Title: System and Method for Environment Perception Based on Radar

Automotive vehicles have been equipped with more and more advanced automated driving functions with the aim of achieving Level 5 autonomous driving goals. Along the way, those advanced automation functions need multiple sensors with various environment sensing capabilities. Automotive radar is one of such sensors that allow automotive vehicles to run safely because of its advantage of all-weather and all-time operation-ability. Starting with the basic functions of target detection and ranging, advanced automotive radar can now produce high resolution point cloud that represents the objects located in the surrounding environment. Based on the point cloud, advanced perception can be carried out to provide information for advanced driving assistance system (ADAS).

Point cloud perception consists of various functions regarding different targets in automotive radar. In general, automotive radar targets can be divided into two categories, i.e., static objects and dynamic objects. Static objects include immobile things such as buildings and road infrastructures, while dynamic objects include moving objects of different types such as cars and pedestrians. Basic perception function is needed to differentiate between dynamic and static categories and advanced perception function needs to predict the semantic types and instance identifications (IDs) of different targets within each category. In general, the information on the category and types of targets may correspond to different reactions and further processing for higher-level functionalities in ADAS. For example, occupancy grid map (OGM) and free space detection (FSD) can be carried out on static objects and target tracking can be performed with dynamic objects. These functions of classification and target instance prediction can be realized through segmentation.

There are two types of conventional segmentation techniques. One is semantic segmentation and the other is instance segmentation. Semantic segmentation can provide the target class information while instance segmentation can provide the target instance ID. Semantic segmentation is usually based on convolutional neural networks for image processing. Instance segmentation is usually performed with object detection involving region-based or proposal-based detectors. Those two segmentation techniques usually are conducted separately. Recently, a new paradigm of segmentation called panoptic segmentation has been proposed in image processing that unifies the typically distinct tasks of semantic segmentation and instance segmentation, providing not only the class information but also the target instance for each point. There are few works on panoptic segmentation and most of them lie in the area of image processing and point cloud generated in Lidar. In those works, the network models have a similar general structure that starts with a feature extraction stage followed by two branches dedicated to semantic classification and instance prediction, respectively. In panoptic segmentation, the instance prediction is more challenging because the instance ID assignment may not be unique and implicitly involves clustering processing. Therefore, the existing works are not end-to-end differentiable and not efficient in computation because of intermediate steps for clustering. In general, those exiting methods can only achieve a frame rate of around or less than 10 frames per second, which is far lower than the implementation requirement.
Automotive radar has many advantages over other sensors like Lidar and cameras. However, the radar measurement may be subject to multipath. Multipath may create ghost targets that affect the perception and subsequent decision on driving actions. Multipath is less studied in the automotive radar domain. Some existing works aimed to deal with multipath separately that increase the complexity and latency of the system.

We present a compact perception based on panoptic segmentation. In our method, we predict the instance label of each object point directly through an ordered labeling procedure in the training and prediction. We also conduct the multipath detection as a segmentation with a separate branch.

Detailed description of each component of your invention, e.g. functions of each component, and interaction between respective components are as below:

1) Figure-1: radar system architecture

   The radar system 001 for which the antenna array is designed has a waveform generator 100 that generates waveforms such as continuous waveforms and pulse radar waveforms. Meanwhile, waveforms generated for different transmitters are orthogonal to each other in order to exploit the benefit of MIMO radar. The orthogonality can be achieved in different ways such as time division multiplexing (TDM) and code division multiplexing (CDM). Our patent is not confined to the specific use of the type of orthogonal waveforms. The waveform undergoes the transmitter circuit 102 before being transmitted from transmit antenna array 104. In the MIMO radar operation, each transmit antenna transmits its own waveform that is orthogonal to the waveforms to be transmitted by other transmitters. The geometry of transmit array can be 1D array or 2D array of desirable array configuration. After being reflected from target, the radar signals are received at the receive antenna array 106 that is connected to the receiver circuit 108. The geometry of receive array can be 1D array or 2D array of desirable array configuration. In fact, each receive antenna array receives a combination of echo signals due to all the transmitted signals. After that, receiver baseband 110 samples the radar signal into digital domain for signal processing. The radar system is controlled by a radar control unit 112 to maintain the basic functionalities of the radar system such as transmitting, receiving, and timing, etc. The radar system 001 may be connected to other devices or sensor network control & processing system 114 so that fusion of information from different sensors can be carried out to improve the perception. The signal processing unit 116 takes the baseband data as input and generates point cloud through the processing of range compression, Doppler processing, and DOA processing. With the point cloud results, the perception unit 118 conducts the environment perception.

2) Figure-2: functionality blocks of signal processing unit

   The function of signal processing unit is to generate point cloud by processing the raw ADC signal. The first step is the range processing 200, then followed by Doppler processing 202. With the range and Doppler processing, the information of the target's range and relative velocity will be achieved. Target detection 204 is to detect targets on the range-Doppler map obtained after range and Doppler processing. For each detected target in the range-Doppler domain, a virtual array will be formed according to the principle of MIMO radar in step 206. With virtual array formation, the step of DOA estimation will be carried out. With the information of range and DOA, the point cloud of targets can be formed.

3) Figure-3: functional blocks in the compact environment perception

   The core of this patent lies in the compact perception based on a novel panoptic segmentation approach. The perception starts with the input of point cloud 300 to the

process of panoptic segmentation 302. The input of the point cloud can be generated from a single radar sensor, or from multiple radar sensors. In the latter case, preprocessing of combining point clouds from different radar sensors is required. The panoptic segmentation generates the point-wise information, consisting of 1) the class information, 2) the object ID of each point, and 3) multipath information. Particularly, different from the existing panoptic segmentation, we also predict the multipath. Overall, the output of panoptic segmentation produces the point clouds belonging to multipath 304, static environment 306, and dynamic objects 308, respectively. The predicted multipath will be removed from the point cloud in the subsequent perception processing. From the point cloud of the static environment, occupancy grid map 310 and free space detection 314 will be carried out. Based on the point cloud of dynamic objects, target tracking and trajectory prediction will be carried out. The information of static and dynamic environments can be applied to higher-level functions such as fusion with other types of sensors and processing in the ADAS system.

4) Figure-4: Block diagram of the panoptic segmentation

Figure-4 illustrated the architecture of the novel panoptic segmentation model designed in this patent. Instead of transforming the point cloud into a different domain such as 2D image, the panoptic segmentation model works directly on the point cloud, which may be generated from one single radar sensor or from multiple radar sensors. The point cloud from radar may contain such raw features as position, Doppler velocity, signal to noise ratio (SNR), radar cross-section (RCS), signal magnitude, and additional information from other sensors such as ego velocity and yaw direction. In total, there are N points and each point has D raw features. A feature encoder, that can be implemented through PointNet model or other models, will process the point cloud to produce the global features, a matrix of extracted features that has N points and $M0$ features extracted for each point. Based on the global feature matrix, three independent functions for instance prediction, semantic classification, and multipath detection will be conducted via three different branches. In the instance network branch, the azimuth value is provided as a feature to the network that is in accordance with the labeling scheme making use of azimuth information for instance indexing. The method to inject this information can be through concatenating the azimuth as another feature or adding the azimuth information to the extracted global features for each point. In one example, the final network uses point-wise azimuth addition instead of concatenation for reduced dimension and complexity. The augmented feature was processed by a network of MLP before producing the instance segmentation matrix of dimension N x M, where M stands for the largest possible instance index. Hence, the network should be able to learn the relationship between azimuth feature and object instance. This method can be generalized to incorporate other features such as elevation if those features are used in the labeling stage.

In the semantic network, the model makes use of the global features directly through a MLP and produces the semantic segmentation matrix of dimension N x S where S is the number of categories such as car, truck, pedestrian, building, etc.

In the multipath network, the augmented features combining the global features from the feature encoder and some handcrafted features such as the number of neighboring points will be processed through MLP to predict the presence of multipath for each point.

In the model implementation, weighted negative likelihood loss (NLL) is used for semantic, instance, and multipath prediction branches. In a usual driving scenario, a

large number of detected points belong to static class hence there is the problem of class imbalance. To solve this problem, the model down-weighs the static class by assigning it a smaller weight so that the network is incentivized to learn moving classes like cars and trucks. The final overall loss function is a weighted sum of losses from instance predictions, semantic predictions, and multipath predictions.

5) Figure-5 illustration of panoptic segmentation labeling
Besides the network structure, the invention has designed a novel labeling scheme for the instance segmentation. The labeling scheme is jointly designed with the panoptic segmentation model. In particular, before the instance index labeling, a clustering step is applied to the point cloud so that the points corresponding to the same target are clustered into one cluster. This can be done, for example, manually by a human labeler or by automatic clustering methods. During the instance labeling, gradually increased instance labels are assigned based on the azimuth value. The azimuth value of the instance can be computed, for example, from the center point of the cluster of the instance. At the same time, during training and inference, azimuth values are provided as a feature to the network in the instance network branch. The method to inject this information can be through, but not limited to, concatenating the azimuth as another feature or adding the azimuth information to the extracted features for each point. The labeling scheme considering the azimuth information can be extended to other features such as elevation.

The patent solves the problem of compact perception, in particular, the problem of panoptic segmentation and multipath detection, using automotive radar. The patent solves this problem through the design of compact segmentation model that predicts the class information, instance information, and multipath information directly.

The main technical difficulty is how to design a network end-to-end differentiable and trainable with a simultaneous prediction of class, instance, and multipath for each point. We solve this problem by treating instance prediction and multipath detection as a classification problem that is based on a specially designed labeling scheme for both training and inference.

The difficulty of making such differences lies in the fact that instance prediction is conventionally solved through clustering-based approaches and there is no direct way of instance prediction. Meanwhile, the solution of multipath is solved separately in the existing methods without considering a compact solution to perception.

The essential components of our invention, i.e. the steps or the units which are indispensable in order for our invention to function as mentioned below:
1) In the radar system 001, all the components in the architect are essential to the system. However, there can be various practical forms for each component. For example, the waveform generator can be either FMCW waveform generator or pulsed waveform generator and even the phase modulated continuous wave (PMCW) generator.
2) The signal processing unit is an essential component of the invention.
3) The perception and panoptic segmentation are essential components of the invention.

About whether it is possible to change or replace any of the components by equivalent components without changing the invention is as below:
1) Referring to Figure-4, the multipath detection branch can be combined with the semantic segmentation branch with necessary modifications such as feature augmentation, Multilayer Perceptron (MLP) or it's derived variants design and semantic class matrix. In this case, the resulting network model will have two branches instead of three branches.
2) In the signal processing components, different signal processing structure can be used. For example, Doppler and angle processing can be swapped or combined into a 2D space-time processing

The compact perception method can be improved by incorporating information from other sensors such as other radar and Lidar. In this way, the perception is a fusion of multiple sensors. Thus with the mentioned components when added to our invention, can provide further functions and even be able to achieve better effect.

Claims

While some embodiments of the invention have been described by the illustrations with diagrams, a person skilled in the art will recognize that the invention can be carried out with many modifications, variations, and adaptions without departing from the scope of this invention defined in the following claims.

1. An efficient method for achieving environment perception based on automotive radar consisting of:
   1) A compact method of environment perception built on the basis of panoptic segmentation
   2) A method of designing and implementing the panoptic segmentation model consisting of one shared encoder stage and three separate functional branches corresponding to instance segmentation, semantic segmentation, and multipath detection, respectively.
   3) A method of data labeling for instance segmentation.
2. The method of claim 1, where the method of compact environment perception consists of:
   1) a basis of efficient panoptic segmentation that takes point cloud from radar sensors as input and outputs the target class information, target instance ID, as well as multipath detection
   2) from the results of panoptic segmentation, static environments are separated from the static classes, by, for example, indexing the points with class types belonging to static classes
   3) from the static environment, the perception functions like the OGM and FSD will be carried out by processing the point cloud belonging to the static environment.
   4) from the results of panoptic segmentation, dynamic environments are separated from the dynamic target classes, by, for example, indexing the points with class types belonging to static classes
   5) from the result of the dynamic environment, target tracking can be carried out by processing the points belonging to dynamic objects
   6) from the result of panoptic segmentation, multipath points are detected and removed.
3. The method of claim 1, where methods of the design and implementation of the efficient panoptic segmentation model consist of:
   1) one feature encoder network that takes raw point cloud along with other information such as ego-motion information as input and produces a matrix of global features
   2) one instance segmentation branch that processes the augmented global features using an MLP and produces the instance information for each point
   3) one semantic segmentation branch that processes the global features using a MLP and produces the semantic information for each point
   4) one multipath branch that processes the global features augmented with additional information and produces the multipath detection
   5) The training of the model on the training data by applying the weighted sum loss functions combining each loss function of the abovementioned individual branch, i.e., instance segmentation branch, semantic segmentation branch, and multipath branch
   6) applying the trained model to real data for panoptic segmentation in real time

4. The method of claim 3, where the feature augmentation in the instance segmentation branch may be
   1) implemented as augmentation with azimuth information, and other possible information such as elevation according to the labeling scheme
   2) implemented through, but not limited to, concatenating the azimuth as another feature or adding the azimuth information to the extracted features for each point

5. The method of claim 1, where the data labeling for instance segmentation consists of the following steps:
   1) clustering step that is applied to the point cloud so that the points corresponding to the same target are clustered into one cluster. This can be done, for example, manually by a human labeler or by automatic clustering methods.
   2) instance labeling step when gradually increased instance labels are assigned based on the azimuth value and the azimuth value of the instance can be computed, for example, from the center point of the cluster of the instance.

6. A system for achieving environment perception based on automotive radar consisting of:
   1) a system of panoptic segmentation model trained on labeled data, consisting of one shared encoder stage and three separate functional branches corresponding to instance segmentation, semantic segmentation, and multipath detection, respectively
   2) a compact system of environment perception built on the system of panoptic segmentation system

7. The system of claim 6, where the system of compact environment perception consists of:
   1) a system of efficient panoptic segmentation that takes point cloud from radar sensors as input and outputs the target class information, target instance ID, as well as multipath detection
   2) from the results of panoptic segmentation, static environments are separated based on point clouds of static classes
   3) from the static environment, the perception functions like the OGM and FSD will be carried out by processing static point clouds.
   4) from the results of panoptic segmentation, dynamic environments are separated based on point clouds of dynamic target classes
   5) from the result of the dynamic environment, target tracking can be carried out on point clouds of dynamic targets
   6) from the result of panoptic segmentation, multipath points are detected and removed.
8. The system of claim 6, where the system of efficient panoptic segmentation model consists of:
   1) one feature encoder network system that takes raw point cloud along with other information as input and produces a matrix of global features
   2) one instance segmentation system that processes the augmented global features using an MLP and produces the instance information for each point
   3) one semantic segmentation system that processes the global features using a MLP and produces the semantic information for each point
   4) one multipath system that processes the global features augmented with additional information and produces the multipath detection

9. The system of claim 8, where the feature augmentation in the instance segmentation system may be
   1) implemented as augmentation with azimuth information, and other possible information such as elevation according to the labeling scheme
   2) implemented in the system through, but not limited to, concatenating the azimuth as another feature or adding the azimuth information to the extracted features for each point
10. A system of the data labeling for instance segmentation consists of:
    1) clustering system that is applied to the point cloud so that the points corresponding to the same target are clustered into one cluster. This can be realized as automatic clustering software.
    2) an instance index assignment system that during the instance labeling assigns gradually increased instance labels to target belonging to the same class based on the azimuth value and the azimuth value of the target instance can be computed, for example, from the center point of the cluster of the instance.