ISSN No:-2456-2165

Navigation in Crowded Spaces Using Trajectory Prediction

Raghav Utpat
Pune Institute of Computer Technology

Sameer Joshi
Pune Institute of Computer Technology

Seema Chandak
Prof., Pune Institute of Computer Technology

Shaunak Joshi
Pune Institute of Computer Technology

Shubham Loya
Pune Institute of Computer Technology

Abstract:- Navigation in cluttered and crowded environments has been an important and difficult problem in technology. This involves accurately predicting pedestrians' movements, dynamically analysing developments in the surroundings, and adjusting the path accordingly. This paper focuses on solving the navigation problem by predicting the trajectories of pedestrians. Humans are identified and tracked using state-of-the-art object detection techniques. R-CNN and YOLO are proven to have the best accuracy and speed to perform the task. We used both social and non-social algorithms to predict trajectories of the detected pedestrians. These trajectories are used to estimate future positions of the pedestrians. Finally these positions are used to calculate the path through the environment.

Keywords:- *Navigation, Realtime, Trajectory Prediction, Deep Learning, Object Detection, Path planning, Long Short Term Memory Network (LSTM), Region based Convolutional Network (R-CNN)*

I. INTRODUCTION

Automation of menial human tasks has been a constant driving force in the history of technology. Humans' ability to navigate their surroundings is a key element of a lot of these tasks. Navigation in a crowded environment involves detecting obstacles, modelling the environment, localizing positions and avoiding obstacles. Doing all these tasks for a dynamic system with changing surroundings has been computationally very expensive and thus difficult. Improvements in object detection, high bandwidth and low latency cloud technology and increased capability of commodity hardware in the past years, now enables us to perform each of these tasks in conjunction and solve the navigation problem for pedestrian environments of medium to low density crowds. Navigation in congested environments can be set out as a three part problem - Object Detection and tracking, Pedestrian trajectory prediction and future position estimation and Path planning in obstacle-rid surroundings. This paper solves the problem by using unique approaches for each sub-statement to achieve overall optimality. Predicted trajectories of the identified pedestrians are used to calculate their next positions, which are then used to model the surroundings and plan an optimum path while avoiding all obstacles.

II. SCOPE

- Identifying pedestrians as obstacles and tracking them across the environment using visual input.
- Implementing state of the art trajectory prediction algorithm to calculate future trajectories of pedestrians with respect to their neighbours.
- Developing a client-side software to generate a map of the environment using predicted positions of all pedestrians.
- Implementing an algorithm to path-find through a cluttered map based on prior environment information, obstacle position and goal by selecting an optimal navigation strategy.
- This system can be extended to various housekeeping and hospitality applications, small scale payload delivery bots.

III. RELATED WORK

Akansha Bathija[6] discusses and proposes a state of the art approach to solve the object detection model which is more suitable for real time tracking and detection using YOLO.

YOLO, which performs object detection by means of a fixed grid regression. All bring different degrees of improvements in detection efficiency over the primary R-CNN and make object recognition more feasible in real-time and accuracy. YOLO is one of the fastest algorithms out there to detect objects. Although it is no longer the most accurate algorithm for object detection, when you need real-time detection without losing too much precision, it is a very good choice.

ISSN No:-2456-2165

Huynh Manh, Gita Alaghband[2] proposed two scene-LSTM models which are used for predicting human trajectories. The Model learns the information regarding the movement of people in the scene, this is a more detailed look in trajectory prediction and performs better in comparison. The model consisted of three parts mainly pedestrian movement, scene model and scene data filter. scene model trains the grid-cells memories in order to characterize the common trajectories in the cells which are the passed through filters to select the target trajectory. It was aimed for more social interactions not only involving humans but also other social objects.

Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet and Li Fei-Fei[1] put forward a novel approach for building a model which can account for the behavior of social objects within a large neighborhood, while predicting the target's path. The use of one LSTM model per person did not encapture the social environment in a scene. To address the limitation they proposed connecting neighboring LSTMs through using a pooling strategy.

Junwei Liang, Lu Jiang,Juan Carlos Niebles, Alexander Hauptmann and Li Fei-Fei [7] worked upon the modeling of paths and activities to benefit future path prediction and its various possible application. Accordingly the method was divided into different parts addressing the path prediction and correlating activities. The person behaviour is typically tested out using the CNNs trained which are then focused onto the LSTM encoders to predict the path the person takes. This approach paves the way for creating a applicable usage of the path planning problem by integrating it with the activity prediction.

Francesco Giuliari, Irtiza Hasan, Marco Cristani and Fabio Galasso [9] put forward a transformer model for trajectory prediction that does not make use of "social" elements of the environment. Rather by assigning individual transformer units to each pedestrian/targets to forecast the resulting trajectory frame by frame. The model is a step ahead of previous attempts since it does not combine or pool different LSTMs thereby giving a better chance at real-time usage and efficiently generalises in certain scenarios.

IV. METHODOLOGY

Object detection and tracking with YOLO and DeepSort

You only look once or YOLO is used for real time object detection. It is one of the fastest algorithms which works on 45 frames per second. Input video is passed to the system and it extracts frames and sends it to the object detector which is YOLO here. YOLO then generates bounding boxes with class ID for each object. Here we need to track objects for predicting their trajectory. So Deep SORT extension of Simple Real time Tracker is used. DeepSort tracks the objects detected using YOLO and the coordinates of objects for a particular time frame are stored in the system which will next be used for trajectory prediction. DeepSort is efficient in real time and cost for running is also low which is favourable for our system.

Model for human trajectory prediction

Until 2016,the major way to calculate trajectory prediction was using a physics and mathematics based approach. These methods were later overtaken by data driven approaches [1],[2],[7],[8] as they can train a machine learning model which can learn the complex patterns of human movement and avoid collisions.

The first model that we are planning on using is an advancement to the basic LSTM model. It takes LSTM models for all the individuals present in the scene and combines the "hidden states"(outputs) of all these LSTM models. This way it helps to take into account the interactions that are present between the humans in the scene. It works on the same principle mentioned in [1].

The second model that we are going to use is a Transformers Model.

We use a transformer network to model each individual and leverage the sequential nature of the task to predict the future motion of the pedestrians. The attention mechanism enables the transformer model to focus on multiple points in a pedestrian trajectory and identify sequence non-linearities. We use the same parameters as the original transformer paper for our implementation.

The metrics that are going to be used to evaluate these models mentioned above are:

ADE(Average Displacement Error) and FDE(Final Displacement Error).

Navigation module

A weighted graph is constructed using the predicted future positions of the targets and the shortest path through the graph is calculated using A* search algorithm.

The second method we use is a Voronoi graph. This method is able to minimize collision risks and works by dividing the environment into regions.

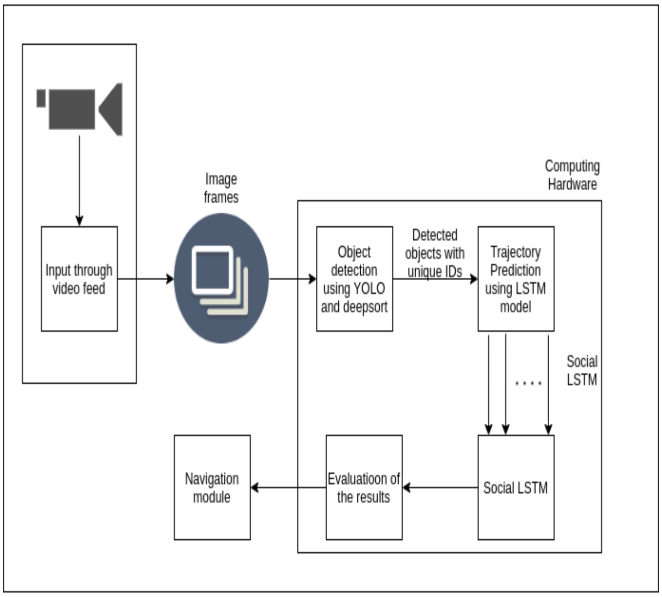

Fig. 1. System Workflow

V. CONCLUSION

The proposed set of methods are able to successfully tackle the three-part problem of real-time online navigation in crowded pedestrian surroundings. While there is still scope for future improvement of efficiency and latency of the technique, the current collection of processes can be effectively applied to a number of navigation problems ranging from a wide variety of industrial payload delivery uses to small scale hospitality and housekeeping situations.

VI. LIMITATIONS

As our system includes two Deep Learning Models, this makes it computationally heavy and expensive, thus raising issues for deployment. As a result, constant updations and efforts are in progress to try and lighten the model, thus making it easy for deployment and ready to use for any individual. Another drawback would be that our model cannot be implemented or executed on any normal machine. A specialized GPU is a must which is required due to the heaviness of the model.

REFERENCES

[1]. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Li, F.-F.; Savarese, S. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 961–971.

[2]. Huynh Manh, Gita Alaghband; Scene-LSTM: A Model for Human Trajectory Prediction; University of Colorado Denver; August 2018; arXiv:1808.04018 [cs.CV]

[3]. Hong Qu, Ke Xing, T. Alexander; An improved genetic algorithm with co-evolutionary strategy for global path planning of multiple mobile robots; Neurocomputing, Volume 120, Pages 509-517, 23 November 2013

[4]. Chen, Z.; Song, C.; Yang, Y.; Zhao, B.; Hu, Y.; Liu, S.; Zhang, J. Robot Navigation Based on Human Trajectory Prediction and Multiple Travel Modes. *Appl. Sci.* **2018**, *8*, 2205

[5]. Choset, Burdick Howei, Joel.;Sensor-based exploration: the hierarchical generalized Voronoi graph.; The International Journal of Robotics Research 19 (2) (2000), pp. 96-125; February 1, 2000

[6]. Visual Object Detection and Tracking using YOLO and SORT, International Journal Of Engineering Research &Technology (IJERT) Volume 08, Issue 11 (November 2019)

[7]. Peeking into the Future: Predicting Future Person Activities and Locations in Videos,Junwei Liang1, Lu Jiang,Juan Carlos Niebles, Alexander Hauptmann, Li Fei-Fei.

[8]. Agrim Gupta et al. "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks". In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018), pp. 2255–2264.

[9]. Transformer Networks for Trajectory Forecasting Francesco Giuliari, Irtiza Hasan, Marco Cristani, Fabio Galasso