Leveraging Dynamic Interaction Models for Anomalous Behavior Detection in 3D Environments

Saleemullah Memon[1,2], Ali Krayani[1], Pamela Zontone[1], Lucio Marcenaro[1],
David Martin Gomez[2], and Carlo Regazzoni[1]
[1]*Department of Engineering and Naval Architecture (DITEN), University of Genova, Italy*
[2]*Intelligent Systems Lab, University Carlos III de Madrid, Spain*
saleemmemon1320@gmail.com, ali.krayani@edu.unige.it,
{pamela.zontone, lucio.marcenaro, carlo.regazzoni}@unige.it, dmgomez@ing.uc3m.es

***Abstract*—This paper contributes a novel approach to learning the dynamic patterns of multiple vehicles moving on a 2-lane road and performing interactions between them. The proposed data-driven, hierarchical, and self-aware approach, called coupled generalized dynamic Bayesian network (C-GDBN), learns from 3D LiDAR sensor observations. A separate GDBN model is learned from the generalized states (GSs), applying an unsupervised growing neural gas (GNG) approach for each vehicle. Then, a higher hierarchy with the global dictionary is formed, which allows the coupling interaction between the learned GDBN models. A coupled Markov jump particle filter (C-MJPF) is proposed during the testing phase, which allows making the inference on the optimum generative model. Our scheme performs better under known and unknown situations in terms of both low and high-level predictions and anomaly detections, which consequently meet the explainable challenges in the current intelligent approaches.**

***Index Terms*—Coupled Generalized Dynamic Bayesian Network (C-GDBN), 3D LiDAR, Self-Awareness, Autonomous Systems**

I. Introduction

With the rapid development of vehicular sensing-based intelligence, an increasing number of proprioceptive and exteroceptive sensors are being seamlessly integrated into vehicles, enhancing their efficiency, sensing capability, and robustness [1], [2]. These could also be integrated with physiological sensors to improve the safety and well-being of drivers [3], [4]. Compared to various vehicle sensors such as GPS (poor indoors), Camera (complex depth estimation), Ultrasonic (short range), RaDAR (low resolution), and Infrared (sensitive to lighting), LiDAR performs better with high-resolution 3D mapping, accurate depth sensing, perceiving a complete 360-degree view as well as detecting tiny objects of the environment [5]–[7]. Taking these features into account, many research problems are investigated in the literature considering LiDAR sensors, such as object detection and tracking [8], occlusion forecasting [9], localization [10], and place recognition [11].

Continuous vehicle tracking and detection remain challenging for autonomous agents due to missing sensor observations. This anomalous situation often occurs when sensors are obstructed by other physical objects, provide unclear and noisy data, malfunction, or operate under harsh weather conditions. However, by harnessing the advanced capabilities of artificial intelligence (AI), deep learning (DL), and machine learning (ML), together with the use of prior sensor information, dynamic abnormal behaviors can be predicted and detected, allowing proactive measures to address these challenges [12].

Most of the existing AI techniques developed to predict this type of abnormal behavior using LiDAR sensors are characterized by a black-box nature, high computational complexity, dependence on large amounts of supervised data, and a lack of self-awareness or explainable capabilities [12], [13]. In contrast, probabilistic reasoning, hierarchical, and data-driven approaches support adaptive learning, offer explainable features by modeling transitions and uncertainties from hidden nodes, and facilitate multisensor fusion [14].

In light of the aforementioned discussions, we are inspired to develop a probabilistic and data-driven approach, i.e., a coupled generalized dynamic Bayesian network (C-GDBN), which learns from the temporal sequences of LiDAR point clouds. The main goal of the proposed C-GDBN is to learn the dynamic interaction among various vehicles moving in the LiDAR's area of interest. For testing, we introduce a coupled Markov jump particle filter (C-MJPF) to perform the predictions (at the state and word levels) and detect anomalies (at the word level). The major contributions of this paper are as follows.

- Learning the dynamic patterns of multiple vehicles, detected by the joint probabilistic data association filter (JPDAF) from the LiDAR sensor observations. Hence, training multiple probabilistic models separately through the unsupervised growing neural gas (GNG) clustering approach and creating their dictionaries.
- Learning the interaction among multiple vehicles and creating a world model (i.e., a global dictionary), also called the C-GDBN model, by considering the sequences of super-state discrete variables (clusters or nodes).
- Re-associating and re-identifying the dynamics of the tracked vehicles obtained by the JPDAF inside the region of interest for continuous tracking.
- Predicting the states at the lower and words at the higher, as well as detecting abnormal behavior at the higher hidden levels from the C-MJPF in both the known and unknown scenarios, after selecting the best generative model(s).

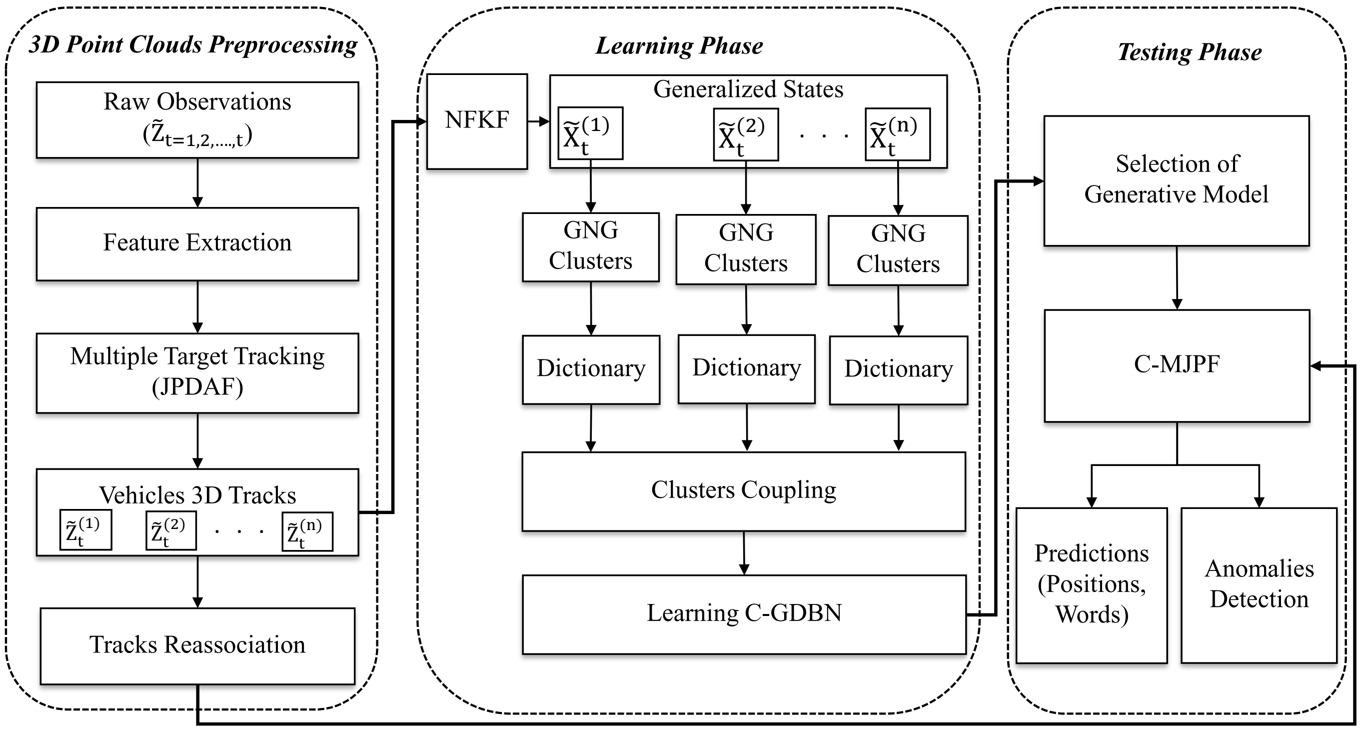


Fig. 1: A schematic representation of the proposed design scheme.

The rest of this paper is structured as follows. The overview of the LiDAR dataset, data pre-processing, multiple vehicle detection and tracking, offline learning, and online testing stages are all introduced in Section II. The simulation results are discussed in Section III. Finally, the conclusions and potential directions for further research are provided in Section IV.

II. Proposed Framework

The proposed framework consists of three stages: the 3D LiDAR dataset and pre-processing stage, the offline learning stage, and the online testing stage, as illustrated in Fig. 1.

A. Dataset Overview & Pre-processing Stage

In this work, we consider the deep-sense 6G dataset [15], which is based on real-world observations and is well-suited to our scenario. The dataset captures a street-level vehicle-to-infrastructure communication scenario, specifically scenario 32, in a wireless outdoor environment. The data is collected on a two-way, two-lane street at College Avenue in Arizona, USA. Here, we have only considered the LiDAR data. The point clouds obtained by the sensor allow accurate mobility patterns and locations of the vehicles, pedestrians, and other static and dynamic objects available in the scene. The LiDAR is mounted on a stationary vehicle, has a 100-meter range, and has a maximum motor spin frequency of 20 Hz.

In the initial phase, essential features of dynamic objects within the scenario are extracted from the sensor's raw time-series observations. Various filtering techniques, including neighbor denoising, ground removal, and cropping, are carefully applied to each point cloud to eliminate outliers, remove the ground, and exclude regions beyond road boundaries. To address the point cloud association problem, a JPDAF approach is utilized, ensuring accurate detection and tracking of multiple vehicles within the LiDAR field of view [5]. The

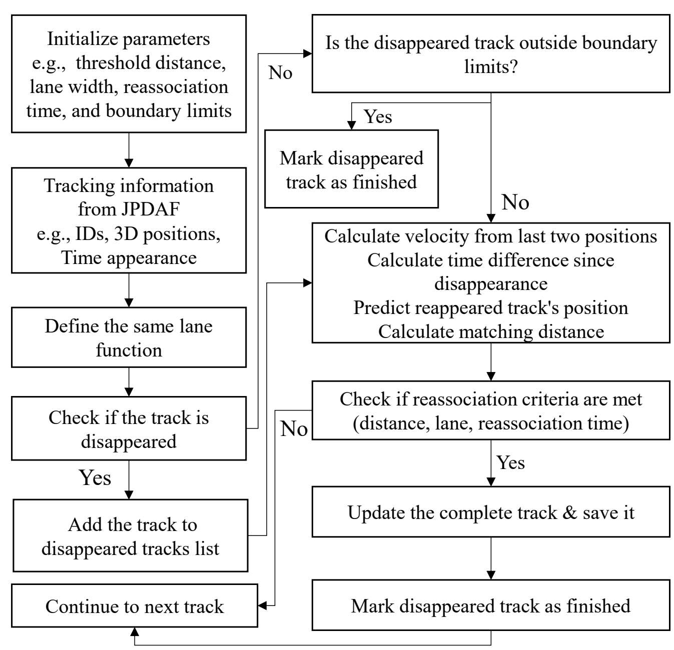
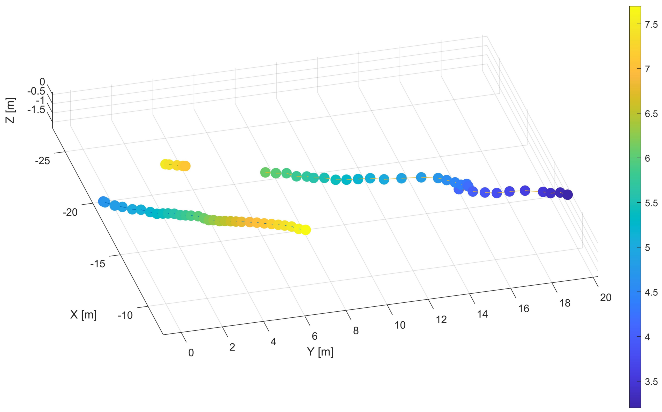
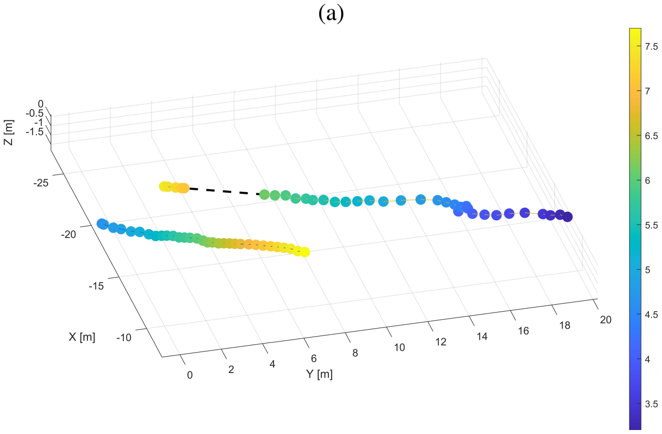


Fig. 2: A flowchart representing the algorithm for reassociation and reidentification of disappeared vehicles.



(a)



(b)

Fig. 3: Anomalous situation caused by the dynamic occlusion moving on the other lane: (a) An example of a vehicle disappearing and reappearing, (b) An example of reassociation and reidentification of the same vehicle after it disappears from the sensing environment.

output of JPDAF consists of 3D tracks of detected vehicles, including their x, y, and z positions, time of appearance, and tracking IDs. In addition to continuous tracks, some discontinuous (broken) tracks due to the abnormal situation are also obtained, which require further processing. The example of an abnormal track is shown in Fig. 3(a) (the color bar represents the time information), where a vehicle initially detected in a lane appears in the scene at time 3.2 but disappears at time 6.1 due to the occlusion moving in another lane. The same vehicle reappears at time 7.1 after passing the obstruction. Such abnormal tracks are recovered using the proposed reassociation and reidentification algorithm, as illustrated in Fig. 2. This algorithm records the disappearance and reappearance times of the blocked tracks, calculates velocity based on the last known positions and the time elapsed since disappearance, predicts missing positions for each time step, and computes the matching distance between the predicted and detected positions when the object reappears. If the matching distance falls within a predefined threshold and the vehicle remains in the same lane, the track will be reassociated with the previously disappeared one. An example of a disappeared track,

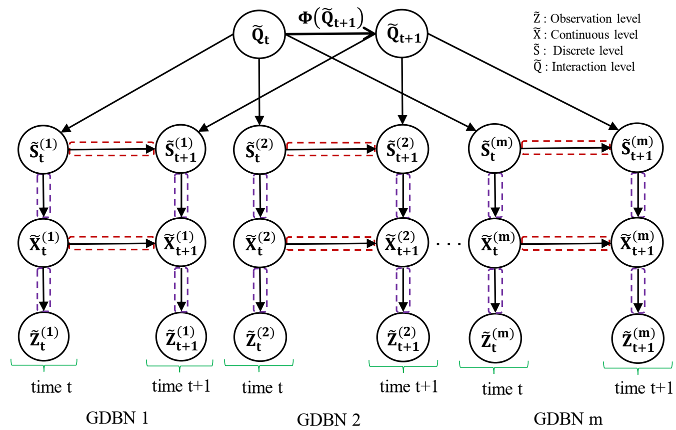


Fig. 4: A graphical representation of learning C-GDBN from 3D LiDAR point clouds.

along with the output of the reassociation and reidentification algorithm, is shown in Fig. 3(a) and Fig. 3(b), respectively.

B. Offline Learning Stage

For the learning stage, we first extract the generalized states (GSs) containing a 6D vector of positions and velocities for all continuous tracks. These GSs are obtained through the null-force Kalman filter (NFKF), which assumes that vehicles are not in motion [5]. To encode and model the dynamic GSs of each tracked vehicle individually, we employ an unsupervised GNG clustering approach. This neural network technique is well-suited for our 3D dynamic environment as it does not require a fixed number of clusters (discrete random variables or nodes) and can incrementally adapt to the structure of incoming data. For each vehicle, we construct a separate dictionary to learn an independent GDBN model, as illustrated in Fig. 4. Each dictionary comprises clusters encoding the GSs, a transition matrix (TM) that captures the probabilities of transitioning between nodes over time, temporal TMs, a covariance matrix, and the mean of each node. Consequently, multiple dictionaries are created to represent different GDBN models.

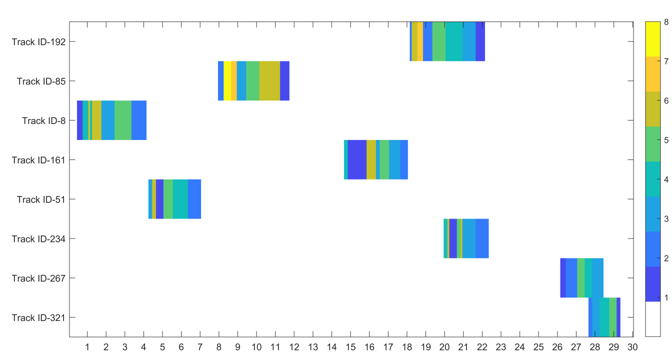


Fig. 5: Clusters timeline view representing the dynamic interactions of tracked vehicles across time.

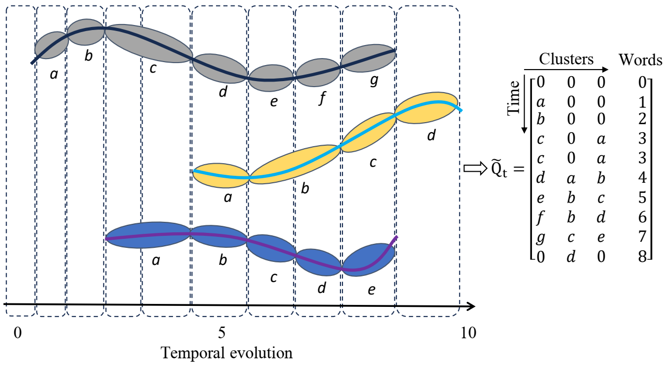


Fig. 6: An example illustrating cluster interaction.

To model interactions among learned GDBN models, we establish a top-level hierarchy (see Fig. 4) that captures the relationships between discrete random variables at each time step. The interactions among the clusters of various tracking vehicles across temporal evolution are illustrated in Fig. 5. These interactions are formed through discrete entities called words (see Fig. 6). Here, a word with value zero indicates the absence of vehicles in the environment, while identical words are assigned when the same sequence of clusters is repeated. Otherwise, words are assigned in ascending order at each time step. The global dictionary for C-GDBN contains all the GDBN dictionaries, the set of words, and a word-level TM that represents the probabilities of transitioning from one word to another.

C. Online Testing Stage

As multiple GDBN models are trained on temporal sequences of LiDAR point clouds, selecting the most suitable model for a given testing track is crucial for better prediction performance. To achieve this, we identify the nearest clusters of a generative model for a defined testing track using Euclidean distance, which measures the distance from the mean of all dictionary nodes. The model with the lowest error (energy) will be selected to make the inference. The closest nodes are then used to map the estimated words, updating the belief of the C-GDBN.

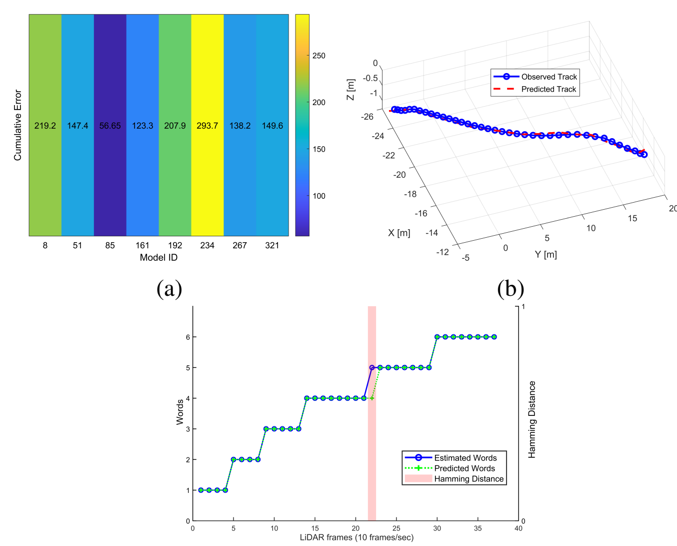
To detect abnormal behavior and perform temporal and hierarchical predictions, we propose a C-MJPF that integrates the particle filter and the Kalman filter. The C-MJPF propagates multiple particles, each representing a hypothesis about the system's current state. Initially, all particles are assigned equal weights. The first word is randomly sampled from the probability distribution to infer at the top level. The transition probabilities of the estimated word then guide the prediction of the next word. Particle weights are updated based on the estimated word at each time step, and finally, the particles are resampled, with those having higher weights being more likely to be selected.

III. Model Performance Analysis & Discussion

This section evaluates the performance of our proposed C-GDBN model. We analyze its effectiveness in terms of optimal generative model selection, state-level predictions, word-level predictions, and word-level anomaly detection. The optimal model is chosen by finding the nearest learned clusters. Here, we use the Euclidean distance between each 3D position of the given testing track and the mean of all the clusters. Cumulative errors are calculated while selecting the generative model. Since the estimated and predicted words are discrete, the Hamming distance is a suitable metric for abnormality detection. A mismatch between the predicted and estimated words results in a binary value of "1", while a match yields "0". In addition, we evaluate performance in both known and unknown scenarios. The same set of tracks used to train the GDBN models is also used for testing in known scenarios, while in unknown scenarios, testing is performed using abnormal tracks that are obstructed by other objects.

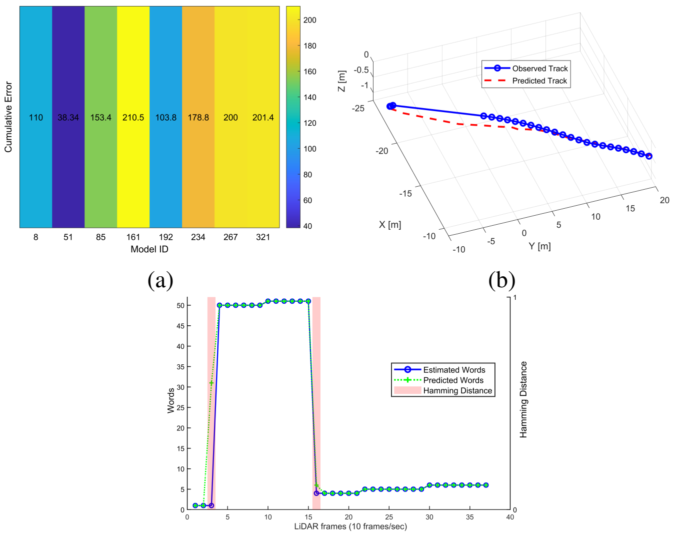
The performance of the model under the known scenario is shown in Fig. 7. For a given testing track, the best generative model chosen is 85 because it achieves fewer errors in finding the nearest clusters compared to other trained models. The low-level state predictions shown in Fig. 7(b) match the observations coming from the LiDAR sensor, and the high-level word predictions shown in Fig. 7(c) match the estimated words that are mapped after finding the nearest clusters of the selected model. The accurate predictions reflect a better understanding and learning abilities of our proposed model.

An example shown in Fig. 8 illustrates the model's performance in an unknown scenario. In this case, the best generative model for a given broken track is 51 because it closely matches the testing data and achieves fewer distance errors (see Fig. 8(a)). Fig. 8(b) shows an abnormal observed track, which disappears due to the blockage after a few observations. The predictions initially match the observations, then deviate



(a) (b) (c)

Fig. 7: Optimal generative model selection and predictions in a normal scenario: (a) Error analysis in selecting the best model for the testing track, (b) Low-level state prediction, (c) High-level word prediction.



(a) (b)

(c)

Fig. 8: Optimal generative model selection and predictions in an abnormal scenario: (a) Error analysis in selecting the best model for the testing track, (b) Low-level state prediction, (c) High-level word prediction.

due to the absence of observations. This anomalous situation occurred for a very short period; however, as soon as the LiDAR point-cloud observations appeared, the model started to give the correct predictions. Similarly, the high-level word prediction under an unknown situation is shown in Fig. 8(c). We can observe that when the observations disappear, the model predicts blockage words (50 and 51 as almost two clusters get missed). The anomaly indicator represented by the hamming distance occurs at LiDAR frames 3 and 16 are due to the model switching. As the observations appear from LiDAR frame 16, the model correctly predicts the normal words.

IV. Conclusion & Future Work

This paper proposes a probabilistic, hierarchical, data-driven, and interactive dynamic Bayesian approach to learning the dynamic point cloud patterns from the 3D LiDAR sensor. Initially, a separate dictionary for each detected vehicle is created, and then a global dictionary is formed, allowing interaction among various vehicles. The optimum model for the C-MJPF is selected during the testing phase to make inferences. Our proposed C-GDBN model allows for better performance in terms of low-level state predictions, as well as high-level word predictions and abnormality detections. Moreover, this interactive approach provides better explainable features because of its hierarchical structure, which models uncertainty and transitions from hidden patterns.

Future extensions of this work could take into account additional data modalities, such as the integration of 3D LiDAR with communication signals. Building interactions between the two would be an interesting and useful approach for self-aware agents. This highlights the concept of integrated sensing and communication (ISAC) for future communication systems.

Acknowledgment

This research was partially funded by the European Union's Horizon Europe research and innovation programme under the Grant Agreement No. 101121134 (B-prepared), and by the European Union under the Italian National Recovery and Resilience Plan (PNRR) of NextGenerationEU partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"), CUP E63C22002040007 - D.D. n.1549 of 11/10/2022. This research was also funded by the Spanish Government under Grants: PID2021-124335OB-C21, PID2022-140554OB-C32, TED2021-129485B-C44, and PDC2022-133684-C31.

References

[1] H. Li, Y. Chen, K. Li, C. Wang and B. Chen, "Transportation Internet: A Sustainable Solution for Intelligent Transportation Systems," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 12, pp. 15818-15829, 2023.

[2] G. Slavic, M. Bracco, L. Marcenaro, D. M. Gómez, C. Regazzoni and P. Zontone, "Joint Data-Driven Analysis of Visual-Odometric Anomaly Signals in Generative Ai-Based Agents," 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp. 264–268, 2024.

[3] T. A. Najafi, A. Affanni, R. Rinaldo, and P. Zontone, "Drivers' mental engagement analysis using multi-sensor fusion approaches based on deep convolutional neural networks," Sensors, vol. 23, no. 17, 2023.

[4] A. L. Grasso, P. Zontone, R. Rinaldo, and A. Affanni, "Advanced necklace for real-time PPG monitoring in drivers," Sensors, vol. 24, no. 18, 2024.

[5] S. Memon, A. Krayani, P. Zontone, L. Marcenaro, D. M. Gomez and C. Regazzoni, "Learning 3D LiDAR Perception Models for Self-Aware Autonomous Systems," 2024 27th International Conference on Information Fusion (FUSION), pp. 1–8, 2024.

[6] L. Zhao, H. Zhou, X. Zhu, X. Song, H. Li and W. Tao, "LIF-Seg: LiDAR and Camera Image Fusion for 3D LiDAR Semantic Segmentation," IEEE Transactions on Multimedia, vol. 26, pp. 1158–1168, 2024.

[7] H. Iqbal, D. Campo, P. Marin-Plaza, L. Marcenaro, D. M. Gómez and C. Regazzoni, "Modeling Perception in Autonomous Vehicles via 3D Convolutional Representations on LiDAR," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 14608–14619, 2022.

[8] Z. Meng, X. Xia, R. Xu, W. Liu and J. Ma, "HYDRO-3D: Hybrid Object Detection and Tracking for Cooperative Perception Using 3D LiDAR," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 8, pp. 4069-4080, 2023.

[9] S. Bannai and K. Suto, "Stereo-Aided Blockage Prediction for mmWave V2X Communications," 2024 International Conference on Computing, Networking and Communications (ICNC), pp. 624-628, 2024.

[10] I. Hroob, B. Mersch, C. Stachniss and M. Hanheide, "Generalizable Stable Points Segmentation for 3D LiDAR Scan-to-Map Long-Term Localization,", IEEE Robotics and Automation Letters, vol. 9, no. 4, pp. 3546-3553, 2024.

[11] J. Ma, G. Xiong, J. Xu and X. Chen, "CVTNet: A Cross-View Transformer Network for LiDAR-Based Place Recognition in Autonomous Driving Environments," IEEE Transactions on Industrial Informatics, vol. 20, no. 3, pp. 4039-4048, 2024.

[12] S. Wu, C. Chakrabarti and A. Alkhateeb, "LiDAR-Aided Mobile Blockage Prediction in Real-World Millimeter Wave Systems," 2022 IEEE Wireless Communications and Networking Conference (WCNC), pp. 2631–2636, 2022.

[13] D. Marasinghe, N. Rajatheva and M. Latva-aho, "LiDAR Aided Human Blockage Prediction for 6G," 2021 IEEE Globecom Workshops (GC Wkshps), pp. 1–6, 2021.

[14] A. Krayani, G. Barabino, L. Marcenaro and C. Regazzoni, "Integrated Sensing and Communication for Joint GPS Spoofing and Jamming Detection in Vehicular V2X Networks," 2023 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–7, 2023.

[15] A. Alkhateeb, et al., "DeepSense 6G: A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset," IEEE Communications Magazine, vol. 61, no. 9, pp. 122–128, 2023.