

Description

FEATURE-LEVEL FUSION METHOD AND SYSTEM FOR COOPERATIVE PERCEPTION BASED ON VEHICLE-INFRASTRUCTURE COOPERATION

Technical Field

The present invention relates to the technical field of cooperative perception for autonomous driving of vehicles, in particular to a feature-level fusion method and system for cooperative perception based on vehicle-infrastructure cooperation.

Background Art

With the deepening of the researches on vehicle autonomous driving technologies, the researches on autonomous driving cooperative perception technologies based on vehicle-road cooperation are gradually increasing. In addition to perceiving surrounding environment information through sensors installed on an autonomous vehicle itself, the autonomous vehicle may also acquire state information of traffic participants on a road through sensors fixedly installed on both sides of the road. By fusing perception information from two different sources of vehicle-end and roadside traffic facilities, the long-distance and non-blind-zone cooperative perception of a vehicle can be realized, and the safety of the vehicle during automatic driving can be improved and guaranteed.

At present, an autonomous driving cooperative perception technology based on vehicle-road cooperation is mainly based on an object-level information fusion scheme, that is, the vehicle itself and the roadside facilities will each carry out 3D detection of an object on the road, and then fuse the obtained 3D object information. However, there is a big problem in this scheme, i.e., the time synchronization consistency of vehicle-end information and roadside-end information is poor, which leads to a significant gap between the fused perception performance and the expectation.

Summary of the Invention

The technical problem to be mainly solved by the present invention is that: the time synchronization consistency of vehicle-end information and roadside-end information is poor.

According to a first aspect, an embodiment provides a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation, including:

acquiring vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information;

performing feature extraction on the vehicle-end point cloud information to generate a vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution;

acquiring roadside-end point cloud feature data, the roadside-end point cloud feature data being determined according to roadside-end point cloud information stored according to the timestamp;

decompressing the roadside-end point cloud feature data, and mapping the decompressed roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution by using feature space correction; and
fusing the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution with the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road cooperative perception.

In one embodiment, the performing feature extraction on the roadside-end information to generate a vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution includes:

performing feature extraction on the roadside-end information by using a point cloud object detection network to generate the vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution of $6C*0.5H*0.5W$, wherein C is a number of channels; H is a height; and W is a width.

In one embodiment, the roadside-end point cloud feature data being determined

according to roadside-end point cloud information stored according to the timestamp includes:

performing feature extraction on the roadside-end point cloud feature data through the roadside-end point cloud information stored according to the timestamp, followed by high-dimensional feature extraction and dimension-reducing compression to obtain roadside-end point cloud feature data of $2C*0.5H*0.5W$, wherein C is the number of channels; H is the height; and W is the width.

In one embodiment, the decompressing the roadside-end point cloud feature data includes:

decompressing the roadside-end point cloud feature data by using a set convolutional kernel of a set channel to generate the decompressed roadside-end point cloud feature data of $6C*0.5H*0.5W$, wherein C is the number of channels; H is the height; and W is the width.

In one embodiment, the mapping the dimension-increased roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution by using feature space correction includes:

calculating a transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural network prediction;

determining corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter; and

determining corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution.

In one embodiment, the calculating a transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural network prediction includes:

determining the feature space conversion parameters by using a convolutional neural network through the following formula:

$$A_{\theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix}$$

in which: A_{θ} is the transformation parameter; θ_{11} is a first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is a second conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{13} is a third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is a fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is a fifth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; and θ_{23} is a sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

In one embodiment, the determining corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter includes:

calculating the corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution by using the following formula:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = A_{\theta} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

in which: (x_i^t, y_i^t) are element coordinates of each element in the vehicle-end point cloud feature space distribution; (x_i^s, y_i^s) are corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution; θ_{11} is the first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is the second conversion parameter for the conversion of the roadside point cloud

feature data into the vehicle-end point cloud feature space distribution; θ_{13} is the third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is the fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is the fifth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; and θ_{23} is the sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

In one embodiment, the determining corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution includes:

determining feature values of the roadside-end point cloud feature data under a coordinate position of the vehicle-end point cloud feature space distribution by using a bilinear interpolation method.

According to a second aspect, an embodiment provides a feature-level fusion system for cooperative perception based on vehicle-infrastructure cooperation, including:

a vehicle-end calculation unit, configured to acquire vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information, perform feature extraction on the vehicle-end point cloud information to generate a vehicle-end pseudo-image, and perform high-dimensional feature extraction on the vehicle-end pseudo-image to generate vehicle-end point cloud feature space distribution; and

a roadside-end calculation unit, configured to acquire roadside-end point cloud information, and store the roadside-end point cloud information according to the timestamp corresponding to the vehicle-end point cloud information to generate roadside-end information, wherein

the roadside-end calculation unit is further configured to perform feature extraction on the roadside-end information to generate a roadside-end pseudo-image, perform high-dimensional feature extraction on the roadside-end pseudo-image, and then perform dimension-reducing compression to generate roadside-end point cloud feature data;

the vehicle-end calculation unit is configured to acquire the roadside-end point cloud feature data, decompress the roadside-end point cloud feature data, and map the decompressed roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution by using feature space correction; and

the vehicle-end calculation unit is configured to fuse the vehicle-end point cloud feature space distribution with the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road cooperation.

In a third aspect, an embodiment provides a computer-readable storage medium, configured to store a program therein, wherein the program can be executed by a processor to implement the above method.

According to the feature-level fusion method and system for cooperative perception based on vehicle-infrastructure cooperation and the computer-readable storage medium in the above embodiments, the vehicle-end point cloud information and the timestamp corresponding to the vehicle-end point cloud information are acquired, and the roadside-end point cloud information is then acquired. The roadside-end point cloud information is stored according to the corresponding timestamp to generate the roadside-end information, and the roadside-end point cloud information is subjected to feature extraction and compressed to generate the roadside-end point cloud feature data, while the vehicle-end point cloud information is subjected to high-dimensional feature extraction to generate the vehicle-end point cloud feature space distribution. The roadside-end point cloud feature data is decompressed, and the decompressed roadside-end point cloud feature data is mapped to and fused with the vehicle-end point cloud feature space distribution, and the fused feature information is processed, thereby realizing vehicle-road cooperative perception. The present application ensures the spatiotemporal synchronization of the vehicle-end point cloud information and the roadside-end point cloud information, eliminates a feature space difference caused by the inconsistency between a roadside end and a vehicle end at a data collection moment, and realizes the alignment of point cloud feature spaces of the vehicle end and the roadside end.

Brief Description of the Drawings

FIG.1 is a technical flowchart of a vehicle-road cooperative perception scheme based

on object-level fusion;

FIG. 2 is a flowchart I of a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation in an embodiment;

FIG. 3 is a technical flowchart of a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation in an embodiment;

FIG. 4 is a schematic diagram of a 1*1 convolutional neural kernel vs. network dimension increase in an embodiment;

FIG. 5 is a schematic diagram of a 1*1 convolutional neural kernel vs. network dimension reduction in an embodiment;

FIG. 6 is a flowchart II of a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation in an embodiment;

FIG. 7 is a schematic diagram of bilinear interpolation in an embodiment;

FIG. 8 is a schematic diagram of a final detection result obtained after the fusion of point cloud features in an embodiment; and

FIG. 9 is a schematic diagram of a feature-level fusion system for cooperative perception based on vehicle-infrastructure cooperation in an embodiment.

Detailed Description of the Invention

The present invention will be further described in detail below in conjunction with specific implementations and accompanying drawings. Associated similar component symbols are adopted for similar elements in different implementations. In the following implementations, many details are described in order to make the present application better understood. However, those skilled in the art can easily recognize that some of these features can be omitted or replaced by other components, materials, or methods in different cases. In some cases, some operations related to the present application are not shown or described in the specification, in order to avoid the core part of the present application being overwhelmed by excessive descriptions. However, it is not necessary for those skilled in the art to describe these relevant operations in detail, and they can fully understand the relevant operations according to the description in the specification and general technical knowledge in the art.

In addition, the features, operations, or characteristics described in the specification

may be combined in any appropriate manner to form various implementations. Meanwhile, the steps or actions in the method description may also be sequentially reversed or adjusted in a manner that is obvious to those skilled in the art. Therefore, various sequences in the specification and accompanying drawings are only intended to clearly describe a particular embodiment and are not meant to be a required order, unless otherwise stated that one of the sequences must be followed.

The serial numbers assigned to the components herein, such as "first" and "second", are only used to distinguish the objects being described, but do not have any sequential or technical meaning. Unless otherwise specified, the terms "connected" and "coupled" in the present application include both direct and indirect connections (coupling).

For vehicle-road cooperative perception, an object-level fusion has been disclosed (the object-level fusion refers to a fusion of prediction results of various modal models to make a final decision). A vehicle-road cooperative perception scheme has a main idea in that LIDARs at a roadside end and a vehicle end first carry out 3D detection and tracking on objects around a vehicle, and then fuse two types of object information to achieve the final object detection. This scheme includes the following steps 1-3.

Step 1: roadside-end perception—laser point cloud 3D object detection.

The roadside-end perception means that by installing a high-definition camera, LIDAR and other sensors, as well as a calculation unit and other devices on traffic poles on both sides of a road, roadside facilities can detect all traffic participants (e.g., pedestrians and vehicles) on the road in real time.

As shown in an area at the upper part of a technical flowchart of a vehicle-road cooperative perception scheme based on object-level fusion in FIG. 1, a roadside-end calculation unit acquires original point cloud data first through high-voltage harness LIDAR. Next, the point cloud data is subjected to feature extraction to obtain a corresponding feature matrix. Then, the obtained feature matrix is detected, the specific operations mainly include regression and classification, and 3D bounding boxes, orientations and object categories of all objects are obtained respectively. Next, moving objects are continuously tracked according to object motion characteristics, and then the velocity information of the moving objects is obtained. Finally, the obtained detection information (including positions, orientations, categories, speeds,

etc.) of all objects is transmitted from the roadside calculation unit to a communication unit RSU (i.e., a road side unit) and broadcast to the outside.

Step 2: vehicle-end perception—laser point cloud 3D object detection.

The vehicle-end perception refers to the detection of traffic participants (e.g., pedestrians) around the vehicle itself by the existing cameras, LIDARs and other sensor devices on the vehicle. However, the vehicle-end perception is easily blocked by other surrounding obstacles, and easily forms a certain range of perception blind zones.

As shown in an area at the lower part of the technical flowchart of a vehicle-road cooperative perception scheme based on object-level fusion in FIG. 1, a vehicle-end laser point cloud 3D object detection process is basically the same as that of roadside-end perception. The vehicle-end calculation unit needs to perform four basic steps of original point cloud data information acquisition, feature extraction, 3D object detection and tracking to obtain 3D bounding boxes, orientations, object categories and speed information of objects around the vehicle. However, the object information obtained by the vehicle-end calculation unit does not need to be transmitted externally.

Step 3: vehicle-end perception—3D object information fusion of the vehicle end and the roadside end.

As shown in an area at the lower part of the technical flowchart of the vehicle-road cooperative perception scheme based on object-level fusion in FIG. 1, a vehicle-end communication unit OBU (i.e., an on board unit) may continuously acquire object information data from the roadside-end communication unit RSU, and transmit the data to the vehicle-end calculation unit. The vehicle-end calculation unit then acquires 3D object information of both the roadside end and the vehicle end at the same time. Finally, the 3D object information of both the roadside end and the vehicle end is fused to obtain final object information, which is outputted as a perception result.

Multi-sensor information fusion is currently based on vehicle-road cooperation (vehicle-road cooperative perception refers to cooperative perception for a surrounding environment of the vehicle by combining the perception information of the roadside end and the perception information of the vehicle end. For example, the acquisition of information on the surrounding environment of the vehicle mainly depends on the vehicle itself, and the acquisition of blind zones around the vehicle

and long-distance environmental information mainly depends on the roadside-end facilities). The main implementation methods for cooperation perception of autonomous driving are shown. Due to a simple data link and a low bandwidth requirement in V2X communication (V2X communication means that the communication unit device RSU configured to send data is installed at the roadside end, and the communication unit device OBU configured to receive data is installed at the autonomous vehicle end. The perception information of the roadside end may be broadcast and transmitted outward by the RSU device, and the OBU device of the autonomous vehicle end may continuously receive information sent from the roadside end, and further transmit the information to a controller of the vehicle itself), the current multi-sensor information fusion method for vehicle-road cooperative perception mainly adopts object-level information fusion, and the specific technical process is as described in the above-mentioned vehicle-road cooperative perception scheme based on object-level fusion. However, this scheme currently has the biggest disadvantage, i.e., the time synchronization consistency of vehicle-end information and roadside-end information is poor, which leads to a significant gap between the fused perception performance and the expectation. The disadvantages of the vehicle-road cooperative perception scheme based on object-level fusion are mainly manifested in two aspects.

In the first aspect, the vehicle-road cooperative perception scheme based on object-level fusion requires that there is no large time interval between the information from two different sources of the roadside end and the vehicle end, otherwise the fused performance will be greatly deteriorated. However, due to the fact that it takes a certain amount of time for a roadside sensing device to detect a 3D object, and because of the time consumption of the data transmission from the roadside-end RSU to the vehicle-end OBU, there is always a large time interval between the perception information of the roadside end received by the autonomous vehicle and perception information detected by a vehicle-end sensor. If this time interval exceeds a certain threshold, data information of the roadside end will no

longer be meaningful to the autonomous vehicle.

In the second aspect, the acquisition of the original environmental information by the vehicle end and the roadside end is not triggered synchronously, and original point cloud data of LIDARs of the vehicle end and the roadside end are not acquired at the same time, which leads to the failure of the same object to align and unify in a space of the vehicle end and the roadside end, always causing some position differences. This also leads to a reason why the performance of sensor information fusion between the roadside end and the vehicle end is often worse than that of the vehicle end alone.

Therefore, the main purposes of the present application are to improve the time synchronization consistency of the vehicle-end information and the roadside-end information, reduce a time difference between the perception information of the vehicle end and the roadside end, and realize the alignment and unification of the original data information in the two spaces of the roadside end and the vehicle end, so as to improve the cooperative perception performance of vehicle-road cooperative automatic driving.

Referring to FIG. 2, some embodiments of the present application provide a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation. This method specifically includes the following steps S100-S400.

Step S100: acquire vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information.

FIG. 3 is a technical flowchart of a feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation. In some embodiments, when the vehicle-end calculation unit acquires each frame of vehicle-end point cloud information, the autonomous vehicle may transmit positioning information of the vehicle itself and a timestamp of the current moment to the vehicle-end communication unit OBU through its internal high-precision positioning system, where the positioning information includes longitude and latitude coordinates and information on attitudes and orientation angles of a vehicle body. The vehicle-end communication unit OBU broadcasts the vehicle-end point cloud information, vehicle

positioning information, and corresponding timestamps to all roadside ends, so as to ensure that roadside-end facilities near the autonomous vehicle can receive the vehicle-end point cloud information, the vehicle positioning information, and the corresponding timestamps. In some embodiments, the vehicle-end point cloud information refers to point cloud information of obstacles acquired by the vehicle end. Step S200: perform feature extraction on the vehicle-end point cloud information to generate a vehicle-end pseudo-image, and performing high-dimensional feature extraction on the vehicle-end pseudo-image to generate vehicle-end point cloud feature space distribution.

In some embodiments, the roadside-end information is subjected to feature extraction by using a point cloud object detection network to generate the vehicle-end pseudo-image, and the vehicle-end pseudo-image is subjected to high-dimensional feature extraction to generate vehicle-end point cloud feature space distribution of $6C*0.5H*0.5W$, wherein C is a number of channels of the pseudo-image; H is a height of the pseudo-image; and W is a width of the pseudo-image.

In some embodiments, the point cloud object detection network performs feature extraction on the roadside-end information by using a PointPillars network. The PointPillars network is a point cloud object detection network, which mainly includes three parts in architecture: Pillar Feature Net, Backbone and Detection Head, wherein Pillar Feature Net is configured to convert point cloud data into a pseudo-image, Backbone is configured to perform feature extraction by using a 2D convolutional neural network, and Detection Head is configured to perform 3D detection on the extracted feature data to obtain object information. The PointPillars network is widely used in the industry because of its small number of parameters and fast network calculation speed.

In step S200, in the course of feature extraction on the vehicle-end point cloud information, the vehicle-end calculation unit first obtains a three-dimensional tensor (i.e., a pseudo-image) with a size of $1C*1H*1W$ from the vehicle-end point cloud information through the Pillar Feature Net subnetwork. Then, the Backbone

subnetwork is used to perform high-dimensional extraction on the pseudo-image, and the vehicle-end point cloud feature space distribution with a feature tensor size of $6C*0.5H*0.5W$ is finally outputted. C is a number of channels of the pseudo-image; H is a height of the pseudo-image; and W is a width of the pseudo-image.

Step S300: acquire the roadside-end point cloud feature data.

In some embodiments, the roadside-end point cloud feature data is determined by performing high-dimensional feature extraction on a pseudo-image correspondingly generated by the roadside-end point cloud information stored according to the timestamp corresponding to the vehicle-end point cloud information, and then performing dimension-reducing compression. The roadside-end point cloud information is subjected to high-dimension extraction on the pseudo-image correspondingly generated by the point cloud object detection network to obtain a roadside-end point cloud feature space of $6C*0.5H*0.5W$. The roadside-end point cloud feature space is subjected to dimension-reducing compression by using a first set convolutional kernel of a first set channel to determine roadside-end point cloud feature data of $2C*0.5H*0.5W$, wherein C is the number of channels; H is the height; and W is the width.

The roadside-end calculation unit is configured to acquire roadside-end point cloud information in real time by using well-installed high-voltage harness LIDAR, and cache the roadside-end point cloud information according to the timestamp corresponding to the vehicle-end point cloud information to generate roadside-end information. In the case of such caching, the roadside-end information may contain both the vehicle-end point cloud information and the timestamp information corresponding to the vehicle-end point cloud information. According to the timestamp corresponding to the acquired vehicle-end point cloud information, the closest frame of the roadside-end information to the vehicle-end point cloud information is captured from the roadside-end information as input data.

The roadside-end point cloud information is stored according to the vehicle-end point cloud information, which ensures that the vehicle-end point cloud information and the

roadside-end point cloud information are basically consistent in the time dimension. Assuming that an operating frequency f of LIDAR is 20 Hz, a maximum original time difference of $\Delta T_{\max}=1/f=50\text{ms}$ between the vehicle-end point cloud information and the roadside-end point cloud information during data acquisition can be guaranteed.

In some embodiments, after the roadside end acquires the positioning information of the vehicle, the roadside-end information is converted from a roadside-end coordinate system to a corresponding autonomous vehicle coordinate system, which can greatly facilitate the fusion of the vehicle-end point cloud information and the roadside-end information.

In some embodiments, in the course of acquiring the roadside-end point cloud feature data in step S300, the roadside-end calculation unit first obtains a three-dimensional tensor (i.e., a pseudo-image) with a size of $1C*1H*1W$ from the roadside-end information through the Pillar Feature Net subnetwork. Then, the Backbone subnetwork is configured to perform high-dimensional extraction on the pseudo-image, and a roadside-end point cloud feature space with a feature tensor size of $6C*0.5H*0.5W$ is finally outputted. C is the number of channels; H is the height; and W is the width.

In some embodiments, in order to reduce a bandwidth occupied by the roadside-end point cloud feature space from the roadside end to the vehicle end, and shorten the transmission time of data from the roadside-end RSU to the vehicle-end OBU, the feature data in the roadside-end point cloud feature space is compressed as necessary. $2*C\ 1*1$ convolutional kernels are used to reduce the dimension of the point cloud feature space, so as to achieve the purpose of data feature compression. Through the $1*1$ convolutional kernels, the feature tensor data in the roadside-end point cloud feature space is changed into the roadside-end point cloud feature data of $2C*0.5H*0.5W$.

Referring to FIG. 4 and FIG. 5, a $1*1$ convolutional kernel actually performs linear combination (information integration) on different channels for each point cloud feature point, and retains an original plane structure of the point cloud feature. In step

S300, the number of 1×1 convolutional kernels is used to regulate the depth of feature output, so as to achieve a function of dimension increase (as shown in FIG. 4) or dimension reduction (as shown in FIG. 5) of feature data. If the number of 1×1 convolutional neural kernels exceeds a number of channels of the original point cloud features, the point cloud feature data may be subjected to dimension increase to achieve feature data decompression. Similarly, if the number of 1×1 convolutional neural kernels exceeds a number of channels of the original point cloud features, the point cloud feature data is subjected to dimension reduction to achieve feature data compression.

After the roadside-end point cloud feature data is acquired, the information is broadcast from the roadside-end RSU to the vehicle-end OBU, and the vehicle-end calculation unit may quickly receive the roadside-end point cloud feature data.

Step S400: decompress the roadside-end point cloud feature data, and map the decompressed roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution by using feature space correction.

In some embodiments, after the vehicle-end calculation unit receives the roadside-end point cloud feature data sent by the roadside end, the data needs to be subjected to necessary data preprocessing. The data preprocessing first includes decompressing the roadside-end point cloud feature data for dimension increase. The decompression of the roadside-end point cloud feature data for dimension increase is mainly used to restore the original feature data information of the roadside end.

In some embodiments, the roadside-end point cloud feature data is subjected to dimension increase by using a second set convolutional kernel of a second set channel to generate the dimension-increased roadside-end point cloud feature data of $6C \times 0.5H \times 0.5W$, wherein C is the number of channels; H is the height; and W is the width.

In some embodiments, the data preprocessing also includes performing feature space correction on the dimension-increased roadside-end point cloud feature data. The feature space correction is mainly used to make necessary corrections to the roadside

point cloud feature data in a vehicle coordinate space, so as to align the roadside point cloud feature data with the vehicle-end point cloud information in the vehicle-end space. Only the preprocessed roadside point cloud feature data can be fused with the vehicle-end point cloud information.

In some embodiments, although two different sources of point cloud data, i.e., the vehicle-end point cloud data and the roadside-end point cloud feature data, are guaranteed to be basically consistent in the time dimension in the step S100 and the step S300, there is always a certain error (<50 ms) between the two types of point cloud data in the time dimension of data acquisition. Due to this time delay, an initial position and attitude of a moving object (e.g., a vehicle) often have a certain offset and rotation error in the two spaces of the vehicle end and the roadside end. Therefore, in order to eliminate this error, the necessary space correction is performed on the roadside-end point cloud feature data. Referring to FIG. 6, the method specifically includes the following steps S401-S403.

Step S401: calculate a transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural network prediction.

In some embodiments, feature space conversion parameters are mainly used to calculate parameters of space transformations. The space transformations from the roadside point cloud feature data to the vehicle-end point cloud feature space distribution mainly include translation and rotation operations. As shown in the following formula, the space transformation parameter A_θ is represented as a matrix of two rows and three columns, and the feature space transformation parameter A_θ is obtained by using a deep learning method of a 2D convolutional neural network. The convolutional neural network module consists of three convolutional layers, one fully connected layer, and one logistic regression layer.

$$A_\theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix}$$

in which: A_θ is a transformation parameter; θ_{11} is a first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is a second conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{13} is a third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is a fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is a fifth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; and θ_{23} is a sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

Step S402: determine corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter.

In some embodiments, according to the feature space transformation parameter A_θ obtained in step S401, by traversing the coordinates of each element in the vehicle-end point cloud feature space distribution, the corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution are calculated by using the following formula, and the position coordinates are stored.

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

in which: (x_i^t, y_i^t) are element coordinates of each element in the vehicle-end point cloud feature space distribution; (x_i^s, y_i^s) are corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space

distribution; θ_{11} is the first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is the second conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{13} is the third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is the fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is the fifth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; and θ_{23} is the sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

Step S403: determine corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution.

In some embodiments, the feature values of the roadside-end point cloud feature data under a coordinate position of the vehicle-end point cloud feature space distribution are determined by using a bilinear interpolation method.

In some embodiments, assuming that the coordinates (x_i^s, y_i^s) corresponding to the roadside-end point cloud feature data calculated in step S402 are all an integer, the corresponding feature values in the roadside-end point cloud feature data may be directly assigned to feature elements corresponding to coordinates (x_i^t, y_i^t) of the vehicle-end point cloud feature space distribution. However, the coordinates (x_i^s, y_i^s) corresponding to the roadside-end point cloud feature data obtained in step S402 are usually decimals. Therefore, the feature values corresponding to coordinates of the vehicle-end point cloud feature space distribution need to be calculated by using the bilinear interpolation method. Referring to FIG. 7, a basic idea of bilinear

interpolation is to estimate a feature value of a point from feature values of four points around this point. The coordinates of a feature point P on X and Y axes are all decimals, and the feature values of the four integer coordinate points Q_{11} , Q_{12} , Q_{21} and Q_{22} around the point P are known. The bilinear interpolation method is used to solve the feature value corresponding to the point P by using the following formula:

$$f(p) = (y_1 - y)(x_2 - x)f(Q_{11}) + (y_2 - y)(x - x_1)f(Q_{21}) + (y - y_1)(x_2 - x)f(Q_{12}) + (y - y_1)(x - x_1)f(Q_{22})$$

in which: P is the feature point; (x, y) are the coordinates of the point P; Q_{11} , Q_{21} , Q_{12} and Q_{22} are four points adjacent to the point P; (x_1, y_1) are coordinates of the point Q_{11} ; (x_1, x_2) are coordinates of the point Q_{12} ; (x_2, y_1) are coordinates of the point Q_{21} ; (x_2, y_2) are coordinates of the point Q_{22} ; and $f(Q_{11})$, $f(Q_{21})$, $f(Q_{12})$ and $f(Q_{22})$ are the feature values corresponding to Q_{11} , Q_{21} , Q_{12} and Q_{22} , respectively.

Step S500: fuse the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution with the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road cooperative perception.

In some embodiments, the vehicle-end point cloud feature space distribution and the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution are cascaded to obtain a new fusion feature. A point cloud feature matrix after the final fusion of the vehicle end and the roadside end has a size of $12C \times 0.5H \times 0.5W$, wherein C is the number of channels; H is the height; W is the width.

In some embodiments, after the vehicle-road fusion is realized, Detection Head is used to perform 3D detection on the fused point cloud features, and centroid coordinates, an outer contour size, as well as object orientation and category information of the 3D object are obtained. FIG. 8 is a schematic diagram of a final detection result obtained after the fusion of point cloud features.

Referring to FIG. 9, some embodiments of the present application provide a feature-level perception fusion system 800 for vehicle-road cooperation. This system includes a vehicle-end calculation unit 810 and a roadside-end calculation unit 820, which will be specifically described below.

The vehicle-end calculation unit 810 is configured to acquire vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information. After performing feature extraction on the vehicle-end point cloud information to generate the vehicle-end pseudo-image, the vehicle-end calculation unit 810 is configured to perform high-dimensional feature extraction on the vehicle-end pseudo-image to generate the vehicle-end point cloud feature space distribution.

The roadside-end calculation unit 820 is configured to acquire roadside-end point cloud information, and store the roadside-end point cloud information according to the timestamp corresponding to the vehicle-end point cloud information to generate roadside-end information. The roadside-end calculation unit 820 is further configured to perform feature extraction on the roadside-end information to generate a roadside-end pseudo-image, perform high-dimensional feature extraction on the roadside-end pseudo-image, and then perform dimension-reducing compression to generate roadside-end point cloud feature data.

The vehicle-end calculation unit 810 is configured to acquire the roadside-end point cloud feature data, decompress the roadside-end point cloud feature data, and map the decompressed roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution by using feature space correction.

The vehicle-end calculation unit 810 is configured to fuse the vehicle-end point cloud feature space distribution with the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road cooperation.

In the present application, a feature-level cooperative perception fusion framework oriented to vehicle-road cooperation is established, in which the PointPillars network with the best real-time performance is used for point cloud feature extraction at both the vehicle end and the roadside end. Firstly, feature-level data fusion is more suitable for the optimization of the entire neural network from the vehicle end to the roadside end. A network of the vehicle end and a network of the roadside end can be trained

and optimized as a whole, which can maximize the advantage of combining the data information from two different sources of the roadside end and the vehicle end, and improve the performance of 3D object detection. Secondly, based on feature-level data fusion, compared with object-level data fusion, the vehicle end will be able to receive the data transmitted by the roadside end within a shorter time to ensure the timeliness of data fusion. In this way, the defect of object-level information fusion that cannot be effectively fused due to high transmission delay is avoided.

In the present application, the 1×1 convolutional kernel is configured to compress and decompress the data of the point cloud features, which can fully reduce the amount of data transmitted from the roadside end to the vehicle end, reduce bandwidth requirements of the roadside-end communication unit RSU and the vehicle-end communication unit OBU, greatly improve a data processing speed, and ensure the real-time performance of algorithm processing. Although the point cloud information from two different sources of the vehicle end and the roadside end keeps basically consistent at an initial moment of data acquisition, there is always a certain error between the two types of point cloud data in the time dimension of data acquisition. In the present application, the convolutional neural network is used to perform feature space correction on the dimension-increased roadside-end point cloud feature data, and the roadside point cloud feature data is restored to a state at the moment when the vehicle-end point cloud data is acquired. In this way, the space consistency of the two types of data is ensured. According to the present application, there is certain delay robustness when the data from two different sources is fused, which indirectly improves the perception performance of object detection after feature fusion.

Those skilled in the art may understand that all or part of the functions of the various methods in the above implementations can be achieved by means of hardware or by means of computer programs. When all or part of the functions in the above implementations are achieved by means of a computer program, the program may be stored in a computer-readable storage medium. The storage medium may include: a read-only memory, a random memory, a magnetic disk, an optical disc, a hard disk,

etc. This program is executed through a computer to achieve the above functions. For example, if the program is stored in a memory of the device, all or part of the above functions can be achieved when the program is executed in the memory by a processor. In addition, when all or part of the functions in the above implementations are achieved by means of a computer program, the program may also be stored in a storage medium such as a server, another computer, a magnetic disk, an optical disk, a flash disk, or a portable hard disk, or saved into a memory of a local device by downloading or copying, or updated for a version of a system of the local device. When a program in the memory is executed through the processor, all or part of the functions in the above implementations can be achieved.

The above specific examples are used to elaborate the present invention, which is only used to help understand the present invention and is not used to limit the present invention. For those skilled in the art to which the present invention belongs, a number of simple deductions, deformations or replacements may be made according to the idea of the present invention.

Claims

1. A feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation, comprising:

acquiring vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information;

performing feature extraction on the vehicle-end point cloud information to generate a vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution;

acquiring roadside-end point cloud feature data, the roadside-end point cloud feature data being determined according to roadside-end point cloud information stored according to the timestamp;

calculating a transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural network prediction;

determining corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter;

determining corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution; and

fusing the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution with the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road cooperative perception.

2. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the performing

feature extraction on the vehicle-end information to generate a vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution comprises:

performing feature extraction on the vehicle-end information by using a point cloud object detection network to generate the vehicle-end point cloud pseudo-image, and performing high-dimensional feature extraction on the vehicle-end point cloud pseudo-image to generate vehicle-end point cloud feature space distribution of $6C*0.5H*0.5W$, wherein C is a number of channels; H is a height; and W is a width.

3. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the roadside-end point cloud feature data being determined according to roadside-end point cloud information stored according to the timestamp comprises:

performing feature extraction on the roadside-end point cloud feature data through the roadside-end point cloud information stored according to the timestamp, followed by high-dimensional feature extraction and dimension-reducing compression to obtain roadside-end point cloud feature data of $2C*0.5H*0.5W$, wherein C is the number of channels; H is the height; and W is the width.

4. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the decompressing the roadside-end point cloud feature data comprises:

decompressing the roadside-end point cloud feature data by using a set convolutional kernel of a set channel to generate the decompressed roadside-end point cloud feature data of $6C*0.5H*0.5W$, wherein C is the number of channels; H is the height; and W is the width.

5. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the calculating the transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural

network prediction comprises:

obtaining the following formula by using the 2D convolutional neural network prediction to determine the transformation parameter:

$$A_{\theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix}$$

in which: A_{θ} is the transformation parameter; θ_{11} is a first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is a second conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{13} is a third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is a fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is a fifth conversion parameter for the conversion of the roadside-end point cloud feature data into the vehicle point cloud feature space distribution; and θ_{23} is a sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

6. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the determining corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter comprises:

calculating the corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution by using the following formula:

$$\begin{pmatrix} \mathbf{x}_i^s \\ y_i^s \end{pmatrix} = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{33} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

in which: (x_i^t, y_i^t) are element coordinates of each element in the vehicle-end point cloud feature space distribution; (x_i^s, y_i^s) are corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution; θ_{11} is the first conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{12} is the second conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{13} is the third conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{21} is the fourth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; θ_{22} is the fifth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution; and θ_{23} is the sixth conversion parameter for the conversion of the roadside point cloud feature data into the vehicle-end point cloud feature space distribution.

7. The feature-level fusion method for cooperative perception based on vehicle-infrastructure cooperation according to claim 1, wherein the determining corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution comprises: determining feature values of the roadside-end point cloud feature data under a coordinate position of the vehicle-end point cloud feature space distribution by using a bilinear interpolation method.

8. A feature-level fusion system for cooperative perception based on

vehicle-infrastructure cooperation, comprising:

a vehicle-end calculation unit, configured to acquire vehicle-end point cloud information and a timestamp corresponding to the vehicle-end point cloud information, perform feature extraction on the vehicle-end point cloud information to generate a vehicle-end pseudo-image, and perform high-dimensional feature extraction on the vehicle-end pseudo-image to generate vehicle-end point cloud feature space distribution; and

a roadside-end calculation unit, configured to acquire roadside-end point cloud information, and store the roadside-end point cloud information according to the timestamp corresponding to the vehicle-end point cloud information to generate roadside-end information, wherein

the roadside-end calculation unit is further configured to perform feature extraction on the roadside-end information to generate a roadside-end pseudo-image, perform high-dimensional feature extraction on the roadside-end pseudo-image, and then perform dimension-reducing compression to generate roadside-end point cloud feature data;

the vehicle-end calculation unit is configured to acquire the roadside-end point cloud feature data, and decompress the roadside-end point cloud feature data; calculate a transformation parameter of the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution by using 2D convolutional neural network prediction; determine corresponding position coordinates of the roadside-end point cloud feature data in the vehicle-end point cloud feature space distribution according to the transformation parameter; and determine corresponding feature values from the roadside-end point cloud feature data according to the position coordinates, so as to complete the mapping of the roadside-end point cloud feature data to the vehicle-end point cloud feature space distribution; and

the vehicle-end calculation unit is configured to fuse the vehicle-end point cloud feature space distribution with the roadside-end point cloud feature data mapped to the vehicle-end point cloud feature space distribution, thereby realizing vehicle-road

cooperative perception.

9. A computer-readable storage medium, configured to store a program therein, wherein the program, when being executed by a processor, implements the method according to any one of claims 1 to 7.