicle was run in the corridors of the university. The testing was done during different times of the day so as to take into consideration different lightning conditions. The car is completely autonomous, i.e. no input from human-driver is taken. The main evaluation parameter of our project is that the vehicle navigates safely, i.e. without any collision. Also, the vehicle should stop upon any sudden visualization of obstruction in the path.

The vehicle was allowed to self-drive for a total period of 1 hour. It was observed that the vehicle was able to transverse without any collision.

VIII.CONCLUSION

The CNN is able to learn meaningful road features from a very sparse training signal. While the vehicle was able to travel autonomously using CNN. We have empirically demonstrated that CNNs are able to learn and the road without manual decomposition into identification of road or lane marking, semantic abstraction, route planning, path planning and control. Even a small amount of training data was adequate to train the car to operate on corridors under various conditions.

The model CNN is able to figure out and memorize the important feature of the path without any need of hardcore typecasting own its own.

Further scope of improvement is still present. The LIDAR used for human detection and avoidance, can be used further. Currently we are working on fusion of sensors such as Lidar, IMU and camera using ROS to further automate the vehicle and to have the capabilities of dynamic obstacle avoidance.

ACKNOWLEDGEMENT

We are thankful to the Electrical Engineering Department and IOT systems lab of Delhi Technological University for giving us the permission to test the vehicle in its corridor as well as providing us with the necessary resources. Also, we are grateful to the TEQIP-III for providing us with funding for our project.

REFERENCES

1. M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," CoRR, vol. abs/1604.07316, 2016.
2. G.N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: A survey," IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 2, pp. 237–267, 2002.

3. Naman Patel, Anna Choromanska, Prashanth Krishnamurthy, Farshad Khorrami, "Sensor Modality Fusion with CNNs for UGV Autonomous Driving in Indoor Environments" in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), , Vancouver, BC, Canada, September 24–28, 2017. pp. 1531-36.
4. Emil-IoanVoisan, Bogdan Paulis, Radu-Emil Precup and Florin Dragan, "ROS-Based Robot Navigation and Human Interaction in Indoor Environment," in 10th Jubilee IEEE International Symposium on Applied Computational Intelligence and Informatics, May 21-23, 2015.
5. A. Elfes "Using Occupancy grids for mobile robot and navigation" in IEEE Journals and Magazines Computer Volume: 22, Issue: 6, June 1989.
6. Dan Xu ; Yen-Lun Chen ; Chuan Lin ; Xin Kong ; Xinyu Wu "Real time dynamic gesture recognition system based on depth perception for robot navigation" in 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO).
7. W. D. Smart and L. P. Kaelbling, "Effective reinforcement learning for mobile robots," in Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on, vol. 4. IEEE, 2002, pp. 3404–3410.
8. Yi Cheng ; Gong Ye Wang "Mobile robot navigation using LIDAR" in 2018 Chinese Control And Decision Conference (CCDC).
9. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987.
10. V. N. Murali and S. T. Birchfield, "Autonomous navigation and mapping using monocular low-resolution grayscale vision," in Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008, pp. 1–8.
11. Patrick Benavidez ; Mo Jamshidi "Mobile robot Navigation and target tracking system" in 2011 6th International Conference on System of Systems Engineering, pp. 299-304, June 27-30, 2011.
12. P. Kahn ; L. Kitchen ; E.M. Riseman "A fast line finder for vision guided robot navigation" in IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 12 , Issue: 11 , Nov 1990.
13. SukkpranhachaiGatesichapakorn ; Jun Takamatsu ; MitiRuchanurucks ROS based Autonomous Mobile Robot Navigation using 2D LiDAR and RGB-D Camera in 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)
14. A. Ghosh, B. Bhattacharya, and S. B. R. Chowdhury, "Sad-gan:Synthetic autonomous driving using generative adversarial networks,"arXiv preprint arXiv:1611.08788, 2016.
15. "Event-based Vision meets Deep Learning on Steering Prediction for Self-driving Cars", Ana I. Maqueda, Antonio Loquercio, Guillermo Gallego, Computer Vision and Pattern Recognition, 2018.

AUTHORS PROFILE

Mr. Arindam Jain, is currently pursuing Electrical Engineering from Delhi Technological University. He is interested towards building embedded hardware. He is particularly interested in Robotics, which he explored by focusing academic coursework around Machine Learning, Deep Learning, Data Analytics, Data Structures & Algorithms, Probabilistic Graphical Models, Reinforcement Learning and Robotics and Machine Vision. He have successfully won IEEE DTU hardware challenges (2017-19) and IET DTU(2019) student chapter hardware. He have research experience through Machine Learning Research Intern at Hero Electronix, where he worked on detecting crying baby sound.

Mr. Ayush Singh, is currently pursuing Electrical Engineering from Delhi Technological University. He is interested in applying Artificial Intelligence Techniques in Electrical Engineering Domain.
He completed a two months internship in the summer of 2018 at IIT-Kanpur in the field of Power System Design of VTOL. He has also presented several research papers in various national and international conferences worldwide. He has also done a part time internship in the field of Natural Language Processing in Avasant; He successfully completed his two months internship at IIT-Delhi in power system protection.

Mr. Deepanshu Bansal, is currently pursuing Electrical Engineering from Delhi Technological University, India. He has worked in the Embedded and Control Department of Autonomous Underwater Vehicles and successfully published a research paper in IJRTE based on the work. He participated in Singapore AUV challenge in 2018 and reached till finals.He successfully completed his two months internship in the field of Electric Vehicles in 2019. He is currently doing research in Navigation of Unmanned Ground Vehicle and has completed a research internship at the University of Quebec, Trois Rivieres in the fieldof autonomous vehicles.

Prof. Madan Mohan Tripathi, received B. Tech. from M. M. M. Engineering College, Gorakhpur, India and Ph.D. from G. B. Technical University, Lucknow, India in 1994 and 2010, respectively. He has worked as Engineer-SC with Institute for Plasma Research, Gandhinagar, India. Presently he is Scientist-D in DOEACC Society, New Delhi, India. His research interests include power system restructuring, ANN application to power system problems and application of IT in power system control and monitoring. He is a member of Plasma Science Society of India (PSSI) and IETE