2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Traffic Scene Prediction via Deep Learning: Introduction of Multi-Channel Occupancy Grid Map as a Scene Representation

Hyeong-Seok Jeon[1], Dong-Suk Kum[2], and Woo-Yeol Jeong[3], *Member, IEEE*

***Abstract*—When predicting future motions of surrounding vehicles for autonomous vehicles, the inter-vehicular interaction must be considered in order to predict future risks and to make safe and intelligent decisions. This becomes critical when it comes to conflicting driving situation such as lane merge, tollgate area, and unsignalized intersections. Previously developed future prediction algorithms show limited performance when handling interactions and conflicts between vehicles because they focused on predicting individual vehicle motion and/or interaction between a single pair of vehicles rather than the entire traffic scene. In this paper, a scene representation method, namely multi-channel Occupancy Grid Map (OGM), is proposed to describe the entire traffic scene, which is then utilized for the deep learning architecture that predicts the future traffic scene or OGM. Multi-channel OGM represents entire traffic scene as a manner of image-like structure from bird's eye view composed with dynamic layer and static layer depicting the occupancy of the dynamic and static objects. By using this 2D traffic scene representation, future prediction can be modeled as a video processing problem, where future time-serial image sequence need to be predicted. In order to predict future traffic scenes based on past traffic scenes, a deep learning architecture is proposed using Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Networks. With the proposed deep learning architecture, future prediction accuracy in highly conflicting traffic situation is guaranteed up to 90 percent with 3 seconds of prediction horizon. A video of the traffic scene prediction results is available online [1].**

I. Introduction

One of the main driving force for the development of the autonomous vehicle is safety; eliminating the accidents caused by human factor. Autonomous vehicles, however, will lead to other types of collisions due to lower level of intelligence than human. In order to reduce these accidents caused by autonomous vehicles, safe and intelligent driving strategy is essential. In particular, the future prediction of the surrounding vehicles is the key to determine safe and reliable maneuver. Most of the time, future motions of surrounding vehicles can be estimated by using simple kinematic models or interacting multiple models [2] [3]. However, during more complex driving situations when intentions of multiple vehicles conflict, the interactions between two or more vehicles becomes very important but very challenging when predicting future motions of these multiple vehicles in real-time.

There are two characteristics in inter-vehicular interactions which increase the complexity of the problem: simultaneously occurring multiple interaction pairs and sequentially varying interaction characteristics according to the each vehicle's driving strategy. Inter-vehicular interaction simultaneously spreads out via conflicting vehicles under multiple vehicles. Additionally, interaction features are continuously changing according to the future driving strategies of the each vehicles. In order to handle these two properties of the inter-vehicular interaction, deep learning based future prediction technologies have been recently proposed. By introducing deep learning based approach, interaction could be partially considered in [4] and [5]. Bayesian network based future prediction in [4] and LSTM based future prediction in [5] could handle single pair of interaction between two vehicles using relative position. However, simultaneously occurring multiple interaction pairs could not be considered because the relative state between only two vehicles is used. Other deep learning approaches were also introduced such as convolutional neural network [6] [7], recurrent neural network [7] [8], and fully connected neural network [9]. The most advanced future prediction technology in terms of interaction handling was [9]. In [9], fully connected neural network is proposed considering multiple vehicles. By doing so, simultaneousness of the interaction pairs could be handled. Even if multiple interaction pairs can be considered, however, the size of the architecture depends on complexity of the traffic scene which demands high computational power in conflicting situation. Additionally, consideration on sequentially changing interaction due to future driving strategy of the each vehicles was out of scope because fully connected neural network is not capable of time dependent feature. As shown in above, previously developed prediction methods could not fully consider time-varying simultaneous multiple interaction pairs because the existing traffic representation method is limited on the individual vehicles.

Therefore, in this paper, we propose the traffic scene representation method, multi-channel OGM, enabling interactive future prediction considering sequentially varying multiple interaction pairs among the multiple surrounding vehicles. With the multi-channel OGM, traffic scene is reformed into image-like snapshot from bird's eye view

*Research supported by Hyundai Motor Group in Republic of Korea under a Future Technology Research Project 2017. The authors would like to thank all partners in the Commercial Vehicle Advanced Electronics Development team of Hyundai Motor Group for supporting this work

1) H. S. Jeon is with the Cho Chun Shik Graduate School of Green Transportation, KAIST, 34141 Daejeon, Republic of Korea (e-mail: 521hsgugu@kaist.ac.kr).

2) D. S. Kum is with the Cho Chun Shik Graduate School of Green Transportation, KAIST, 34141 Daejeon, Republic of Korea (corresponding author e-mail: dskum@kaist.ac.kr).

3) W. Y. Jeong CV Advanced Electronics, Hyundai Motor Group, 18280 Hwaseong, Republic of Korea (e-mail: wooyeol.jeong@hyundai.com).

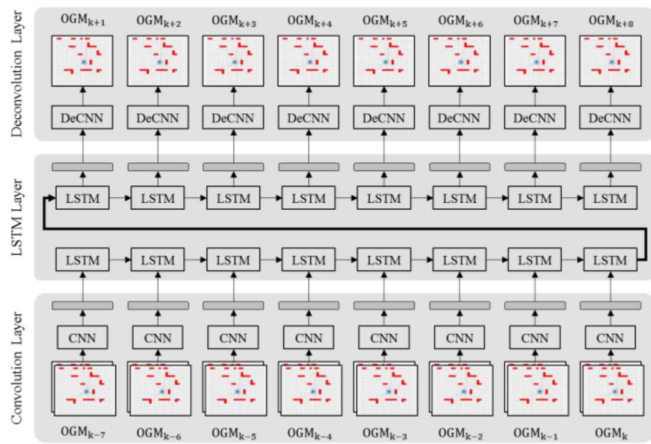

Figure 1. The proposed deep learning architecture for interactive future prediction using multi-channel occupancy grid map frame work

implicitly containing characteristics of the multiple interaction pairs. With this scene representation framework, video processing approach can be applicable in future prediction algorithm in order to interpret time-serial image sequence. In this paper, deep learning architecture with CNN and LSTM is developed as shown in fig 1. CNN is applied in order to extract implicit characteristics of multiple interaction pairs from multi-channel OGM structure. Using this interaction feature, finally, LSTM network based time-serial future prediction is conducted. LSTM network generate feature map of future traffic scene which will be decoded into future OGM structure.

The remainder of this paper is organized as follows. Section II describes the representation method of traffic scene via multi-channel OGM. Section III presents the deep learning architecture for interactive future prediction for autonomous driving vehicle. Detailed training procedure of the proposed deep learning architecture is stated in section IV. Trained deep learning architecture is simulated with traffic scenario whose interaction feature is emphasized in section V. Finally, conclusion is discussed in Section VI.

II. Traffic Scene Representation Method: Multi-Channel Occupancy Grid Map

Prior to the development of a deep learning architecture for future prediction, its inputs and outputs must be defined first. In order to handle multiple inter-vehicular interaction pairs, signals that can describe the entire traffic scene around the ego-vehicle would be ideal for inputs and outputs. In this paper, multi-channel occupancy grid map (OGM) is proposed as a traffic scene representation, which will be used as a basis for inputs and outputs of the deep learning algorithm. One of the main advantages of the multi-channel OGM is that it can efficiently describe the entire traffic scene, which can inherently describe all the interactions between traffic participants. Another key feature of the OGM is its suitability for the deep learning architecture because it is essentially an image, taken from bird's eye view, the deep learning techniques of which have recently been extensively studied and advanced in computer vision.

The concept of OGM is introduced in mobile robotics area mainly for mapping environment and static obstacles [10] [11]. In vehicle application, however, conventional OGM has not been widely utilized because the maneuver of the roadway vehicles are highly dependent on the road infrastructure. In the proposed multi-channel OGM, dependency between future maneuver of dynamic object and road information can be interpreted as well by dividing dynamic layer and static layer.

The multi-channel OGM with 1,000 grid cells, 50 rows in longitudinal direction and 20 columns in lateral direction as shown in fig 2, is virtually attached to the ego vehicle under alignment condition between the center of OGM and the center of ego vehicle to use relative position of each surrounding vehicles. Each grid cell has 2 meters of length and 0.54 meters of width.

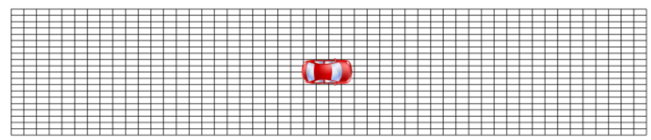

Figure 2. Frame structure of the multi-channel occupancy grid map whose center is aligned with the ego vehicle position.

The position of the dynamic object and information of the road infrastructure is included in different layer, dynamic layer and static layer respectively. After constructing each layer, finally, two layer is stacked over following depth direction as fig 3 in order to build multi-channel OGM which has image-like structure. Traffic scene representation with the suggested multi-channel OGM enables future prediction considering simultaneous interaction and properties of roadway traffic participants.

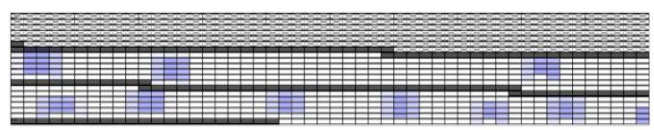

Figure 3. Sample of the multi-channel OGM structure consist with dual channel, dynamic layer and static layer.

A. Dynamic Layer

In dynamic layer, first channel of the multi-channel OGM structure, the position of the all of the surrounding vehicles is stored. Since the designated resolution of the grid cell is not sufficient to exquisitely describe the vehicle position, the value of each grid cell is determined based on occupancy ratio between 0 and 1 as shown in Eq.(1).

$$\alpha_{i,j} = \sum_{k=1}^{n} \frac{A(k|i,j)}{A(i,j)} \quad \begin{matrix} (0 \le i < 50) \\ (0 \le j < 20) \end{matrix} \tag{1}$$

where i and j are the index of each grid cell. Occupied area of (i, j)-th grid cell by k-th vehicle and is denoted as $A(k|i,j)$. By adding up occupied area divided by the area of each grid cell for all of the n surrounding vehicles, occupancy ratio of each grid cell is calculated as shown in fig 4.

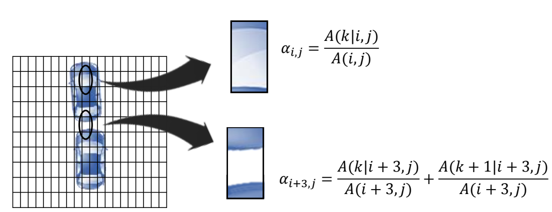

Figure 4. Occupancy ratio calculation under existence of multiple surrounding vehicles.

In fig 5, sample of the dynamic layer of the multi-channel OGM is shown. By applying the occupancy ratio to calculate the value of each grid cell, insufficient degree of precision of representation method due to sparsely divided OGM can be overcome while the computational power is maintained.

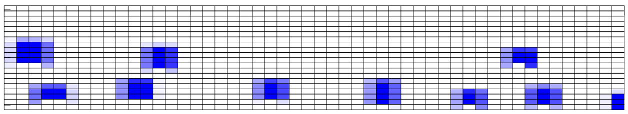

Figure 5. Dynamic layer of multi-channel occupancy grid map with occupacny ratio under multi-surrounding vehicles.

B. Static Layer

In traffic environment, maneuver of the vehicles including the ego vehicle and the surrounding vehicles is guided by the infrastructure information such as lane marking. Therefore, dependency properties between dynamic object and the road infrastructure have to be included in scene representation framework. In the proposed multi-channel OGM based scene representation method, static layer channel is introduced as a second channel of the multi-channel OGM. By introducing static layer, dynamic maneuver of the objects can be constrained based on static infrastructural information.

In static layer channel, position of the all of infrastructure can be included such as lane marking, undrivable region and drivable region without concerning about dynamic obstacles. To determine the value of each grid cell of static layer, deterministic rule-based approach is selected as shown in table 1 with visualization rule for sample of the static layer in fig 6. Road infrastructural data is converted into deterministic integer form in order to improve the performance of prediction algorithm in various situation by considering relationship between vehicle maneuver and road infrastructural information.

TABLE I

DECISION RULE FOR THE VALUE OF GRID CELL IN STATIC LAYER

Infrastructure	Integer Value	Visualization rule
Drivable Region	0	□
Lane Marking	1	■
Undrivable Region	5	⊠

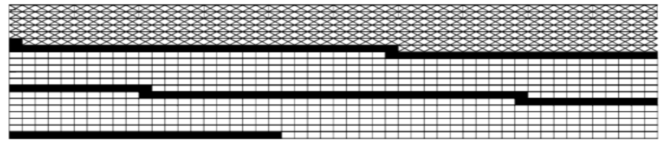

Figure 6. Static layer of multi-channel occupancy grid map with rule-based road infrastructure representation

III. DEEP LEARNING BASED FUTURE TRAFFIC PREDICTION

By introducing multi-channel OGM in section II, complex multiple interaction pairs are included in one snapshot. To interpret the multiple interaction pair information in multi-channel OGM, deep learning architecture is developed as shown in fig 7. The first part of the proposed deep learning architecture is based on convolutional neural networks (CNN) which can extract implicitly represented interaction feature from image-like input structure, multi-channel OGM. CNN is commonly used in computer vision area in order to classify or detect objects from camera images. This network is essential for interpreting the simultaneous multiple interaction pairs among vehicles. Secondly, characteristics of the future traffic scene is predicted via long short-term memory (LSTM) network using interaction feature, extracted by CNN. LSTM network is widely used in video processing in order to analyze the time serial future video, sequential structure of the image [12]. Because of the capavility of the LSTM network in handling of the time serial data structure, LSTM network is used in the proposed deep learning architecture. By using multi-channel OGM, traffic scene can be modeled as time serial image-like structure whose structure is similar to video. By applying LSTM network to generate time serial future traffic feature, sequentially varying inter-vehicular interaction pairs can be handled. Finally, deconvolution operation is applied to decode predicted traffic feature into the future OGM.

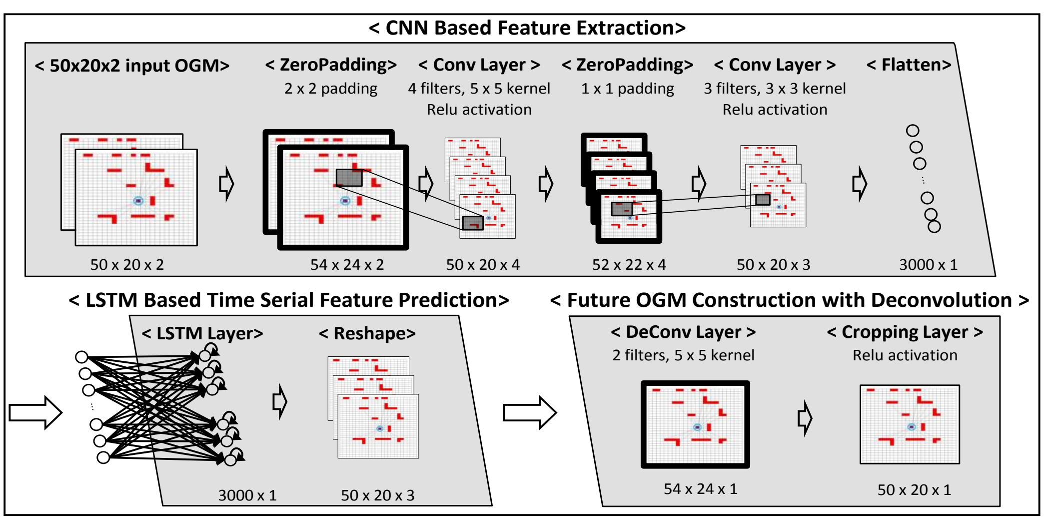

Figure 7. Overall architecture of the multi-vehicle future prediction deep learning structure using multi-channel OGM based on convolutional neural network and long short-term memory network. CNN can extract simultaneous interaction feature and relationship between vehicle maneuver and road information from multi-channel OGM. In LSTM network, time-serial prediction is available considering time dependency of interaction feature.

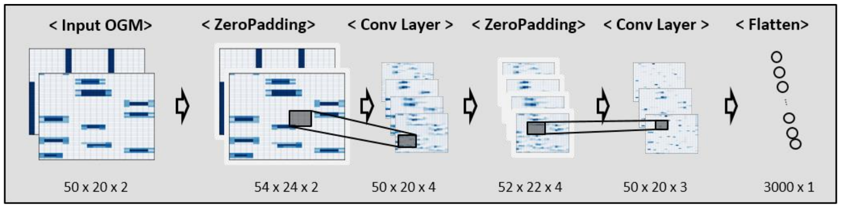

Figure 8. Interaction feature extraction via CNN from multi-channel OGM who contains inter-vehicular interaction feature implicitly.

A. *Interaction Feature Extraction using CNN*

In convolution layer, features of simultaneously occurring multiple interaction pairs among the surrounding vehicles are encoded into feature map. By introducing static layer as a second layer of multi-channel OGM which is convoluted with dynamic layer as well, characteristics of maneuver of the dynamic object constrained by static infrastructure can be encoded as well. The output of convolutional neural network, feature matrix, is converted to feature vector by flattening in order to be shaped as an input structure of the LSTM networks for time-serial future prediction which is capable of handling sequentially varying inter-vehicular interaction.

Overall convolution neural network architecture for multiple interaction feature extraction from multi-channel OGM is composed of two convolution layers and two zero-padding layers as shown in fig 8. First convolution layer composed with 5 by 5 sized 4 filters is designated as the second layer of convolutional neural network following the first zero-padding layer. This process is repeated twice in order to consider gradually varying interaction strength according to the distance from each vehicle. In the first convolution operation, feature of the interaction pairs between nearby objects are extracted. In the second convolution, feature of weaker interaction pairs between distant objects are extracted. As a result of the second convolution layer, 50 by 20 by 3 sized feature map is generated. Afterward, feature map is converted to feature vector via flatten layer.

B. *Time-serial Future Traffic Prediction using LSTM*

Long Short-Term Memory (LSTM) network is one of the Recurrent Neural Network (RNN) which can handle the time serial data structure by calculating the output at time step k+1 using the input at k+1 and the output at previous time step k. In this conventional RNN architecture, however, long term dependencies are not able to be trained because of the vanishing gradient problem [13]. To overcome vanishing gradient problem, advanced RNN structure is introduced such as Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) networks [14].

In the proposed deep learning architecture, LSTM network is introduced in order to consider sequentially varying feature of the interaction pairs due to agent's driving strategy. By suggesting multi-channel OGM whose structure is close to video, time serial image-like structure, LSTM based video processing frame work can be applied in future prediction for traffic situation. Encoder-Decoder LSTM architecture is applied in the proposed future prediction algorithm as shown in fig 9 because time-serial future traffic scene have to be predicted using time-serial historical multi-channel OGM.

In LSTM layers, feature vector of future scene is predicted up to 4 seconds with 0.5 seconds time difference between each time steps from 8 time steps of historical feature vector with 0.5 seconds time difference including current traffic state.

C. *Future OGM Generation via Deconvolution Layer*

At the end of the deep learning architecture for interactive time-serial future prediction using multi-channel OGM, deconvolution operation is applied after LSTM networks. The aim of the deconvolution operation is to generate future traffic OGM from feature map of future traffic scene, decoding process.

IV. Training of The Deep Learning Architecture

In order to extract the characteristics of implicitly included multiple interaction pairs in multi-channel OGM, deep learning architecture is developed in section III. In this section, training process of the proposed deep learning architecture using NGSIM database [15] is introduced with ADAM optimizer [16]. Throughout the training process, Keras library is used with Tensorflow backend [17] [18].

A. *NGSIM trajectory database*

The traffic flow data in NGSIM database was collected via several video cameras mounted on the top of the towers. By applying image processing algorithms on to recorded traffic flow video, the trajectory database with 18 variables related to vehicle state including positions and dimensions which are essential variables to build multi-channel OGM is constructed. The trajectory data on interstate 80 (I-80) in San Francisco bay, California consists of a six-lane road is utilized to train the proposed deep learning architecture among four data sets. I-80 trajectory dataset is collected during 30 minutes with 10 FPS. Because of the image-processing error, the cubic smoothing spline method is employed in [19]. This smoothing method generates low fluctuating trajectory data from the raw trajectory data provided which is used in training process.

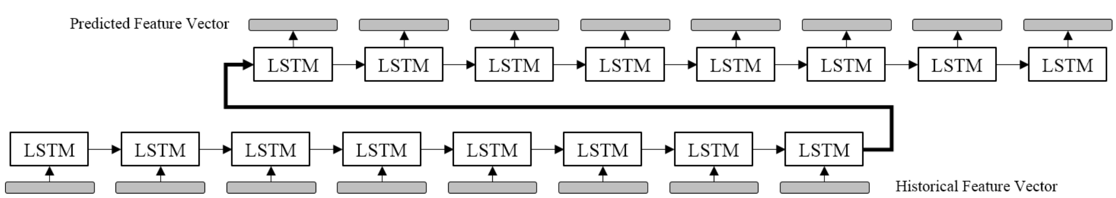

Figure 9. LSTM based time-serial future prediction using historical feature vector extracted from CNN structure.

B. Network training with ADAM optimizer

Multi-channel OGM dataset generated using NGSIM trajectory data is divided into two groups: training data set, test data set. 20% of overall multi-channel OGM dataset, composed with 45,621 data, is designated as the test set which is used for validation of trained deep learning architecture. In order to avoid overfitting without dropout, additionally, two batch normalization layers are placed after conv2 layer and conv4 layer. By introducing batch normalization to deep learning architecture large learning rate can be selected which leads faster training process and increased performance of network [20]. Overall configuration of the proposed deep learning network including two batch normalization layers is shown in table II.

With this 11 layered deep learning architecture, training process is done under 10 epochs with pre-determined batch size and early stopping function regarding to validation loss which is calculated using loss function as shown in Eq.(2).

$$J(\theta) = mean(|y^{pred}(\theta) - y^{target}(\theta)|) \quad (2)$$

where y_{pred} and y_{target} represent the value of predicted time-serial OGM and target OGM, ground truth of prediction output.

TABLE II
DEEP LEARNING ARCHITECTURE WITH CNN AND LSTM NETWORK FOR INTERACTIVE FUTURE PREDICTION USING MULTI-CHANNEL OGM

Layer index	Layer name	Activation function
Conv1	Zero-padding layer	-
Conv2	Convolution layer	ReLU
BN1	Batch Normalization layer	-
Conv3	Zero-padding layer	-
Conv4	Convolution layer	ReLU
BN2	Batch Normalization layer	-
Conv5	Flatten layer	-
LSTM1	LSTM layer	-
LSTM2	Reshape layer	-
Deconv1	Deconvolution layer	-
Deconv2	Cropping layer	ReLU

V. PREDICTION PERFORMANCE EVALUATION

Defining performance measure for the prediction model is not trivial due to the newly adopted scene representation, OGM. In this study, mean absolute percentage error, which is extended from the loss function in Eq. (2), is selected as a performance metric. Accuracy of each time step k is decided by average value of the difference between predicted value and ground truth value over every grid cell as shown in Eq.(3).

$$Accuracy_k(\%) = \frac{\sum_{j=1}^{M_j}\sum_{i=1}^{M_i}\left[\left(1-\left|y_{i,j}^{k,pred}-y_{i,j}^{k,target}\right|\right)\right]}{M_j \times M_i} \times 100 \quad (3)$$

where $y_{i,j}^{k,pred}$ and $y_{i,j}^{k,target}$ represents the occupancy ratio value of the predicted OGM and ground truth OGM at time step k of the grid cell in *Mi* by *Mj* multi-channel OGM which is indexed with *i* and *j*. With this performance metric, deep learning architecture for future prediction considering consecutively occurring multiple interaction pairs is evaluated over validation dataset which is not used for training. As a result of evaluation, 90 percentage of the accuracy is guaranteed up to 3 seconds of prediction horizon as shown in fig 10.

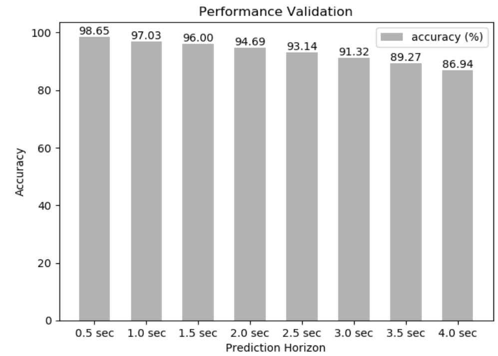

Figure 10. Accuracy of predicted occupancy ratio of all of the grid cells in multi-channel OGM structure up to 4 seconds.

In order to ensure the prediction performance during a complex and interactive scene, a highly interactive traffic scenario is selected, and its prediction results are analyzed in detail. In this scenario, there are three main agents, which cause multiple interaction pairs interacting with each other. These interaction pairs due to conflicting vehicles are caused by a left-lane changing vehicle (a) in fig 13. Notations of each grid cell in fig 13 is shown in fig 11.

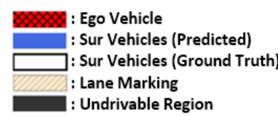

Figure 11. Notations of the grid cell of multi-channel OGM shown in fig 13

As shown in fig 13, future maneuver of the vehicle (b) and vehicle (c) affected by vehicle (a) is quite accurately predicted by the developed deep learning architecture. The overall prediction accuracy of the proposed algorithm across the simulation time is shown in fig 12. There are two key interaction pairs in this situation, vehicle (a) to (b) and vehicle (a) to (c), which are simultaneously occurring. By introducing multi-channel OGM based scene representation framework, future prediction algorithm for autonomous vehicle can handle these simultaneously occurring inter-vehicular interaction pairs. In this scenario, there are two main events as depicted in fig. 13. The first event (event 1) is lane change maneuver of the vehicle (a), and the second event (event 2) is the interactive maneuver of the vehicle (b) and vehicle (c). Figure 12 illustrates that the event 1 is well-predicted without any accuracy reduction. Furthermore, the future maneuver influenced by the surrounding vehicle's maneuver, event 2, is accurately predicted without any accuracy loss up to 3 seconds of prediction horizon. However, long-term (4 sec) prediction of the interactive maneuver leads to slight reduction in prediction accuracy.

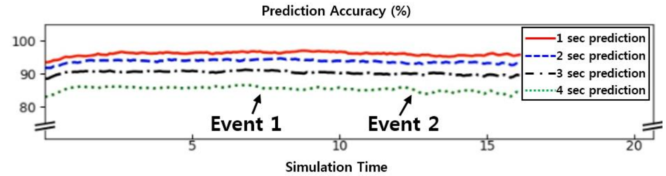

Figure 12. Prediction accuracy change with special traffic events: active lane change maneuver of vehicle (a) (Event 1) and interactive maneuver of vehicle (b) and vehicle (c) affected by vehicle (a) (Event 2)

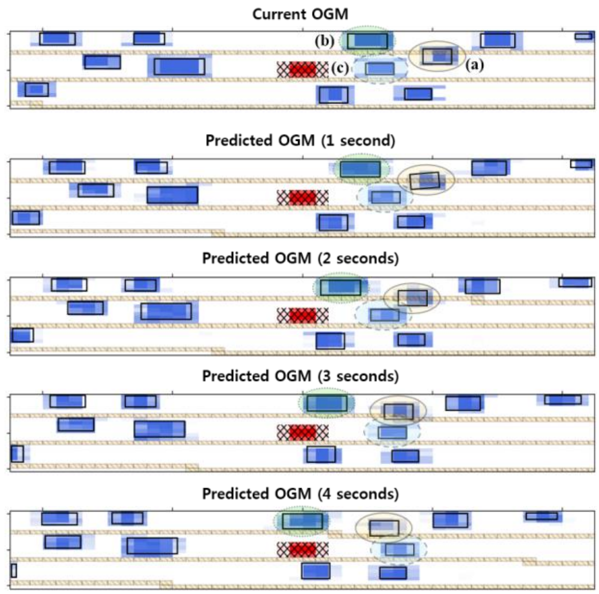

Figure 13. Time-serial interactive prediction result as a manner of the multi-channel OGM. Vehicle (b) is decelerating because of lane change maneuver of the vehicle (a) with deceleration. In case of vehicle (c), free headway is appeared due to vehicle (a)'s lane change which lead acceleration maneuver. All of above maneuvers with simultaneously and sequtianlly varying inter-vehicular interaction are effectively predicted due to the proposed integrated scene representation framework with CNN+LSTM based deep learning architecture. Video is available online [1]

VI. CONCLUSION

In this paper, the overall framework of scene representation is proposed in order to handle multiple interaction pairs in future prediction algorithm via deep learning architecture. Consequently, multi-channel occupancy grid map is built composed with dynamic layer and static layer. The proposed multi-channel OGM is the entire traffic scene representation method describing the occupancy of the dynamic objects and infrastructure. By introducing integrated scene representation method, simultaneously occurring multiple interaction pairs can be illustrated in one image from bird's eye view. By introducing CNN in the future prediction, explicitly illustrated multiple interaction pairs can be extracted into feature map. Afterward, future occupancy of the objects is predicted via LSTM network which is capable of handling sequentially varying interaction feature due to conflicting driving strategy of the vehicles. In highly interactive traffic scenario, future maneuver due to inter-vehicular interaction is accurately predicted via the proposed deep learning architecture using multi-channel OGM [1]. However, the proposed deep learning architecture is primary architecture for interactive future prediction. Thus, in future work, deep learning architecture will be developed in order to cope with various traffic scenario. Additionally, we will also focus on the comparison with previously developed prediction algorithm by suggesting appropriate performance metric which can be applicable to all of the prediction algorithms.

REFERENCES

[1] https://youtu.be/V0cwWvSqvtQ

[2] P. Kumar, M. Perrollaz, S. Lefever and C. Laugier, "Learning-Based Approach for Online Lane Change Intention Prediction," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, Golda Coast, Australia, Jun. 2013.

[3] S. Yoon and D. Kum, "The Multilayer Perceptron Approach to Lateral Motion Prediction of Surrounding Vehicles for Autonomous Vehicles," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, Gothenburg, Sweden, Jun. 2016.

[4] M. Schreier, V. Willert and J. Adamy, "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment ini Arbitrary Road Environmnets," *IEEE Transactions on Intelligent Transportation Systems,* vol. 17, no. 10, pp. 2751-2766, Oct. 2016.

[5] B. Kim, C. Kang, S. Lee, H. Chae, J. Kim, C. Chung and J. Choi, "Probabilistic Vehicle Trajectory Prediction over Occupancy Grid Map via Recurrent Neural Network," *arXiv preprint arXiv:1704.07049,* Sep. 2017.

[6] D. Lee, Y. P. Kwon, S. McMains and J. K. Hedrick, "Convolution Neural Network-based Lane Change Intention Prediction of Surrounding Vehicles for ACC," in *Transactions on Computational Science XIII*, pp.125-146, Apr. 2011.

[7] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr and M. Chandraker, "DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, Apr. 2017.

[8] A. Zyner, S. Worrall, J. Ward and E. Nebot, "Long Short Term Memory for Driver Intent Prediction," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, CA, USA, Jun. 2017.

[9] D. Lenz, F. Diehl, M. T. Le and A. Knoll, "Deep Neural Networks for Markovian Interactive Scene Prediction in Highway Scenarios," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, CA, USA, Jun. 2017.

[10] A. Elfes, "Occupancy Grids: A Stochastic Spatial Representation for Active Robot Perception," in *Proceedings of the Sixth Conference on Uncertainty in AI*, MA, USA, Jul. 1990.

[11] A. Elfes, "Using Occupancy Grids for Mobile Robot Perception and Navigation," *Computer,* vol. 22, no. 6, pp. 46-57, Jun. 1989.

[12] N. Srivastava, E. Mansimov and R. Salakhutdinov, "Unsupervised Learning of Video Representations using LSTMs," in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, Jul. 2015.

[13] Y. Bengio, P. Simard and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult," *IEEE Transactions on Neural Networks,* vol. 5, no. 2, pp. 157-166, Mar. 1994.

[14] J. Chung, C. Culcehre, K. Cho and Y. Dengio, "Enpirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," *arXiv preprint arXiv:1412.3555,* Dec. 2014.

[15] FHWA, "Next Generation Simulation (NGSIM) Program," Federal Highway Administration, 2006. [Online]. Available: http://ngsim-community.org/. [Accessed 11 6 2017].

[16] D. P. Kingma and J. L. Ba, "ADAM: A Method for Stochastic Optimization," in *the 3rd International Conference for Learning Representations (ICLR)*, CA, USA, May. 2015.

[17] F. Chollet, "Keras," 2015. [Online]. Available: http://keras.io.

[18] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu and X. Zheng, "TensorFlow: A System for Large-Scale Machine Learning," in *the Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, GA, USA, Nov. 2016.

[19] J. Kim, "Probabilistic Collision Risk Assessment via Lane-based Motion Prediction of Surrounding Vehicles," in *Master Thesis*, Korea Advanced Institute of Science and Technology, Daejeon, Korea, 2016.

[20] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning (PMLR)*, Lille, France, Jul. 2015.