A Method for Predicting Vehicles Motion Based on Road Scene Reconstruction and Neural Networks in Real Time

Prof. D. Sc. Nina Krapukhina, PhD student Nikolay Kamenov
The National University of Science and Technology MISIS Moscow, Russia

***Abstract*—The suggested method helps predicting vehicles movement in order to give the driver more time to react and avoid collisions on roads. The algorithm is dynamically modelling the road scene around the vehicle based on the data from the on-board camera. All moving objects are monitored and represented by the dynamic model on a 2D map. After analyzing every object's movement, the algorithm predicts its possible behavior.**

Keywords—prediction; neural networks; road scene reconstruction; real time

I. INTRODUCTION

Intensive development of big cities leads to increase in the number of vehicles and traffic flows. Road situations become complex and unpredictable, which results in higher probability of collisions and accidents. The drivers make decisions in such complex situations based on the information about the other traffic participants' position and approximate estimation of their movement dynamics. Driving a vehicle in a big city can become a challenging task, and without additional safety systems - almost impossible for an ordinary driver. The development of information technology has brought us various drivers' assistance systems [1, 2].

Unlike other robotized fixed systems where the movement happens in specific pre-determined trajectory (conveyor systems or in production), vehicles on a road have a freedom of movement in quite dense traffic flows and complex traffic situations. Therefore, the probability of a dangerous situation becomes much higher. This is the reason why safety becomes the priority in infrastructure development as well as in vehicles design. In some cases, an ordinary driver faces the challenges that can be hard to solve without vehicle's safety systems (ABC, TCS etc. [3]). To improve the driver's alertness and thus safety, a lot of systems that work in passive or active mode have been developed (Line following, Anti lock brakes, Forward collision avoidance etc.[4]). The suggested method aims at increasing the system safety level and can be used as an independent mobile application or build in a driver-assistance system.

II. PREVIOUS WORK

Developing automatic and smart vehicles has become one of the popular areas in the recent years (google car and tesla [5]). Road scene recognition and its adequate assessment, which any human performs automatically, is a serious challenge for technical systems that have to interpret the moving objects around the vehicle. A lot of manufacturers equip their vehicles with obstacle recognition systems, that warn about potentially dangerous objects. (Volvo, Mercedes, Porsche etc.)

Choice of the technical systems to provide data on the road scene is always a compromise between the price of the system's development and implementation, its safety level and efficiency. High requirements towards these systems and fast-developing technology create a field for developing a variety of algorithms to solve the issue. There are different approaches to collecting data about a road scene, depending on sensors collecting the information, methods of extracting the information, algorithms for objects detecting and identification, as well as the type of computing facilities used.

In order to improve the system's efficiency and adequate decision-making, quite often the information is collected from various types of sensors, which makes the system much more complex. The best source of information for a person is his vision, giving the biggest amount of information in a time unit. Equally important role in technical systems is devoted to a technical vision system. The development of technology and computing capacity enables us to use technical vision systems in road scene recognition. Using mono-camera for this task poses low requirements for equipment, but requires very powerful software. Based on the image the software needs to identify vertical and horizontal lines, areas, boundaries, objects, etc. [6].

Another known approach to solving this issue is using stereoscopic cameras that can provide the opportunity for 3D reconstruction of a road scene. The drawbacks include complex settings and equipment [7, 8].

Some articles suggest using additional sensors to improve the preciseness of road scene identification. These sensors provide data on vehicle's positioning (GPS, accelerometer, gyroscopes). To assure better reliability and lower calculation load, some other equipment is suggested for road scene recognition. Radars, LIDAR and other ultrasound distance measuring equipment provides very limited but highly precise information. [9, 10, 11]

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Another well-known approach to road scene recognition involves usage of neural networks. In this approach, the requirements for initial image processing are not very high, but the results of these algorithms depend a lot on the neural network structure and the methods of training. The main problem in using a wide variety of sensors usually lays in their synchronization and ensuring the stability of the system [12]. Road scene reconstruction takes part in many driver's assistance systems (ADAS).

The main goal of implementing the abovementioned systems is providing higher level of safety on the road and lowering the informational load on the driver. The main functions of the systems: identify various static and moving objects, predict obstacles and warn the driver, and in some cases – interfere in the driving process. As a standard valuation for danger level many systems use the calculated Time To Collision (TTC) with the object that is located on the vehicle's route or around the vehicle [13, 14]. The analysis of the abovementioned articles showed that the main goal there is to correctly calculate the time to collision and estimate the right course of actions to lower the impact.

The main issue in this kind of systems is that, on the one hand, the optimal choice of action requires precise information about the road scene and enough time for the driver to react, but on the other hand, it requires more time for precise reconstruction of the road scene and estimate of the danger.

III. GOAL

The aim of this work is developing methods and algorithms for detecting and predicting the development of a potentially dangerous situation. The specifics of this work is that we analyze the road scene and predict the movement dynamics on early stages, when the collision can still be avoided and the driver has more time for making a decision.

In this approach, we use one on-board camera and a series of sensors providing the information about the vehicle (speed, wheels turning angle, vehicle tilt angle) as the source of information. Based on the video flow analysis all moving objects in the camera field of view are identified and classified (vehicles, pedestrians, cyclists and others). Their positioning relative to the vehicle in consideration is calculated and displayed on a 2D virtual coordinate system, acting as a model of the road scene around the vehicle. Based on the information on the objects' trajectories, the system estimates their possible behavior and the probability of a potentially dangerous situation.

IV. ALGORITHM WORK PRINCIPLES

A. Reducing the need for calculation

In order to eliminate the problems with the mono-camera in processing large amounts of data, some sources recommend the region of interest approach (ROI) [15]. Our method suggests using dynamically changing area dependent on the vehicle's movement parameters, rather than static zones of the frame that do not always reflect the real traffic situation. Earlier we suggested a method for identifying the dynamically changing area based on the information about the vehicle's speed and wheels turning angle. The method allows us to cut down the picture processing time proportional to the dynamically changing actual area without losing the important information.

B. Detecting and classifying moving objects

For detecting moving objects on the road, the suggested method detects the changes in ansparse optical flow[16]. On the first stage, the video frames are converted into grey scale. Then on the basis of a well-known SIFT algorithm [17] special points are generated to be monitored in the next frame. The difference in the points coordinates $P_{xy}(f_n)$ and $P_{xy}(f_{n+1})$ in the two consequent frames creates the displacement vectors for every point. These vectors are used as the input data in RANSAC algorithm [18] for approximation and measuring the road scene global motion. All the points assigned to the static objects of the road scene have the vectors with very close characteristics. All the points assigned to the moving objects have the vectors that differ in direction and/or value from the static objects' vectors. The filters are applied to put the points in the groups with similar characteristics. Earlier we suggested the approach of classifying the objects based on their size and possible speed, but after additional analysis and research we found it necessary to change the criteria.

TABLE I.

Object type	Object height	Height to width ration	Max speed
Pedestrians	1m – 2.2m	1:3 – 1:1	14km/h
Cyclists	1.3m – 1.7m	1:1 – 1:2	20km/h
Motorcyclists	1.3m – 1.7m	2:1 – 1:2	150km/h
Vehicles	1.3m – 2m	4:1 – 3:4	150km/h
Buses / Trucks	2 m – 3.5m	13:1 – 1:1	100km/h
Others	Does not belong to the other categories (speed not exceeding 60km/h)		

C. Calculating location of each object

In monitoring every potential moving object our method includes the phase for measuring the distance and azimuth to every object. This phase is implemented based on the information about the vehicle's body position, known characteristics of the camera and its preliminary calibration test. Therefore our method allows identifying the vertical horizon, as well as the approximate angle of the road incline. The bounding box is created around every group of points representing a potential object. The size and coordinates of the bounding box are used for measuring the precise positioning of the moving object in relation to the vehicle.

$$D = h \tan\left[\alpha + \beta + \gamma - \tan^{-1}\left(\frac{\left|M_{pix} - \frac{H_{pix}}{2}\right|}{f\ VD}\right)\right] \quad (1)$$

Where:

h – height of the camera [m];

α – minimal vertical angle of the field of vision; [degrees]

2β – angle of the vertical coverage;

γ – vehicle tilt angle;

H_{pix}– maximal picture height [pixels];

M_{pix}– the lowest point of the bounding box [pixels];

VD – ratio of vertical size of the virtual space to the frame size;

f – distance to the virtual space [m];

D – distance to the object [m]

$$O = D \tan\theta \tag{2}$$

Where:

Θ –reference angle between the vehicle's direction and the main axis of the object.

$$S = D \frac{S_{pix}}{H_{pix}} \tag{3}$$

Where:

S_{pix}– size of the bounding box [pixels];

H_{pix} – maximal size of the image on the virtual surface [pixels];

S – actual size of the object [m].

Information about the distance and relative angle of arrival to every moving object in relation to the vehicle allows generating the coordinates for each of them and displaying the objects on a 2D matrix describing the road scene from the bird's-eye view. The information about an object's position works as feedback for its monitoring in the next frame, as its position is already known from the previous frame. The identified objects are projected and connected to the ones already existing in the virtual road model. If the object demonstrates a shift larger than the set limit m, it is considered to be the new object and is added to the virtual map.

Based on the information about the vehicle's speed and changes in the moving objects coordinates the algorithm calculates the approximate speed and direction for each of the identified objects. On every step the algorithm saves the current vector of the object and uses it later for calculating its trajectory and behavior. The system keeps saving the information about the object even in the case of temporary loss of vision or its disappearance from the frame. This way the system can keep the objects in the virtual monitoring mode (Fig1) [19], which can play significant role in the reconstruction of the road scene.

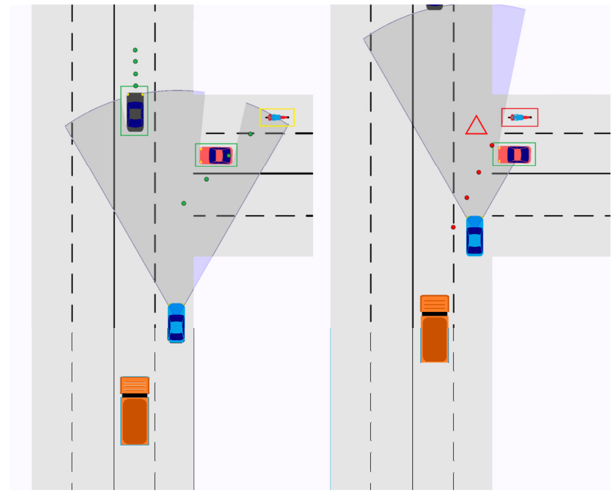

Fig. 1. Virtual monitoring mode for motorcycle. (followed even out of sign).

In most cases the trajectory curvature of other participants in the road scene, moving in the same direction as the vehicle, has almost linear character. The problem occurs when we need to predict the future movements of the objects that are moving perpendicular or do not change their position in relation to the vehicle (when both are moving in the same direction with the constant speed). In this situation it is very difficult or completely impossible to predict the movement with the algorithm of spline approximation.

D. *Usage of neural network*

For moving objects there is physical relation between the previous and the following coordinates of the object, and we suggest using this relation to predict the trajectory based on the neural networks data. A well-trained neural network can approximate any movement even in the small vicinity, which cannot be approximated by the function. We suggest using coordinates of the moving object as the input variables in the neural network, and predicted coordinates as the output. Similar works have been published, but they differ in the type of usage and mode of the object's movement. [20, 21, 22]

One of the main challenges in the neural network usage is the process of the training and obtaining the data necessary for it. The advantage of using neural networks in this case is that the information for the network training can be obtained from the current video flow. The training is based on the actual data. The input data are the known coordinates for the objects' movement trajectory extracted from the earlier time units. The obtained data are compared against the actual coordinates of the moving object in the next time unit and are used for neural network training.

Let's form several assumptions for solving the problem and minimizing the complexity. Let the objects be moving in a 2D space of x and y axis by a determined but unknown to us trajectory. In equal time units the position of the moving object in relation to the vehicle (x_o, y_o) is registered by the model of the road scene. The objects of the model represent real moving objects in the physical space that have actual physical characteristics (speed, acceleration, physical coordinates).

  ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

For us to predict the object's movement we need the data on its position, speed, direction and acceleration. Since the inertial characteristics of the object's movement are implicitly presented in its trajectory, we suggest building a neural network based on the object's coordinates only. This will cut down the volume of calculations and time for predicting the coordinates.

The chosen neural network has multi-layered perceptron structure. The input information illustrates 6 positions of the object Oi in the virtual space in 5 time units

$$P_{Oi}(t)(X_{Oi}(t), Y_{Oi}(t)) \qquad (4)$$

where t ϵ [-5, 0]. The structure has 2 concealed layers – the first one has 24 nodes and the second one has 48. The output consists of 30 nodes.

The number of inputs and outputs, as well as the time slot for receiving the result, play important role in training and getting the correct predictions. There are optimal ratios between the prediction preciseness and the time slot used as data foundation for predicting.

The data about the moving object's coordinates are normalized on the neural network input, so the peak value exceeds the activation function maximum. In our model we use bipolar sigmoid activation function [23] in the range of (-1.0; 1.0). For normalized input signals every actual coordinates value is divided by coefficient p, and the result of the neural network is multiplied by the same coefficient. Since the maximum distance for prediction is taken as 99 meters, the coefficient for coordinates values normalization in the range of (-1; 1) is taken as p = 100. Then all the values for the object's position to the left of the vehicle fall in the range of (-1; 0).

All trajectories are relative to the host vehicle

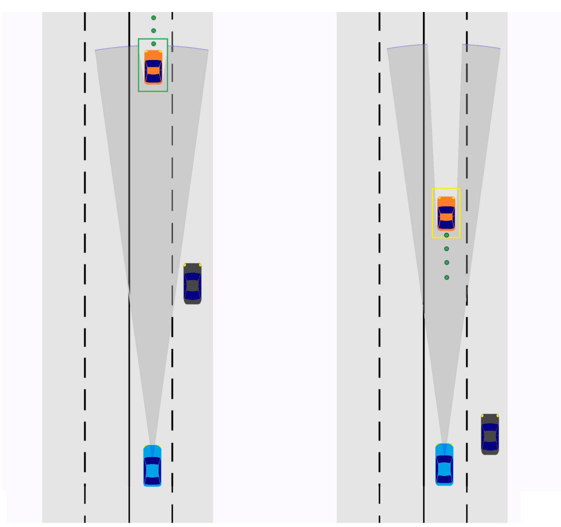

Fig. 2. Prediction of possible danger at high speed. Car in front is slowing down.

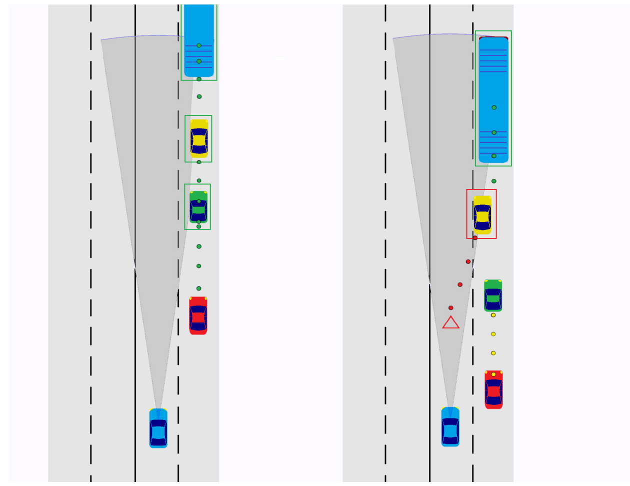

Fig. 3. Prediction of dangerous situation at high speeds. Danger overtaking. Green car is in virtual monitoring mode.

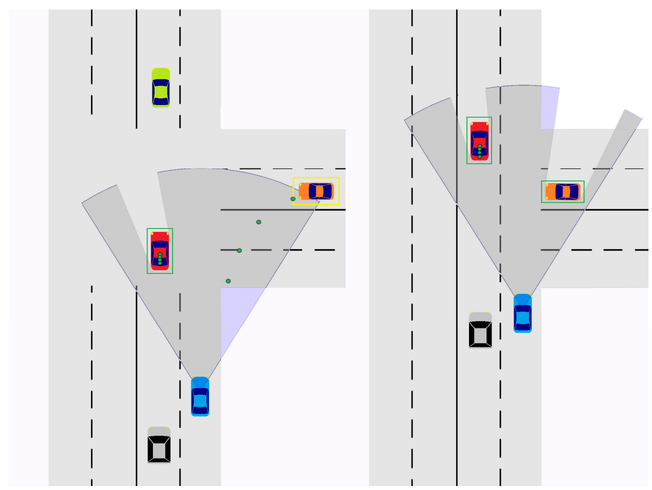

Fig. 4. Prediction of possible dangerous situation at medium speeds. Car from the right is closing.

E. Classification of possible dangerous situations

The suggested approach classifies potential situations into 3 groups for each of the detected objects.

- No danger – the moving object is far enough from the vehicle's trajectory and its predicted behavior poses no threat.
- Attention, possible danger – the object is far enough from the vehicle's trajectory, but its predicted movement shows dangerous approach in the vicinity of the vehicle. (The safety zone radius can be adjusted depending on the driver's preferences and the vehicle type).
- Danger – the object is in the way of the vehicle and/or their trajectories cross. The direction or speed of the vehicle needs to be adjusted to avoid a dangerous situation.

  ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

V. RESULTS

Data showing algorithm accuracy and speed is describedin tables below. Images from software visual interface are shown after the tables. At this stage of development algorithm does not differentiate road lines.

TABLE II.

Time of algorithm's work	
Receiving the image	18ms
Identifying the region of interest (ROI) (times per second)	1ms
Generating special points (if necessary)	9ms
Calculating the video flow (average)	48ms
Processing the data	9ms
Identifying the parameters for the moving objects and updating the road scene model	2ms
Predicting the objects' movement based on the neural network and evaluating the possibility for dangerous crossing of trajectories	3ms
Checking the preciseness of the prediction from the previous time slot (every 3 frames)	1ms
Total time for the algorithm cycle - 80 ms (91 ms if necessary)	

TABLE III.

Characteristics of the algorithm	
Detecting moving objects within ROI	96.753%
Preciseness of detecting for every object	87.135%
Preciseness of the object's parameters calculation	
Up to 10 m	98.689%
Up to 20 m	96.934%
Up to 40 m	93.225%
Up to 60 m	86.160%
Up to 100 m	80.250%
Preciseness of the neural network prediction	97.865%

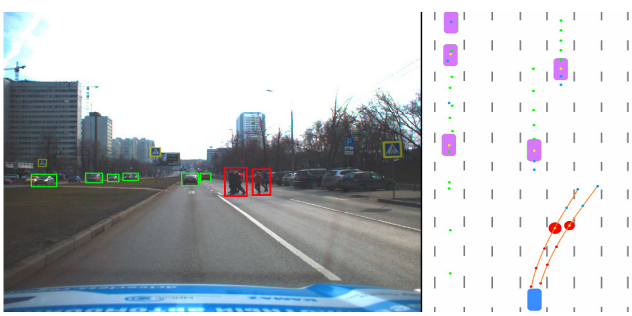

Fig. 5. Prediction of accident if object or vehicle in consideration doesn't change speed or trajectory.

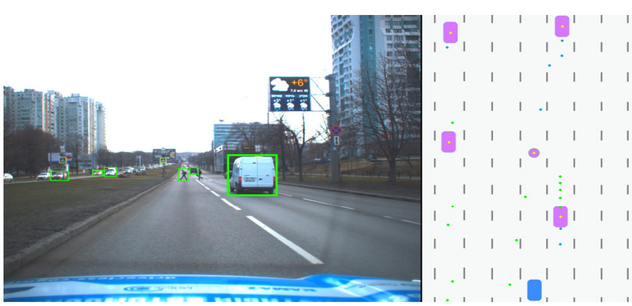

Fig. 6. Prediction for clear situation, if pedestrians continue with same speed on same trajectory.

VI. CONCLUSION

Thanks to using dynamically changing region of interest for processing the image on objects identification stage, as well as multi-layered neural network for predicting the trajectories of the various objects using just the objects' coordinates, the volume of calculations and time for trajectory prediction has been cut down significantly. This allows extra time for decision-making to the driver or the driver's assistance system (DAS).

The suggested algorithm has the following functions for predicting dangerous situations:

- Detects moving objects around the vehicle and dynamically models the road scene.
- All moving objects are classified based on the processed video flow, their calculated size and approximate speed.
- Neural network uses the accumulated data about the identified moving objects' coordinates for predicting the trajectories crossing and potential collision.

REFERENCES

[1] Bo Wahlberg "Automated Transport Systems (ATS)" KTH Royal Institute of Technology, Sweden, 2000

[2] Gereon Meyer, Jadranka Dokic, Beate Müller " European Roadmap Smart Systems for Automated Driving", The European Technology Platform on Smart Systems Integration, Berlin, April, 2015

[3] David Burton, Amanda Delaney, Stuart Newstead, David Logan, Brian Fildes "Evaluation of Anti-lock Braking Systems Effectiveness"; Royal Club of Victoria (RACV) Ltd; April 2004, ISBN 1 875963 39 1

[4] Batavia, P.H. "Driver-adaptive lane departure warning systems" CMU-RI-TR-99-25, Thesis 1999

[5] Neal E. Boudette "Building a Road Map for the Self-Driving Car" The New York Times, March 2017

[6] Diana Gornea, Dan Popescu, Grigore Stamatescu, Radu Fratila „Mono-camera robotic system for tracking moving objects" 9th IEEE Conference on Industrial Electronics and Applications 2014, pp:1820 – 1825

[7] N. Molton, S. Se, J. M. Brady, D. Lee, and P. Probert, "A stereo vision-based aid for the visually impaired," Image Vis. Comput., vol. 16, Mar. 1998, pp. 251 – 263

[8] Zhencheng Hu, and Keiichi Uchimura „U-V-Disparity: An efficient algorithm for Stereovision Based Scene Analysis" Proceedings of Intelligent Vehicles Symposium IEEE June 2005 pp: 48 - 54

[9] Mohammad Sohrab Hossan Monsi, "Laser Radar for Precise Vehicle Velocity Measurement," Kasel University, Kasel, PhD Thesis 978-3-89958-736-4, 2009.

[10] Garcia F., Cerri P., Broggi A., Armingol J.M., de la Escalera A. "Vehicle Detection Based on Laser Radar" Lecture Notes in Computer Science, vol 5717. Springer, Berlin 2009

[11] Daniel Gohring, Miao Wang, Michael Schnurmacher, Tinosch Ganjineh "Radar/Lidar sensor fusion for car-following on highways" 5th International Conference on Automation, Robotics and Applications (ICARA), 2011 pp: 407 – 412

[12] Nico Kaempchen, Klaus Dietmayer, "Data synchronization strategies for multi-sensor fusion", University of Ulm, Department of Measurement, Control and Microtechnology Albert-Einstein-Allee 41, D-89081 Ulm, Germany, Nov 2015

[13] Richard van der Horst & Jeroen Hogema, "Time to-collision and collision aboidance systems", Institute for Road Safety Research, SWOV,pp: 59 – 66, 1997 ISBN: 90-6807-293-5

[14] Kusano, K. and Gabler, H., "Method for Estimating Time to Collision at Braking in Real-World, Lead Vehicle Stopped Rear-End Crashes for

  ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Use in Pre-Crash System Design," SAE Int. J. Passeng. Cars – Mech. Syst. 4(1):435-443, doi:10.4271/2011-01-0576. 2011

[15] P. Kapsalas ; K. Rapantzikos ; A. Sofou ; Y. Avrithis – “Regions of interest for accurate object detection” Content-Based Multimedia Indexing CBMI, pp:147 – 154 2008

[16] Philip Lenz, Julius Ziegler, Andreas Geiger, Martin Roser “Sparse scene flow segmentation for moving object detection in urban environments” 4th Intelligent Vehicles Symposium 2011 pp: 926 - 932

[17] U.S. Patent 6,711,293, "Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image", David Lowe's patent for the SIFT algorithm, March 23, 2004

[18] M. A. Fischler, R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, Vol 24,1981, pp 381-395

[19] Senchenko R.V., Kamenov N.V., Krapukhina N.V., “The approach of transport security using multi-agent simulation and machine vision methods,”, In Proceedings of the XXIV International Conference Problems Of Safety Management Of Complex Systems, Moscow Russia 2016, pp: 329 – 332

[20] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, Silvio Savarese; “Social LSTM: Human Trajectory Prediction in Crowded Spaces”, “The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 961-971

[21] Anurag Sharma, Ashish Chaturvedi “Gradient Descent Feed Forward Neural Networks for Forecasting the Trajectories” International Journal of Advanced Science and Technology Vol. 34, 2011

[22] Pierre Payeur, Hoang Le-Huy, Clement M. Gosslin “Trajectory Prediction for moving objects using artifical neural networks” IEEE Transactions on industrial electronics, vol. 42, 1995 pp: 147 – 158

[23] B. Kosko, Neural Networks and Fuzzy Systems. Englewood Cliffs, NJ: Prentice-HALL, 1992

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731