

SPECIFICATION

TITLE

THREE-DIMENSIONAL TARGET DETECTION METHOD AND APPARATUS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The present disclosure claims the benefit of a priority of Chinese Patent Application entitled “*Three-Dimensional Target Detection Method and Apparatus*”, No. 202010258375.3, filed on April 3, 2020, the entire contents of which are incorporated herein by reference.

TECHNICAL FIELD

[0002] The present disclosure relates to the technical field of image detection, and specifically to a three-dimensional target detection method, a driving control method, an apparatus, an electronic device and a storage medium.

BACKGROUND

[0003] Recently, three-dimensional (3D) object detection becomes a very important task in computer vision. For example, 3D object detection can be applied in scenes such as unmanned driving. In a 3D object detection, point cloud data in the field of view can be acquired by a LIDAR, and then the acquired point cloud data can be input to a neural network to obtain three-dimensional pose data of all objects in the field of view.

SUMMARY

[0004] In view of this, the present disclosure at least provides a three-dimensional target detection method, a driving control method, an apparatus, an electronic device and a storage medium.

[0005] In a first aspect, there is provided a three-dimensional target detection method, including:

acquiring a binocular image, which includes a left-eye image and a right-eye image; performing a target detection on the left-eye image and the right-eye image respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image; generating a target disparity map of the object to be detected based on the first area image and the second area image, and calculating point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image, and the second area image; and determining

three-dimensional pose data of the object to be detected based on the point cloud data.

[0006] With the above method, the first area image of the object to be detected is obtained from the left-eye image and the second area image of the object to be detected is obtained from the right-eye image; the target disparity map for each object to be detected is determined based on the first area image and the second area image; and then the point cloud data of each object to be detected is calculated based on the target disparity map. By generating the point cloud data based on the disparity map in this manner, the point cloud data can be acquired without the need of a LIDAR, denser point cloud data can be generated, and then the three-dimensional pose data of the object to be detected can be determined based on the point cloud data generated from the disparity map.

[0007] In a possible implementation, before generating the target disparity map of the object to be detected, the method further includes:

determining an abscissa adjustment value based on the first area image and the second area image; and

adjusting an abscissa value of a pixel in the first area image or an abscissa value of a pixel in the second area image based on the abscissa adjustment value, so that abscissa values of corresponding boundary points in the first area image and in the second area image coincide.

[0008] In the above implementation, by determining the abscissa adjustment value, the abscissa adjustment value can represent a rough disparity value between the first area image and the second area image, and then the disparity value between the first area image and the second area image can be roughly corrected based on the abscissa adjustment value, so that when the target disparity map for each object to be detected is generated based on the adjusted first area image and the adjusted second area image subsequently, the search range of the area image can be reduced, and the amount of calculation in the process of determining the target disparity map can be reduced.

[0009] In a possible implementation, said generating the target disparity map based on the first area image and the second area image includes:

determining a contour image of the object to be detected in the first area image or in the second area image, and generating an area disparity map corresponding to the object to be detected based on the first area image and the second area image; and

capturing the target disparity map from the area disparity map based on the contour image.

[0010] In the above implementation, with the contour image of the object to be detected, the target disparity map of the object to be detected can be captured from the area disparity map, and useful disparity maps related to the target are further retained, so that the point cloud data

of the object to be detected determined based on the target disparity map can better represent the object to be detected, which can improve the accuracy of detection.

[0011] In a possible implementation, said generating the area disparity map corresponding to the object to be detected based on the first area image and the second area image includes:

5 inputting the first area image and the second area image to a disparity estimation network for disparity estimation, to obtain the area disparity map.

[0012] In the above implementation, the area disparity map of the object to be detected is determined by the trained disparity estimation network. The area disparity map of the object to be detected can be determined quickly while the accuracy of the obtained area disparity map can be ensured.

[0013] In a possible implementation, the disparity estimation network is trained by:

 acquiring shape data of different objects to be detected of a same type, a plurality of sets of binocular image samples with annotated three-dimensional pose data of the objects to be detected, and point cloud data corresponding to the binocular image samples, wherein the binocular image sample includes a left-eye image sample and a right-eye image sample;

 determining, based on the shape data of different objects to be detected of the same type, average shape data of the objects to be detected of said type;

 obtaining, for one set of binocular image samples, a target disparity map corresponding to said set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples; and

 training the disparity estimation network by using the binocular image samples and the target disparity map corresponding to the binocular image samples as training samples, to obtain a trained disparity estimation network.

[0014] Here the target disparity map corresponding to each set of binocular image samples can be determined based on the average shape data, the three-dimensional pose data of the objects to be detected included in each set of binocular image samples, and the point cloud data corresponding to each set of binocular image samples. The target disparity map obtained from the above implementation can more accurately represent the disparity information of the object to be detected; further, when the disparity estimation network is trained with training samples composed of binocular image samples and the target disparity map corresponding to each set of binocular image samples, the accuracy of the obtained disparity estimation network is higher.

[0015] In a possible implementation, said obtaining, for one set of binocular image samples, the target disparity map corresponding to said set of binocular image samples based on the

average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples includes:

generating a transformed binocular image sample for the one set of binocular image samples, wherein the transformed binocular image sample matches the average shape data and the three-dimensional pose data, and

generating the target disparity map corresponding to the binocular image samples, based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample.

[0016] In the above implementation, the transformed binocular image sample is generated without changing the three-dimensional pose data of the object to be detected in the binocular image sample, so that the shape data in the transformed binocular image sample matches the average shape data. Since the average shape data can represent a shape characteristic of the object to be detected, and the surface of the object shape has continuity, the target disparity map corresponding to the binocular image samples obtained based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample is more accurate and denser.

[0017] In a possible implementation, said generating the target disparity map corresponding to the binocular image samples based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample includes:

adjusting shape data of the transformed binocular image sample based on the point cloud data corresponding to the binocular image samples, to obtain an adjusted binocular image sample, wherein shape data of the adjusted binocular image sample matches a shape corresponding to the point cloud data;

generating a depth map corresponding to the binocular image samples based on the adjusted binocular image sample; and

generating the target disparity map corresponding to said set of binocular image samples based on the depth map.

[0018] In the above implementation, the shape data of the transformed binocular image sample is adjusted based on the point cloud data corresponding to the binocular image sample, so that the shape corresponding to the adjusted shape data matches a shape corresponding to the shape data contained in the binocular image sample before the transformation, that is, the adjusted shape data matches the real shape of the object to be detected in the binocular pattern sample. Therefore, the depth map corresponding to the binocular image sample that is generated based on the adjusted binocular image sample has a higher accuracy and density, and

the accuracy of the generated target disparity map can be improved.

[0019] In a possible implementation, the three-dimensional pose data includes at least one of following information:

length, height, and width of a three-dimensional detection frame of the object to be detected, coordinate information of a center point of the three-dimensional detection frame, or an included angle between a reference plane of the three-dimensional detection frame and a set reference line.

[0020] In a second aspect, there is provided a driving control method, including:

acquiring a binocular road image collected by a travelling device during travelling;

performing a three-dimensional target detection on the binocular road image by using the three-dimensional target detection method in the first aspect, to obtain three-dimensional pose data of a target object included in the binocular road image; and

controlling the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

[0021] See also the above descriptions on the method for the descriptions on the effect of the following apparatus and electronic device, which will not be repeated here.

[0022] In a third aspect, there is provided a three-dimensional target detection apparatus, including:

a binocular image acquisition module, configured to acquire a binocular image which includes a left-eye image and a right-eye image;

a detection module, configured to perform a target detection on the left-eye image and the right-eye image respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image;

a point cloud data determination module, configured to generate a target disparity map of the object to be detected based on the first area image and the second area image, and calculate point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image, and the second area image; and

a three-dimensional pose data determining module, configured to determine three-dimensional pose data of the object to be detected based on the point cloud data.

[0023] In a possible implementation, the apparatus further includes:

an abscissa adjustment value determining module, configured to determine an abscissa adjustment value based on the first area image and the second area image;

an adjustment module, configured to adjust an abscissa value of a pixel in the first area image or an abscissa value of a pixel in the second area image based on the abscissa adjustment value, so that abscissa values of corresponding boundary points in the first area

image and in the second area image coincide.

[0024] In a possible implementation, in response to the target disparity map being generated based on the first area image and the second area image, the point cloud data determining module is configured to:

5 determine a contour image of the object to be detected in the first area image or in the second area image, and generate an area disparity map corresponding to the object to be detected based on the first area image and the second area image;

 capture the target disparity map from the area disparity map based on the contour image.

10 **[0025]** In a possible implementation, in response to the area disparity map corresponding to the object to be detected being generated based on the first area image and the second area image, the point cloud data determining module is configured to:

 input the first area image and the second area image to a disparity estimation network for disparity estimation, to obtain the area disparity map.

15 **[0026]** In a possible implementation, the apparatus further includes: a network training module;

 the network training module is configured to train the disparity estimation network by the following steps:

 acquiring shape data of different objects to be detected of a same type, a plurality of
20 sets of binocular image samples with annotated three-dimensional pose data of the objects to be detected, and point cloud data corresponding to the binocular image samples; wherein the binocular image sample includes a left-eye image sample and a right-eye image sample;

 determining, based on the shape data of different objects to be detected of the same type, average shape data of the objects to be detected of said type;

25 for one set of binocular image samples, obtaining a target disparity map corresponding to said set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples; and

30 training the disparity estimation network by using the binocular image samples and the target disparity map corresponding to the binocular image samples as training samples, to obtain a trained disparity estimation network.

[0027] In a possible implementation, in response to that, for one set of binocular image samples, a target disparity map corresponding to said set of binocular image samples is
35 obtained based on the average shape data of the objects to be detected, the three-dimensional

pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples, the network training module is configured to:

generate a transformed binocular image sample for the one set of binocular image samples, wherein the transformed binocular image sample matches the average shape data and the three-dimensional pose data; and

generate the target disparity map corresponding to the binocular image samples, based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample.

[0028] In a possible implementation, in response to the target disparity map corresponding to the binocular image samples being generated based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample, the network training module is configured to:

adjust shape data of the transformed binocular image sample based on the point cloud data corresponding to the binocular image samples, to obtain an adjusted binocular image sample, wherein shape data of the adjusted binocular image sample matches a shape corresponding to the point cloud data;

generate a depth map corresponding to the binocular image samples based on the adjusted binocular image sample; and

generate the target disparity map corresponding to said set of binocular image samples based on the depth map.

[0029] In a possible implementation, the three-dimensional pose data includes at least one of the following information:

length, height, and width of a three-dimensional detection frame of the object to be detected, coordinate information of a center point of the three-dimensional detection frame, or an included angle between a reference plane of the three-dimensional detection frame and a set reference line.

[0030] In a fourth aspect, there is provided a driving control apparatus, including:

a road image acquisition module, configured to acquire a binocular road image collected by a travelling device during travelling;

a three-dimensional target detection module, configured to perform a three-dimensional target detection on the binocular road image by using the three-dimensional target detection method in the first aspect, to obtain three-dimensional pose data of a target object included in the binocular road image; and

a control module, configured to control the travelling device based on the

three-dimensional pose data of the target object included in the binocular road image.

[0031] In a fifth aspect, there is provided electronic device, including: a processor, a memory, and a bus, wherein the memory stores machine-readable instructions executable by the processor, when the electronic device runs, the processor communicates with the memory through the bus, and when being executed by the processor, the machine-readable instructions perform steps in the three-dimensional target detection method in the first aspect or any embodiment, or steps in the driving control method in the second aspect or any embodiment.

[0032] In a sixth aspect, there is provided a computer readable storage medium, storing a computer program thereon, wherein when being executed by a processor, the computer program performs steps in the three-dimensional target detection method in the first aspect or any embodiment, or steps in the driving control method in the second aspect or any embodiment.

[0033] In a seventh aspect, there is provided a computer program, wherein the computer program includes computer readable codes, and when the computer readable codes are executed in an electronic device, a processor in the electronic device performs the three-dimensional target detection method in the first aspect or any embodiment, or implements the driving control method in the second aspect or any embodiment.

[0034] In order to make the above objectives, features and advantages of the present disclosure more comprehensible, preferred embodiments are described in detail below in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] In order to explain technical solutions of embodiments of the present disclosure more clearly, the drawings needed in the embodiments will be briefly introduced below. The drawings here are incorporated into the specification and constitute a part of the specification. These drawings show embodiments conforming to the present disclosure, and are used to explain the technical solutions of the present disclosure together with the specification. It should be understood that the following drawings only show some embodiments of the present disclosure, and therefore should not be regarded as the limitation to the scope. For those of ordinary skill in the art, other related drawings can be obtained according to these drawings without creative work.

[0036] FIG. 1 shows a schematic flowchart of a three-dimensional target detection method according to an embodiment of the present disclosure;

[0037] FIG. 2 shows a schematic flowchart of a specific method for generating a target disparity map based on a first area image and a second area image, in the three-dimensional

target detection method according to an embodiment of the present disclosure;

[0038] FIG. 3 shows a schematic flowchart of a specific method for training a disparity estimation network, in the three-dimensional target detection method according to an embodiment of the present disclosure;

5 **[0039]** FIG. 4 shows a schematic flowchart of a specific method for obtaining, for a set of binocular image samples, a target disparity map corresponding to the set of binocular image samples based on average shape data of the object to be detected, three-dimensional pose data of the object to be detected included in the set of binocular image samples, and point cloud data corresponding to the set of binocular image samples, in the three-dimensional target detection
10 method according to an embodiment of the present disclosure;

[0040] FIG. 5 shows a schematic flowchart of a specific method for generating a target disparity map corresponding a binocular image sample based on point cloud data corresponding to the binocular image sample and the transformed binocular image sample, in the three-dimensional target detection method according to an embodiment of the present
15 disclosure;

[0041] FIG. 6 shows a schematic flowchart of a driving control method according to an embodiment of the present disclosure;

[0042] FIG. 7 shows a schematic framework diagram of a three-dimensional target detection apparatus according to an embodiment of the present disclosure;

20 **[0043]** FIG. 8 shows a schematic framework diagram of a driving control apparatus according to an embodiment of the present disclosure;

[0044] FIG. 9 shows a schematic structural diagram of an electronic device 900 according to an embodiment of the present disclosure;

[0045] FIG. 10 shows a schematic structural diagram of another electronic device 1000
25 according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0046] In order to make the objectives, technical solutions, and advantages of the embodiments of the present disclosure clearer, the technical solutions in the embodiments of
30 the present disclosure will be described clearly and completely in conjunction with the drawings in the embodiments of the present disclosure. Obviously, the described embodiments are only a part of the embodiments of the present disclosure, but not all of the embodiments. Generally, the components of the embodiments of the present disclosure described and illustrated in the drawings herein can be arranged and designed in a variety of different
35 configurations. Therefore, the following detailed description of the embodiments of the present

disclosure provided in the drawings is not intended to limit the scope of the claimed present disclosure, but merely represents selected embodiments of the present disclosure. Based on the embodiments of the present disclosure, all other embodiments obtained by those skilled in the art without creative work shall fall within the protection scope of the present disclosure.

5 **[0047]** At present, point cloud data in a scene can be acquired by LIDAR, and then based on the acquired point cloud data, three-dimensional pose data of each object included in the scene can be determined, that is, a three-dimensional detection frame for each object is determined. The embodiments of the present disclosure provide a three-dimensional target detection method, which can reduce the dependence on LIDAR in acquiring the point cloud data of the
10 object, and reduce the cost of acquiring the point cloud data.

[0048] The present disclosure can be applied to scenes such as target detection and target positioning for an image or video. For example, in traffic scenes such as autonomous driving and assisted driving, by acquiring binocular images in real time during the driving, the three-dimensional pose data of target objects such as vehicles and pedestrians on the road can
15 be determined, thereby assisting the driver in driving.

[0049] The three-dimensional target detection method according to embodiments of the present disclosure can be applied on a server or a terminal device with a central processor. The server may be a local server or a cloud server, etc., and the terminal device may be a smart phone, a tablet computer, a smart TV, an on-board computer, etc., which is not limited in the
20 present disclosure.

[0050] In order to facilitate the understanding of the embodiments of the present disclosure, a three-dimensional target detection method disclosed in the embodiments of the present disclosure is first introduced in detail.

[0051] Referring to FIG. 1, which is a schematic flowchart of a three-dimensional target
25 detection method according to an embodiment of the present disclosure, the method includes S101-S104, wherein:

[0052] S101: acquiring a binocular image, which includes a left-eye image and a right-eye image.

[0053] S102: performing target detection on the left-eye image and the right-eye image,
30 respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image.

[0054] S103: generating a target disparity map of the object to be detected based on the first area image and the second area image, and calculating point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image and
35 the second area image.

[0055] S104: determining three-dimensional pose data of the object to be detected based on the point cloud data.

[0056] In the above method, by obtaining the first area image of the object to be detected from the left-eye image and the second area image of the object to be detected from the right-eye image, the target disparity map of the object to be detected is determined based on the first area image and the second area image, and then the point cloud data of the object to be detected is calculated based on the target disparity map. This method of generating the point cloud data based on the disparity map can reduce the dependence on LIDAR in acquiring the point cloud data, can generate denser point cloud data, and then can determine the three-dimensional pose data of the object to be detected based on the point cloud data generated from the disparity map.

[0057] S101-S104 are described below.

[0058] For S101:

[0059] Here, the binocular image can be acquired from a storage device storing binocular images, or the binocular image can be acquired from an imaging device (such as a binocular camera) in real time. Generally, the binocular image may include a left-eye image and a right-eye image.

[0060] For S102:

[0061] In the embodiment of the present disclosure, a first detection frame in the left-eye image and a second detection frame in the right-eye image for each object to be detected from at least one object to be detected can be obtained by target detections by an object detection neural network on the left-eye image and the right-eye image respectively; the first area image of each object to be detected is captured from the left-eye image based on the first detection frame, and the second area image of each object to be detected is captured from the right-eye image based on the second detection frame.

[0062] Exemplarily, when a size of the first detection frame is different from a size of the second detection frame, a larger size can be selected as the size of the corresponding detection frame in capturing the area image. For example, if the first detection frame is of a first size, the second detection frame is of a second size, and the first size is larger than the second size, then the size of the second detection frame is adjusted to the first size, and the first area image and the second area image are obtained based on the first detection frame and the resized second detection frame, wherein the first area image and the second area image have the same size.

[0063] Here, the object detection neural network can be trained by a plurality of sets of detection training samples, wherein each set of detection training samples include a left-eye image sample and a right-eye image sample labeled with an object. Exemplarily, a plurality of

sets of detection training samples are input to the object detection neural network, and the object detection neural network is trained until the object detection neural network meets the set condition.

[0064] For S103:

5 **[0065]** In the embodiment of the present disclosure, the pixel value of a pixel in the disparity map may be the difference between the abscissa value corresponding to the pixel in the first area image and the abscissa value corresponding to the pixel in the second area image. For example, if for a pixel A in the disparity map, the abscissa value corresponding to the pixel A in the first area image is x_1 , and the abscissa value corresponding to the pixel A in the second
10 area image is x_2 , the pixel value of the pixel A in the disparity map is $x_1 - x_2$, wherein x_1 and x_2 may be the abscissa values of the pixel A in the image coordinate system.

[0066] Here, by a back projection formula, the point cloud data of each object to be detected can be calculated based on the target disparity map of each object to be detected, camera internal parameters, the first area image and the second area image.

15 **[0067]** In an alternative embodiment, before generating the target disparity map of the object to be detected, the method further includes:

determining an abscissa adjustment value based on the first area image and the second area image; and

adjusting an abscissa value of a pixel in the first area image or an abscissa value of a
20 pixel in the second area image based on the abscissa adjustment value, so that abscissa values of corresponding boundary points in the first area image and in the second area image coincide.

[0068] The coordinate adjustment value can be determined based on the first area image and the second area image. Here, the abscissa adjustment value can be determined based on the abscissa value of a position point on the first area image and the abscissa value of the position
25 point on the second area image, wherein the position point may be any position point on the area images. For example, the position point may be the vertex, the center point, the midpoint on the boundary line, etc. on the area images. Exemplarily, a first abscissa value of a vertex position at the upper left corner of the first area image and a second abscissa value of a vertex position at the upper left corner of the second area image can be determined, and a difference
30 value between the first abscissa value and the second abscissa value is solved and determined as the abscissa adjustment value. Here, the coordinate values of the pixel on the first area image and the coordinate values of the pixel on the second area image are located in the same image coordinate system.

[0069] Exemplarily, the abscissa adjustment value can also be determined by determining a
35 difference between a first abscissa value of the left boundary line of the first area image and a

second abscissa value of the left boundary line of the second area image. Alternatively, the abscissa adjustment value can be determined by determining a difference between a first abscissa value of the right boundary line of the first area image and a second abscissa value of the right boundary line of the second area image; wherein the abscissa values of arbitrary position points on the left boundary line or on the right boundary line in the same area image are the same.

[0070] Further, the abscissa value of the pixel in the first area image or in the second area image can be adjusted based on the abscissa adjustment value. Exemplarily, if the first abscissa value of the left boundary line in the first area image is 60, and the second abscissa value of the left boundary line in the second area image is 30, it is determined that the abscissa adjustment value is 30, and the abscissa value of the pixel in the first area image can be adjusted based on the abscissa adjustment value of 30, that is, 30 is subtracted from each of abscissa values of multiple pixels in the first area image, so that the abscissa values of the boundary points on the left boundary line in the adjusted first area image coincide with the abscissa values of the boundary points on the left boundary line in the second area image; alternatively, the abscissa values of pixels in the second area image can be adjusted based on the abscissa adjustment value of 30, that is, 30 is added to each of abscissa values of multiple pixels in the second area image, so that the abscissa values of the boundary points on the left boundary line in the adjusted second area image coincide with the abscissa values of the boundary points on the left boundary line in the first area image.

[0071] Here, the area disparity map can be obtained based on the first area image and the second area image after adjusting the abscissa values. For example, the first area image and the second area image where the abscissa values of the corresponding boundary points coincide can be input to the disparity estimation network for disparity estimation, to obtain the area disparity map of the object to be detected.

[0072] In the above embodiment, by determining the abscissa adjustment value, the abscissa adjustment value can represent a rough disparity value between the first area image and the second area image, and then the disparity value between the first area image and the second area image can be roughly corrected based on the abscissa adjustment value, so that when the target disparity map of any object to be detected is generated based on the adjusted first area image and the adjusted second area image, the search range of the area image can be reduced, and the amount of calculation in the process of determining the target disparity map can be reduced.

[0073] In an alternative embodiment, referring to FIG. 2, generating the target disparity map based on the first area image and the second area image may include:

[0074] S201: determining a contour image of the object to be detected in the first area image or in the second area image, and generating an area disparity map corresponding to the object to be detected based on the first area image and the second area image of the object to be detected.

5 **[0075]** S202: capturing the target disparity map from the area disparity map based on the contour image.

[0076] Exemplarily, the contour image of each object to be detected in the first area image or in the second area image can be determined by an object detection neural network. For example, the object detection neural network performs a target detection on the left-eye image to obtain a first detection frame in the left-eye image for each of at least one object to be detected and the contour image of that object to be detected in the first detection frame, i.e., obtain the contour image of the object to be detected in the first area image.

10 **[0077]** Here, after obtaining, based on the first area image and the second area image for each object to be detected, the area disparity map corresponding to that object to be detected, the target disparity map of that object to be detected can be captured from the corresponding area disparity map based on the contour image of the object to be detected. Alternatively, after obtaining, based on the first area image and the second area image where the abscissa values coincide, the area disparity map corresponding to the object to be detected, the target disparity map of the object to be detected can be captured from the corresponding area disparity map based on the contour image of the object to be detected.

15 **[0078]** In the above embodiment, with the contour image of the object to be detected, the target disparity map of the object to be detected can be captured from the area disparity map, and useful disparity maps related to the target are further retained, so that the point cloud data of the object to be detected determined based on the target disparity map can better represent the object to be detected, which can improve the accuracy of detection.

20 **[0079]** In an alternative embodiment, generating the area disparity map corresponding to the object to be detected based on the first area image and the second area image includes: inputting the first area image and the second area image to a disparity estimation network for disparity estimation, to obtain the area disparity map.

25 **[0080]** Exemplarily, the first area image and the second area image of the object to be detected can be input to the disparity estimation network, and the disparity estimation network determines a position of the pixel in the first area image and the corresponding position in the second area image based on a minimum loss amount for the pixel value of each constructed pixel, and then obtains the area disparity map of the object to be detected based on the abscissa value of the position in the first area image for each pixel and the abscissa value of the

corresponding position in the second area image for the corresponding pixel. Alternatively, the first area image and the second area image of the object to be detected, where the abscissa values coincide, can be input to the disparity estimation network to obtain the area disparity map of the object to be detected.

5 **[0081]** In the above embodiment, the area disparity map of the object to be detected is determined by the disparity estimation network. The area disparity map of the object to be detected can be determined quickly while the accuracy of the obtained area disparity map can be ensured.

[0082] In an alternative embodiment, referring to FIG. 3, the disparity estimation network can
10 be trained by the following steps:

[0083] S301: acquiring shape data of different objects to be detected of a same type, a plurality of sets of binocular image samples with annotated three-dimensional pose data of objects to be detected, and point cloud data corresponding to each set of binocular image samples, wherein each set of binocular image samples include left-eye image samples and
15 right-eye image samples.

[0084] S302: determining, based on the shape data of different objects to be detected of the same type, average shape data of the objects to be detected of said type.

[0085] S303: obtaining, for each set of binocular image samples, a target disparity map corresponding to the set of binocular image samples based on the average shape data of the
20 objects to be detected, the three-dimensional pose data of the objects to be detected included in the set of binocular image samples, and the point cloud data corresponding to the set of binocular image samples.

[0086] S304: training the disparity estimation network by using the binocular image samples and the target disparity map corresponding to each set of binocular image samples as training
25 samples, to obtain the trained disparity estimation network.

[0087] S301 is described, taking a case where an object to be detected is a vehicle as an example. Here, the vehicle type may include a mini vehicle type, a light vehicle type, a medium vehicle type, a large vehicle type, etc.; or the vehicle type may also include a saloon car type, a sports car type, an off-road vehicle type, etc. Specifically, the types of objects to be
30 detected can be classified according to actual conditions, and the above is only an exemplary description. Here, different objects to be detected of the same type may be different vehicles of the shape data (where the different vehicles may be vehicles of different brands and/or different models) under the type of light vehicles.

[0088] Here, the shape data of the object to be detected can be represented by a truncated
35 signed distance function. Exemplarily, the object to be detected can be divided into n small

cubes, with each small cube corresponding to a value, the magnitude of which may be the minimum of distance values between the cube and respective surfaces of the object to be detected; n values corresponding to the object to be detected can be obtained, and the n values constitute the truncated signed distance function values for the object to be detected, wherein n is a positive integer. The shape data of each of different objects to be detected of the same type includes n values, and the value of n can be set as required.

[0089] In the embodiment of the present disclosure, the shape data of different objects to be detected of the same type, a plurality of sets of binocular image samples, and the point cloud data corresponding to each set of binocular image samples can be acquired; wherein each set of binocular image samples include left-eye image samples and right-eye image samples, and the left-eye image sample and the right-eye image sample in each set of binocular image samples are labeled with the three-dimensional pose data of each object to be detected. Here, the point cloud data corresponding to each set of binocular image samples is sparse point cloud data. For example, the point cloud data corresponding to each set of binocular image samples may be data estimated by a convolutional neural network model, wherein the convolutional neural network may be a neural network trained based on the training data set. Alternatively, the point cloud data may also be point cloud data obtained by LIDAR.

[0090] S302 is described. The shape data of different objects to be detected of the same type can be averaged to obtain the average shape data of the objects to be detected of this type. Here, the average shape data of the objects to be detected of this type corresponds to an average shape.

[0091] The truncated signed distance function value corresponding to each object to be detected can be averaged to obtain the average shape data of the objects to be detected of this type. Exemplarily, if the objects to be detected of the same type include an object to be detected A, an object to be detected B, and an object to be detected C, the shape data of the object to be detected A is $\{a_1, a_2, \dots, a_n\}$, the shape data of the object to be detected B is $\{b_1, b_2, \dots, b_n\}$, and the shape data of the object to be detected C is $\{c_1, c_2, \dots, c_n\}$, then the average shape data of the objects to be detected of this type can be determined as $\{m_1, m_2, \dots, m_n\}$, wherein the calculation process of $m_1, m_2, \dots, m_n$ may be:

$$m_1 = \frac{(a_1+b_1+c_1)}{3}; m_2 = \frac{(a_2+b_2+c_2)}{3}; m_n = \frac{(a_n+b_n+c_n)}{3}.$$

[0092] S303 is described. For each set of binocular image samples, the type of the object to be detected included in the binocular image samples can be determined, and then the average shape data corresponding to the objects to be detected of this type can be determined; after that, the target disparity map corresponding to the set of binocular image samples can be obtained

based on the average shape data of the objects to be detected of this type, the three-dimensional pose data of the objects to be detected included in the set of binocular image samples, and the point cloud data corresponding to the set of binocular image samples.

[0093] In the above embodiment, the target disparity map corresponding to each set of binocular image samples can be determined based on the average shape data, the three-dimensional pose data of the objects to be detected included in each set of binocular image samples, and the point cloud data corresponding to each set of binocular image samples. The target disparity map obtained based on the above embodiment can more accurately represent the disparity information of the object to be detected; further, when training the disparity estimation network with training samples composed of binocular image samples and the target disparity map corresponding to each set of binocular image samples, the accuracy of the obtained disparity estimation network is higher.

[0094] In an alternative embodiment, referring to FIG. 4, obtaining, for any set of binocular image samples, the target disparity map corresponding to the set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in the set of binocular image samples, and the point cloud data corresponding to the set of binocular image samples may include:

[0095] S401: generating a transformed binocular image sample for the one set of binocular image samples, wherein the transformed binocular image sample matches the average shape data and the three-dimensional pose data.

[0096] In the embodiment of the present disclosure, for each set of binocular image samples, the average shape data corresponding to the objects to be detected in the set of binocular image samples is determined, then the size and the direction of the average shape corresponding to the average shape data are adjusted based on the three-dimensional pose data corresponding to the object to be detected in the binocular image sample, and the transformed binocular image sample is obtained based on the adjusted average shape and the set of binocular image samples. The shape of the object to be detected in the transformed binocular image sample is the average shape, and the three-dimensional pose data of that object to be detected is the same as the three-dimensional pose data before the transformation.

[0097] S402: generating the target disparity map corresponding to the binocular image samples based on the point cloud data corresponding to the binocular image samples and the transformed binocular image sample.

[0098] In the above embodiment, the transformed binocular image sample is generated without changing the three-dimensional pose data of the object to be detected in the binocular image samples, so that the shape data in the transformed binocular image sample matches the

average shape data. Since the average shape data can represent the shape feature of the object to be detected and the surface of the object shape has continuity, the target disparity map corresponding to the binocular image samples, obtained based on the point cloud data corresponding to the binocular image samples and the transformed binocular image sample, is more accurate and denser.

[0099] Exemplarily, as shown in FIG. 5, generating the target disparity map corresponding to the binocular image samples based on the point cloud data corresponding to the binocular image samples and the transformed binocular image sample may include:

[00100] S501: adjusting the shape data of the transformed binocular image sample based on the point cloud data corresponding to the binocular image samples, to obtain an adjusted binocular image sample, wherein shape data of the adjusted binocular image sample matches the shape corresponding to the point cloud data.

[00101] Here, the shape data of the transformed binocular image sample can be adjusted based on the point cloud data corresponding to the binocular image samples, to obtain the adjusted binocular image sample. Specifically, the value of the truncated signed distance function corresponding to a position on the surface of the object to be detected is zero, that is, the target values of the truncated signed distance function of point cloud points corresponding to the point cloud data may be zero. Therefore, the shape data of the transformed binocular image sample can be adjusted based on the real values and the target values of the truncated signed distance function of the point cloud points corresponding to the point cloud data, to obtain the adjusted binocular image sample, so that the shape data of the adjusted binocular image sample matches the shape corresponding to the point cloud data.

[00102] S502: generating a depth map corresponding to the binocular image samples based on the adjusted binocular image sample.

[00103] In the embodiment of the present disclosure, the depth map corresponding to the binocular image samples can be generated based on the adjusted binocular image sample and a renderer.

[00104] S503: generating the target disparity map corresponding to the set of binocular image samples based on the depth map.

[00105] Exemplarily, the target disparity map corresponding to the set of binocular image samples can be generated by the depth map. For example, each pixel in the depth map can be used to generate the target disparity map corresponding to the set of binocular image samples in a manner of a geometric relationship transformation.

[00106] Here, an area disparity map corresponding to the set of binocular image samples can be generated based on the depth map, and then the target disparity map corresponding to the set

of binocular image samples can be generated based on the area disparity map and the contour image of the object to be detected.

[00107] In the above embodiment, the shape data of the transformed binocular image sample is adjusted based on the point cloud data corresponding to the binocular image sample, so that the shape corresponding to the adjusted shape data matches the shape corresponding to the shape data contained in the binocular image sample before the transformation, that is, the adjusted shape data matches the real shape of the object to be detected in the binocular pattern sample. Therefore, the depth map corresponding to the binocular image sample that is generated based on the adjusted binocular image sample has a higher accuracy and density, and the accuracy of the generated target disparity map can be improved.

[00108] Following the description of S303, the description of S304 is continued. After the target disparity map corresponding to each set of binocular image samples is obtained, the disparity estimation network is trained by using the binocular image samples and the target disparity map corresponding to each set of binocular image samples as training samples, until the trained disparity estimation network meets a condition, and the trained disparity estimation network is obtained. The condition may be that the accuracy of the disparity estimation network is larger than a set accuracy threshold, or that the loss amount of the disparity estimation network is smaller than a set loss amount threshold, or the likes.

[00109] For S104:

[00110] Exemplarily, the point cloud data can be input to the three-dimensional detection neural network to determine the three-dimensional pose data of each object to be detected. The structure of the three-dimensional detection neural network can be set according to actual needs, and there is no specific limitation here.

[00111] Exemplarily, the three-dimensional pose data includes at least one of the following information: length, height, and width of a three-dimensional detection frame of an object to be detected, coordinate information of a center point of the three-dimensional detection frame, or an included angle between a reference plane of the three-dimensional detection frame and a set reference line. The reference line and the reference plane for the three-dimensional detection frame can be set according to actual conditions; for example, the reference plane may be a front surface, a back surface, a left-side surface, a right-side surface, etc. of the three-dimensional detection frame; and the reference line may be the abscissa axis of the image coordinate system.

[00112] Based on the same concept, the embodiment of the present disclosure also provides a driving control method. As shown in FIG. 6, which is a schematic flowchart of the driving control method according to the embodiment of the present disclosure, the method includes

S601-S603, wherein:

[00113] S601: acquiring a binocular road image collected by a travelling device during travelling;

[00114] S602: performing a three-dimensional target detection on the binocular road image by using the three-dimensional target detection method provided in the embodiment of the present disclosure, to obtain three-dimensional pose data of a target object included in the binocular road image;

[00115] S603: controlling the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

[00116] Exemplarily, the travelling device may be an autonomous driving vehicle, a vehicle equipped with an Advanced Driving Assistance System (ADAS), a robot, or the likes. The binocular road image may be a binocular image collected by the travelling device in real time during travelling. The target object may be any object and/or any target that may appear on the road. For example, the target object may be an animal, a pedestrian, etc. appearing on the road, or another vehicle on the road.

[00117] When controlling the travelling device, the travelling device can be controlled to accelerate, decelerate, turn, brake, etc.; or voice prompt information can be played to prompt the driver to control the travelling device to accelerate, decelerate, turn, brake, etc.

[00118] Those skilled in the art can understand that in the above method of the specific embodiment, the writing order of steps does not mean a strict execution order making any limitation on the implementation process. The specific execution order of the steps should be determined based on the function and possible inner logic thereof. Without violating logic, different embodiments of the present disclosure can be combined with each other, and the parts described emphatically by different embodiments are different and may be learned by referring to the records of corresponding implementations.

[00119] Based on the same concept, an embodiment of the present disclosure also provides a three-dimensional target detection apparatus. As shown in FIG. 7, which is a schematic framework diagram of the three-dimensional target detection apparatus according to the embodiment of the present disclosure, the apparatus includes a binocular image acquisition module 701, a detection module 702, a point cloud data determination module 703, a three-dimensional pose data determination module 704, an abscissa adjustment value determination module 705, an adjustment module 706, and a network training module 707. Specifically:

[00120] The binocular image acquisition module 701 is configured to acquire a binocular image, which includes a left-eye image and a right-eye image;

[00121] The detection module 702 is configured to perform a target detection on the left-eye image and the right-eye image, respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image;

5 **[00122]** The point cloud data determination module 703 is configured to generate a target disparity map of the object to be detected based on the first area image and the second area image, and calculate point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image and the second area image;

[00123] The three-dimensional pose data determining module 704 is configured to determine
10 three-dimensional pose data of the object to be detected based on the point cloud data.

[00124] In a possible embodiment, the apparatus further includes:

[00125] The abscissa adjustment value determining module 705, configured to determine an abscissa adjustment value based on the first area image and the second area image;

[00126] The adjustment module 706, configured to adjust an abscissa value of a pixel in the
15 first area image or an abscissa value of a pixel in the second area image based on the abscissa adjustment value, so that abscissa values of corresponding boundary points in the first area image and in the second area image coincide.

[00127] In a possible embodiment, in response to the target disparity map being generated based on the first area image and the second area image, the point cloud data determining
20 module 703 is configured to:

determine a contour image of the object to be detected in the first area image or in the second area image, and generate an area disparity map corresponding to the object to be detected based on the first area image and the second area image;

capture the target disparity map from the area disparity map based on the contour
25 image.

[00128] In a possible embodiment, in response to the area disparity map corresponding to the object to be detected being generated based on the first area image and the second area image, the point cloud data determining module 703 is configured to input the first area image and the second area image to a disparity estimation network for disparity estimation, to obtain the area
30 disparity map.

[00129] In a possible embodiment, the apparatus further includes: the network training module 707.

[00130] The network training module 707 is configured to train the disparity estimation network by the following steps:

35 acquiring shape data of different objects to be detected of a same type, a plurality of

sets of binocular image samples with annotated three-dimensional pose data of the objects to be detected, and point cloud data corresponding to each set of binocular image samples; each set of binocular image samples include left-eye image samples and right-eye image samples;

determining, based on the shape data of different objects to be detected of the same type, average shape data of the objects to be detected of said type;

obtaining, for each set of binocular image samples, a target disparity map corresponding to said set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples;

training the disparity estimation network by using the binocular image samples and the target disparity map corresponding to each set of binocular image samples as training samples, to obtain the trained disparity estimation network.

[00131] In a possible embodiment, in response to that, for each set of binocular image samples, a target disparity map corresponding to said set of binocular image samples is obtained based on the average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples, the network training module 707 is configured to:

generate a transformed binocular image sample for each set of binocular image samples, wherein the transformed binocular image sample matches the average shape data and the three-dimensional pose data;

generate the target disparity map corresponding to the binocular image samples, based on the point cloud data corresponding to the binocular image sample and the transformed binocular image sample.

[00132] In a possible embodiment, in response to the target disparity map corresponding to the binocular image samples being generated based on the point cloud data corresponding to the binocular image sample and the transformed binocular image sample, the network training module 707 is configured to:

adjust the shape data of the transformed binocular image sample based on the point cloud data corresponding to the binocular image samples, to obtain an adjusted binocular image sample, wherein shape data of the adjusted binocular image sample matches the shape corresponding to the point cloud data;

generate a depth map corresponding to the binocular image samples based on the adjusted binocular image sample;

generate the target disparity map corresponding to said set of binocular image samples based on the depth map.

[00133] In a possible embodiment, the three-dimensional pose data includes at least one of the following information:

length, height, and width of a three-dimensional detection frame of the object to be detected, coordinate information of a center point of the three-dimensional detection frame, or an included angle between a reference plane of the three-dimensional detection frame and a set reference line.

[00134] The embodiment of the present disclosure also provides a driving control apparatus. As shown in FIG. 8, which is a schematic framework diagram of the driving control apparatus according to the embodiment of the present disclosure, the apparatus includes a road image acquisition module 801, a three-dimensional target detection module 802, and a control module 803. Specifically:

[00135] The road image acquisition module 801 is configured to acquire a binocular road image collected by a travelling device during travelling;

[00136] The three-dimensional target detection module 802 is configured to perform a three-dimensional target detection on the binocular road image by using the three-dimensional target detection method described in the above embodiment, to obtain three-dimensional pose data of a target object included in the binocular road image;

[00137] The control module 803 is configured to control the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

[00138] In some embodiments, the functions or templates contained in the apparatuses provided in the embodiments of the present disclosure can be used to execute the methods described in the above method embodiments. The specific implementations can be learned by referring to the description of the above method embodiments, and will not be described redundantly for the purpose of conciseness.

[00139] Based on the same technical concept, the embodiment of the present disclosure also provides an electronic device. As shown in FIG. 9, which is a schematic structural diagram of an electronic device 900 according to the embodiment of the present disclosure, the electronic device 900 includes a processor 901, a memory 902, and a bus 903; wherein, the memory 902 stores executable instructions and includes an internal storage 9021 and an external memory 9022; the internal storage 9021 here is also called an internal memory, and is configured to temporarily store operational data in the processor 901 and data exchanged with the external memory 9022 such as a hard disk. The processor 901 exchanges data with the external memory 9022 via the internal storage 9021. When the electronic device 900 runs, the processor 901

communicates with the memory 902 through the bus 903, so that the processor 901 executes the following instructions:

acquiring a binocular image, which includes a left-eye image and a right-eye image;

performing a target detection on the left-eye image and the right-eye image,
5 respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image;

generating a target disparity map of the object to be detected based on the first area image and the second area image, and calculating point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image and the
10 second area image; and

determining three-dimensional pose data of the object to be detected based on the point cloud data.

[00140] Based on the same technical concept, the embodiment of the present disclosure also provides another electronic device. As shown in FIG. 10, which is a schematic structural
15 diagram of an electronic device 1000 according to the embodiment of the present disclosure, the electronic device 1000 includes a processor 1001, a memory 1002, and a bus 1003; wherein, the memory 1002 stores executable instructions and includes an internal storage 10021 and an external memory 10022; the internal storage 10021 here is also called an internal memory, and is configured to temporarily store operational data in the processor 1001 and data exchanged
20 with the external memory 10022 such as a hard disk. The processor 1001 exchanges data with the external memory 10022 via the internal storage 10021. When the electronic device 1000 runs, the processor 1001 communicates with the memory 1002 through the bus 1003, so that the processor 1001 executes the following instructions:

acquiring a binocular road image collected by a travelling device during travelling;

performing a three-dimensional target detection on the binocular road image by using
25 the three-dimensional target detection method provided in the above embodiment, to obtain three-dimensional pose data of a target object included in the binocular road image; and

controlling the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

[00141] In addition, the embodiments of the present disclosure also provide a computer readable storage medium on which a computer program is stored, and when the computer program is executed by a processor, the steps of the three-dimensional target detection method described in the above method embodiment are performed.

[00142] The computer program product of the three-dimensional target detection method
35 provided by the embodiments of the present disclosure includes a computer readable storage

medium storing program code, and instructions included in the program code can be used to perform the steps of the three-dimensional target detection method described in the above method embodiment, of which the details can be learned by referring to the above method embodiment and will not be described redundantly here.

5 **[00143]** In addition, the embodiments of the present disclosure also provide another computer readable storage medium on which a computer program is stored, and when the computer program is executed by a processor, the steps of the driving control method described in the above method embodiment are performed.

10 **[00144]** The computer program product of the driving control method provided by the embodiments of the present disclosure includes a computer readable storage medium storing program code, and instructions included in the program code can be used to perform the steps of the driving control method described in the above method embodiment, of which the details can be learned by referring to the above method embodiment and will not be described redundantly here.

15 **[00145]** The embodiments of the present disclosure also provide a computer program which includes computer readable code, and when the computer readable code is executed in an electronic device, a processor in the electronic device performs the three-dimensional target detection method or implements the driving control method.

20 **[00146]** Those skilled in the art can clearly understand that, for the convenience and conciseness of description, the specific working process of the system and apparatus described above can be learned by referring to the corresponding process in the foregoing method embodiment, and will not be described redundantly here. In the several embodiments provided by the present disclosure, it should be understood that the disclosed system, apparatus, and method can be implemented in other ways. The apparatus embodiments described above are
25 merely schematic. For example, the division of the units is only a logical function division, and there may be other division manners in actual implementation. Still for example, multiple units or components can be combined or integrated into another system, or some features can be ignored or not executed. In addition, the displayed or discussed mutual coupling or direct coupling or communication connection may be through some communication interfaces, and
30 indirect coupling or communication connection of apparatuses or units may be in electrical, mechanical or other forms.

[00147] The units described as separate components may be or may not be physically separated, and the components displayed as units may be or may not be physical units, that is, the components can be located in one place, or can be distributed on multiple network units.
35 Some or all of the units can be selected according to actual needs to achieve the objectives of

the solutions of the embodiments.

[00148] In addition, the functional units in the embodiments of the present disclosure can be integrated into one processing unit, or each unit can exist alone physically, or two or more units can be integrated into one unit.

5 **[00149]** If the function is implemented in the form of a software functional unit and sold or used as an independent product, it can be stored in a non-volatile computer readable storage medium executable by a processor. Based on this understanding, the technical solution of the present disclosure essentially, or a part that contributes to the prior art, or a part of the technical solution can be embodied in the form of a software product, and the computer software product
10 is stored in a storage medium, including several instructions for causing a computer device (which may be a personal computer, a server, or a network device, etc.) to execute all or part of steps of the methods described in the embodiments of the present disclosure. The aforementioned storage medium includes various medium that can store program code, such as U disk, mobile hard disk, Read-Only Memory (ROM), Random Access Memory (RAM),
15 magnetic disk or optical disk.

[00150] The above are only specific embodiments of the present disclosure, but the protection scope of the present disclosure is not limited thereto. Within the technical scope disclosed in the present disclosure, any person skilled in the art can easily conceive that changes or substitutions shall be encompassed in the protection scope of this disclosure. Therefore, the
20 protection scope of the present disclosure should be subject to the protection scope of the claims.

CLAIMS

What is claimed is:

1. A three-dimensional target detection method, characterized by comprising:

5 acquiring a binocular image, which includes a left-eye image and a right-eye image;

performing a target detection on the left-eye image and the right-eye image respectively, to obtain a first area image of an object to be detected in the left-eye image and a second area image of the object to be detected in the right-eye image;

10 generating a target disparity map of the object to be detected based on the first area image and the second area image, and calculating point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image, and the second area image; and

determining three-dimensional pose data of the object to be detected based on the point cloud data.

15 2. The method according to claim 1, characterized in that, before generating the target disparity map of the object to be detected, the method further comprises:

determining an abscissa adjustment value based on the first area image and the second area image; and

20 adjusting an abscissa value of a pixel in the first area image or an abscissa value of a pixel in the second area image based on the abscissa adjustment value, so that abscissa values of corresponding boundary points in the first area image and in the second area image coincide.

3. The method according to claim 1 or 2, characterized in that, said generating the target disparity map based on the first area image and the second area image comprises:

25 determining a contour image of the object to be detected in the first area image or in the second area image, and generating an area disparity map corresponding to the object to be detected based on the first area image and the second area image; and

30 capturing the target disparity map from the area disparity map based on the contour image.

4. The method according to claim 3, characterized in that, said generating the area disparity map corresponding to the object to be detected based on the first area image and the second area image comprises:

35 inputting the first area image and the second area image to a disparity estimation network

for disparity estimation, to obtain the area disparity map.

5. The method according to claim 4, characterized in that, the disparity estimation network is trained by:

5 acquiring shape data of different objects to be detected of a same type, a plurality of sets of binocular image samples with annotated three-dimensional pose data of the objects to be detected, and point cloud data corresponding to the binocular image samples, wherein the binocular image sample includes a left-eye image sample and a right-eye image sample;

10 determining, based on the shape data of different objects to be detected of the same type, average shape data of the objects to be detected of said type;

15 obtaining, for one set of binocular image samples, a target disparity map corresponding to said set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples; and

 training the disparity estimation network by using the binocular image samples and the target disparity map corresponding to the binocular image samples as training samples, to obtain a trained disparity estimation network.

20 6. The method according to claim 5, characterized in that, said obtaining, for one set of binocular image samples, the target disparity map corresponding to said set of binocular image samples based on the average shape data of the objects to be detected, the three-dimensional pose data of the objects to be detected included in said set of binocular image samples, and the point cloud data corresponding to said set of binocular image samples comprises:

25 generating a transformed binocular image sample for the one set of binocular image samples, wherein the transformed binocular image sample matches the average shape data and the three-dimensional pose data, and

30 generating the target disparity map corresponding to the binocular image samples, based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample.

7. The method according to claim 6, characterized in that, said generating the target disparity map corresponding to the binocular image samples based on the transformed binocular image sample and the point cloud data corresponding to the binocular image sample comprises:

adjusting shape data of the transformed binocular image sample based on the point cloud data corresponding to the binocular image samples, to obtain an adjusted binocular image sample, wherein shape data of the adjusted binocular image sample matches a shape corresponding to the point cloud data;

5 generating a depth map corresponding to the binocular image samples based on the adjusted binocular image sample; and

 generating the target disparity map corresponding to said set of binocular image samples based on the depth map.

10 8. The method according to claim 1, characterized in that, the three-dimensional pose data includes at least one of following information:

 length, height, and width of a three-dimensional detection frame of the object to be detected, coordinate information of a center point of the three-dimensional detection frame, or an included angle between a reference plane of the three-dimensional detection frame and a set
15 reference line.

 9. A driving control method, characterized by comprising:

 acquiring a binocular road image collected by a travelling device during travelling;

 performing a three-dimensional target detection on the binocular road image by using the
20 three-dimensional target detection method according to any one of claims 1-8, to obtain three-dimensional pose data of a target object included in the binocular road image; and

 controlling the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

25 10. A three-dimensional target detection apparatus, characterized by comprising:

 a binocular image acquisition module, configured to acquire a binocular image which includes a left-eye image and a right-eye image;

 a detection module, configured to perform a target detection on the left-eye image and the right-eye image respectively, to obtain a first area image of an object to be detected in the
30 left-eye image and a second area image of the object to be detected in the right-eye image;

 a point cloud data determination module, configured to generate a target disparity map of the object to be detected based on the first area image and the second area image, and calculate point cloud data of the object to be detected based on the target disparity map of the object to be detected, the first area image, and the second area image; and

35 a three-dimensional pose data determining module, configured to determine

three-dimensional pose data of the object to be detected based on the point cloud data.

11. A driving control apparatus, characterized by comprising:

a road image acquisition module, configured to acquire a binocular road image collected

5 by a travelling device during travelling;

a three-dimensional target detection module, configured to perform a three-dimensional target detection on the binocular road image by using the three-dimensional target detection method according to any one of claims 1-8, to obtain three-dimensional pose data of a target object included in the binocular road image; and

10 a control module, configured to control the travelling device based on the three-dimensional pose data of the target object included in the binocular road image.

12. An electronic device, characterized by comprising: a processor, a memory, and a bus,

wherein the memory stores machine-readable instructions executable by the processor,

15 when the electronic device runs, the processor communicates with the memory through the bus, and when being executed by the processor, the machine-readable instructions perform steps in the three-dimensional target detection method according to any one of claims 1-8 or steps in the driving control method according to claim 9.

20 13. A computer readable storage medium, characterized by storing a computer program thereon, wherein when being executed by a processor, the computer program performs steps in the three-dimensional target detection method according to any one of claims 1-8 or steps in the driving control method according to claim 9.

25 14. A computer program, characterized in that, the computer program includes computer readable codes, and when the computer readable codes are executed in an electronic device, a processor in the electronic device performs the three-dimensional target detection method according to any one of claims 1-8 or implements the driving control method according to claim 9.