



(12) 发明专利

(10) 授权公告号 CN 110060230 B

(45) 授权公告日 2021.11.26

(21) 申请号 201910049152.3

(22) 申请日 2019.01.18

(65) 同一申请的已公布的文献号
申请公布号 CN 110060230 A

(43) 申请公布日 2019.07.26

(73) 专利权人 商汤集团有限公司
地址 中国香港沙田科学园科技大道东12号
海滨大楼1座2楼

(72) 发明人 潘柏文 亚历克斯 安东尼安
奥德·奥利瓦 周博磊

(74) 专利代理机构 北京思源智汇知识产权代理
有限公司 11657
代理人 毛丽琴

(51) Int. Cl.
G06T 7/00 (2017.01)
G06T 7/10 (2017.01)
G06N 3/04 (2006.01)

(56) 对比文件

CN 105894549 A, 2016.08.24
CN 105894549 A, 2016.08.24
CN 108876707 A, 2018.11.23
CN 108571974 A, 2018.09.25
CN 109191515 A, 2019.01.11
CN 108389251 A, 2018.08.10
CN 105574545 A, 2016.05.11
CN 103617606 A, 2014.03.05
CN 108985194 A, 2018.12.11
CN 107578022 A, 2018.01.12
Krishna Regmi and Ali Borji. Cross-View Image Synthesis using Conditional GANs.《2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition》.2018,

徐谦 等. 基于深度学习图像语义分割的机器人环境感知.《吉林大学学报(工学版)》.2019,

审查员 白立鹏

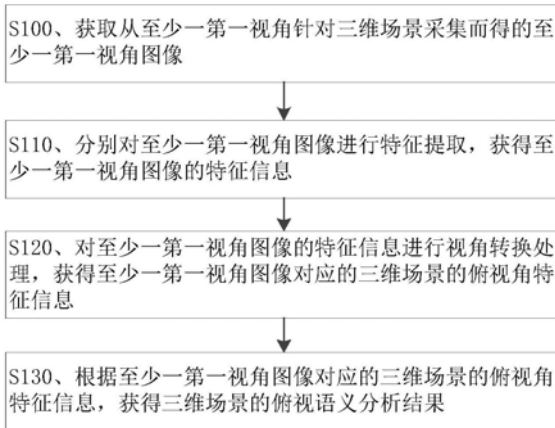
权利要求书7页 说明书22页 附图6页

(54) 发明名称

三维场景分析方法、装置、介质及设备

(57) 摘要

本申请的实施方式公开了一种三维场景分析方法、装置、电子设备、计算机可读存储介质以及计算机程序, 其中的三维场景分析方法包括: 获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像, 所述第一视角与俯视角成一定角度; 分别对所述至少一第一视角图像进行特征提取, 获得至少一第一视角图像的特征信息; 对所述至少一第一视角图像的特征信息进行视角转换处理, 获得至少一第一视角图像对应的三维场景的俯视角特征信息; 根据所述至少一第一视角图像对应的三维场景的俯视角特征信息, 获得所述三维场景的俯视角语义分析结果。



CN 110060230 B

1. 一种三维场景分析方法,其特征在于,包括:

获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像,所述第一视角与俯视角成一定角度;

分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;

对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息,其中,所述第一视角图像的特征信息包括:第一视角图像的特征图,所述第一视角图像对应的三维场景的俯视角特征信息包括:第一视角图像对应的三维场景的俯视角特征图;

根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果。

2. 根据权利要求1所述的方法,其特征在于,所述至少一第一视角包括:多个第一视角;所述多个第一视角中不同第一视角对应所述三维场景的采集区域至少部分不同。

3. 根据权利要求1所述的方法,其特征在于:

多个第一视角图像可形成所述三维场景的非全景图像;或者

多个第一视角图像可形成所述三维场景的全景图像。

4. 根据权利要求1所述的方法,其特征在于,所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像、从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图、以及从第一视角针对三维场景的至少局部区域采集而得的深度图中的至少一个。

5. 根据权利要求4所述的方法,其特征在于,在所述第一视角图像包括从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图的情况下,所述获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像包括:

对从第一视角针对三维场景的至少局部区域采集而得的二维图像进行语义分割处理;

根据所述语义分割处理的结果,获取所述二维图像的语义图。

6. 根据权利要求5所述的方法,其特征在于,所述根据所述语义分割处理的结果,获得所述二维图像的语义图,包括:

基于所述语义分割处理的结果,生成所述二维图像的语义图;或者

基于所述语义分割处理的结果,生成待处理语义图,其中,所述待处理语义图采用第一语义表示,所述第一语义对应第一场景域预定的语义;

根据预设的第一语义与第二语义的映射关系,将所述待处理语义图从第一语义表示转换为第二语义表示,所述转换后的语义图被作为所述二维图像的语义图;

其中,所述第二语义对应第二场景域预定的语义。

7. 根据权利要求1至6中任一项所述的方法,其特征在于,

所述分别对三维场景的至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息,包括:分别对所述三维场景的至少一第一视角图像进行编码处理,根据编码处理的结果获取至少一第一视角图像的特征信息;

所述根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:对所述至少一第一视角图像对应的三维场景的俯视角特

征信息进行解码处理,根据解码处理的结果获取所述三维场景的俯视语义分析结果。

8. 根据权利要求1至6中任一项所述的方法,其特征在于:

所述对所述至少一第一视角图像的特征信息进行视角转换处理,以获得至少一第一视角图像对应的三维场景的俯视角特征信息,包括:

将至少一第一视角图像的特征图转换为至少一第一视角图像的特征向量;

将所述至少一第一视角图像的特征向量分别转换为至少一第一视角图像对应的三维场景的俯视角特征向量;

将至少一第一视角图像的俯视角特征向量分别转换为特征图,获得至少一第一视角图像对应的三维场景的俯视角特征图。

9. 根据权利要求8所述的方法,其特征在于,所述第一视角图像对应的三维场景的俯视角特征图与所述第一视角图像的特征图的大小相同。

10. 根据权利要求7所述的方法,其特征在于,所述俯视语义分析结果包括:语义俯视图;

所述根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:

对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

11. 根据权利要求10所述的方法,其特征在于,在所述至少一第一视角图像为多个第一视角图像的情况下,所述对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图,包括:

将多个第一视角图像各自对应的三维场景的俯视角特征信息叠加,获得三维场景的综合俯视角特征信息;

对所述综合俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

12. 根据权利要求1至6中任一项所述的方法,其特征在于,所述分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息,对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息,根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:

将所述至少一第一视角图像输入神经网络系统;

由所述神经网络系统对所述至少一第一视角图像进行特征提取,对所述特征提取获得的至少一第一视角图像的特征信息进行视角转换处理,并根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息,获得并输出所述三维场景的俯视语义分析结果。

13. 根据权利要求12所述的方法,其特征在于,所述神经网络系统是利用从第一视角针对属于第二场景域的三维场景的至少局部区域采集而得的二维图像的语义图样本训练获得的。

14. 根据权利要求12所述的方法,其特征在于,所述神经网络系统包括:编码器、视角转换器以及解码器;

经所述编码器对所述至少一第一视角图像进行编码处理,并输出编码处理获得的至少一第一视角图像的特征信息;

经所述视角转换器对所述至少一第一视角图像的特征信息进行视角转换处理,并输出至少一第一视角图像对应的三维场景的俯视角特征信息;

经所述解码器根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息,获得并输出所述三维场景的俯视语义分析结果。

15. 根据权利要求14所述的方法,其特征在于,所述视角转换器包括:至少一个多层感知机;

在所述至少一第一视角图像为多个第一视角图像的情况下,所述经所述视角转换器对所述至少一第一视角图像的特征信息进行视角转换处理,包括:

经所述视角转换器中的多个多层感知机分别对其对应的第一视角图像的特征信息进行视角转换处理;不同的多层感知机对应不同的第一视角图像。

16. 根据权利要求14所述的方法,其特征在于,所述神经网络系统的训练过程包括:

将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统中的编码器;

经由所述编码器进行特征提取,以获得至少一第一视角图像样本的特征信息,由所述神经网络系统中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理,获得至少一第一视角图像样本对应的三维场景的俯视角特征信息,并由所述神经网络系统中的解码器对所述至少一第一视角图像样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析;

根据所述俯视语义分析的结果与所述三维场景的语义俯视图样本之间的损失,对所述神经网络系统中的网络参数进行调整。

17. 根据权利要求6所述的方法,其特征在于:

所述第一场景域包括:真实三维场景;

所述第二场景域包括:虚拟三维场景。

18. 一种基于三维场景分析的控制方法,其特征在于,所述方法包括:

基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像;

采用如权利要求1-17任一所述的三维场景分析方法,获得所述三维场景的俯视语义分析结果;

根据所述俯视语义分析结果,对所述智能设备进行控制。

19. 根据权利要求18所述的方法,其特征在于,所述根据所述俯视语义分析结果,对所述智能设备进行控制,包括下述至少之一:

根据所述俯视语义分析结果,进行路径规划,并根据所述路径规划的结果,对所述智能设备进行控制;

根据所述俯视语义分析结果,进行导航规划,并根据所述导航规划的结果,对所述智能设备进行控制;

根据所述俯视语义分析结果,进行避障规划,并根据所述避障规划的结果,对所述智能设备进行控制;

根据所述俯视语义分析结果进行所述智能设备的泊车控制;

根据所述俯视语义分析结果,进行目标抓取规划,并根据所述目标抓取规划的结果,对所述智能设备进行控制。

20. 一种三维场景分析装置,其特征在于,包括:

获取图像模块,用于获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像,所述第一视角与俯视角成一定角度;

特征提取模块,用于分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;

视角转换模块,用于对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息,其中,所述第一视角图像的特征信息包括:第一视角图像的特征图,所述第一视角图像对应的三维场景的俯视角特征信息包括:第一视角图像对应的三维场景的俯视角特征图;

俯视语义分析模块,用于根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果。

21. 根据权利要求20所述的装置,其特征在于,所述至少一第一视角包括多个第一视角;所述多个第一视角中不同第一视角对应所述三维场景的采集区域至少部分不同。

22. 根据权利要求20所述的装置,其特征在于:

多个第一视角图像可形成所述三维场景的非全景图像;或者

多个第一视角图像可形成所述三维场景的全景图像。

23. 根据权利要求20所述的装置,其特征在于,所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像、从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图、以及从第一视角针对三维场景的至少局部区域采集而得的深度图中的至少一个。

24. 根据权利要求23所述的装置,其特征在于,在所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图的情况下,所述获取图像模块包括:

第一子模块,用于对从第一视角针对三维场景的至少局部区域采集而得的二维图像进行语义分割处理;

第二子模块,用于根据所述语义分割处理的结果,获取所述二维图像的语义图。

25. 根据权利要求24所述的装置,其特征在于,所述第二子模块进一步用于:

基于所述语义分割处理的结果,生成所述二维图像的语义图;或者

基于所述语义分割处理的结果,生成待处理语义图,其中,所述待处理语义图采用第一语义表示,所述第一语义对应第一场景域预定的语义;

根据预设的第一语义与第二语义的映射关系,将所述待处理语义图从第一语义表示转换为第二语义表示,所述转换后的语义图被作为所述二维图像的语义图;

其中,所述第二语义对应第二场景域预定的语义。

26. 根据权利要求20至25中任一项所述的装置,其特征在于,所述特征提取模块进一步用于:

分别对所述三维场景的至少一第一视角图像进行编码处理,根据编码处理的结果获取至少一第一视角图像的特征信息;

所述俯视语义分析模块进一步用于：

对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理，根据解码处理的结果获取所述三维场景的俯视语义分析结果。

27. 根据权利要求20至25中任一项所述的装置，其特征在于：

所述视角转换模块进一步用于：

将至少一第一视角图像的特征图转换为至少一第一视角图像的特征向量；

将所述至少一第一视角图像的特征向量分别转换为至少一第一视角图像对应的三维场景的俯视角特征向量；

将至少一第一视角图像的特征向量分别转换为特征图，获得至少一第一视角图像对应的三维场景的俯视角特征图。

28. 根据权利要求27所述的装置，其特征在于，所述第一视角图像对应的三维场景的俯视角特征图与所述第一视角图像的特征图的大小相同。

29. 根据权利要求26所述的装置，其特征在于，所述俯视语义分析结果包括：语义俯视图；

所述俯视语义分析模块进一步用于：

对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理，根据解码处理的结果获取三维场景的语义俯视图。

30. 根据权利要求29所述的装置，其特征在于，在所述至少一第一视角图像为多个第一视角图像的情况下，所述俯视语义分析模块进一步用于：

将多个第一视角图像各自对应的三维场景的俯视角特征信息叠加，获得三维场景的综合俯视角特征信息；

对所述综合俯视角特征信息进行解码处理，根据解码处理的结果获取三维场景的语义俯视图。

31. 根据权利要求20至25中任一项所述的装置，其特征在于，所述特征提取模块、视角转换模块以及俯视语义分析模块利用神经网络系统实现。

32. 根据权利要求31所述的装置，其特征在于，所述神经网络系统是利用从第一视角针对属于第二场景域的三维场景的至少局部区域采集而得的二维图像的语义图样本训练获得的。

33. 根据权利要求31所述的装置，其特征在于，所述神经网络系统包括：编码器、视角转换器以及解码器：

所述编码器用于，对所述至少一第一视角图像进行编码处理，并输出编码处理获得的至少一第一视角图像的特征信息；

所述视角转换器用于，对所述至少一第一视角图像的特征信息进行视角转换处理，并输出至少一第一视角图像对应的三维场景的俯视角特征信息；

所述解码器用于，根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息，获得并输出所述三维场景的俯视语义分析结果。

34. 根据权利要求33所述的装置，其特征在于，所述视角转换器包括：至少一个多层感知机；

在所述至少一第一视角图像为多个第一视角图像的情况下，经所述视角转换器中的多

个多层感知机分别对其对应的第一视角图像的特征信息进行视角转换处理;不同的多层感知机对应不同的第一视角图像。

35. 根据权利要求33所述的装置,其特征在于,所述装置还包括:

样本输入模块,用于将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统中的编码器;

经由所述编码器进行特征提取,以获得至少一第一视角图像样本的特征信息,由所述神经网络系统中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理,获得至少一第一视角图像样本对应的三维场景的俯视角特征信息,并由所述神经网络系统中的解码器对所述至少一第一视角图像样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析;

调整模块,用于根据所述俯视语义分析的结果与所述三维场景的语义俯视图样本之间的损失,对所述神经网络系统中的网络参数进行调整。

36. 根据权利要求25所述的装置,其特征在于:

所述第一场景域包括:真实三维场景;

所述第二场景域包括:虚拟三维场景。

37. 一种基于三维场景分析的控制装置,其特征在于,所述装置包括:

获取模块,用于基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像;

特征提取模块,用于分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;

视角转换模块,用于对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息,其中,所述第一视角图像的特征信息包括:第一视角图像的特征图,所述第一视角图像对应的三维场景的俯视角特征信息包括:第一视角图像对应的三维场景的俯视角特征图;

俯视语义分析模块,用于根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果;

控制模块,用于根据所述俯视语义分析结果,对所述智能设备进行控制。

38. 根据权利要求37所述的装置,其特征在于,所述控制模块包括下述至少之一:

第一控制模块,用于根据所述俯视语义分析结果,进行路径规划,并根据所述路径规划的结果,对所述智能设备进行控制;

第二控制模块,用于根据所述俯视语义分析结果,进行导航规划,并根据所述导航规划的结果,对所述智能设备进行控制;

第三控制模块,用于根据所述俯视语义分析结果,进行避障规划,并根据所述避障规划的结果,对所述智能设备进行控制;

第四控制模块,用于根据所述俯视语义分析结果进行所述智能设备的泊车控制;

第五控制模块,用于根据所述俯视语义分析结果,进行目标抓取规划,并根据所述目标抓取规划的结果,对所述智能设备进行控制。

39. 一种电子设备,包括:

存储器,用于存储计算机程序;

处理器,用于执行所述存储器中存储的计算机程序,且所述计算机程序被执行时,实现上述权利要求1-19中任一项所述的方法。

40.一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时,实现上述权利要求1-19中任一项所述的方法。

41.一种计算机程序,包括计算机指令,当所述计算机指令在设备的处理器中运行时,实现上述权利要求1-19中任一项所述的方法。

三维场景分析方法、装置、介质及设备

技术领域

[0001] 本申请涉及计算机视觉技术,尤其是涉及一种三维场景分析方法和装置、基于三维场景分析的控制方法和装置、电子设备、计算机可读存储介质以及计算机程序。

背景技术

[0002] 在智能驾驶、机器人操控以及避障导航等多种应用中,通常需要涉及到三维场景分析技术。对三维场景的分析往往需要先获得构建出三维场景的三维信息,然后,利用获得的三维信息来分析三维场景。

发明内容

[0003] 本申请实施方式提供一种三维场景分析技术方案以及基于三维场景分析的控制技术方案。

[0004] 根据本申请实施方式其中一个方面,提供一种三维场景分析方法,包括:获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像,所述第一视角与俯视角成一定角度;分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息;根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果。

[0005] 在本申请一实施方式中,所述至少一第一视角包括:多个第一视角;所述多个第一视角中不同第一视角对应所述三维场景的采集区域至少部分不同。

[0006] 在本申请又一实施方式中,多个第一视角图像可形成所述三维场景的非全景图像;或者多个第一视角图像可形成所述三维场景的全景图像。

[0007] 在本申请再一实施方式中,所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像、从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图、以及从第一视角针对三维场景的至少局部区域采集而得的深度图中的至少一个。

[0008] 在本申请再一实施方式中,在所述第一视角图像包括从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图的情况下,所述获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像包括:对从第一视角针对三维场景的至少局部区域采集而得的二维图像进行语义分割处理;根据所述语义分割处理的结果,获取所述二维图像的语义图。

[0009] 在本申请再一实施方式中,所述根据所述语义分割处理的结果,获得所述二维图像的语义图,包括:基于所述语义分割处理的结果,生成所述二维图像的语义图;或者,基于所述语义分割处理的结果,生成待处理语义图,其中,所述待处理语义图采用第一语义表示,所述第一语义对应第一场景域预定的语义;根据预设的第一语义与第二语义的映射关系,将所述待处理语义图从第一语义表示转换为第二语义表示,所述转换后的语义图被作

为所述二维图像的语义图;其中,所述第二语义对应第二场景域预定的语义。

[0010] 在本申请再一实施方式中,所述分别对三维场景的至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息,包括:分别对所述三维场景的至少一第一视角图像进行编码处理,根据编码处理的结果获取至少一第一视角图像的特征信息;所述根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取所述三维场景的俯视语义分析结果。

[0011] 在本申请再一实施方式中,所述第一视角图像的特征信息包括:第一视角图像的特征图;所述第一视角图像对应的三维场景的俯视角特征信息包括:第一视角图像对应的三维场景的俯视角特征图;所述对所述至少一第一视角图像的特征信息进行视角转换处理,以获得至少一第一视角图像对应的三维场景的俯视角特征信息,包括:将至少一第一视角图像的特征图转换为至少一第一视角图像的特征向量;将所述至少一第一视角图像的特征向量分别转换为至少一第一视角图像对应的三维场景的俯视角特征向量;将至少一第一视角图像的俯视角特征向量分别转换为特征图,获得至少一第一视角图像对应的三维场景的俯视角特征图。

[0012] 在本申请再一实施方式中,所述第一视角图像对应的三维场景的俯视角特征图与所述第一视角图像的特征图的大小相同。

[0013] 在本申请再一实施方式中,所述俯视语义分析结果包括:语义俯视图;所述根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

[0014] 在本申请再一实施方式中,在所述至少一第一视角图像为多个第一视角图像的情况下,所述对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图,包括:将多个第一视角图像各自对应的三维场景的俯视角特征信息叠加,获得三维场景的综合俯视角特征信息;对所述综合俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

[0015] 在本申请再一实施方式中,所述分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息,对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息,根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果,包括:将所述至少一第一视角图像输入神经网络系统;由所述神经网络系统对所述至少一第一视角图像进行特征提取,对所述特征提取获得的至少一第一视角图像的特征信息进行视角转换处理,并根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息,获得并输出所述三维场景的俯视语义分析结果。

[0016] 在本申请再一实施方式中,所述神经网络系统是利用从第一视角针对属于第二场景域的三维场景的至少局部区域采集而得的二维图像的语义图样本训练获得的。

[0017] 在本申请再一实施方式中,所述神经网络系统包括:编码器、视角转换器以及解码器;经所述编码器对所述至少一第一视角图像进行编码处理,并输出编码处理获得的至少一第一视角图像的特征信息;经所述视角转换器对所述至少一第一视角图像的特征信息进

行视角转换处理,并输出至少一第一视角图像对应的三维场景的俯视角特征信息;经所述解码器根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息,获得并输出所述三维场景的俯视语义分析结果。

[0018] 在本申请再一实施方式中,所述视角转换器包括:至少一个多层感知机;在所述至少一第一视角图像为多个第一视角图像的情况下,所述经所述视角转换器对所述至少一第一视角图像的特征信息进行视角转换处理,包括:经所述视角转换器中的多个多层感知机分别对其对应的第一视角图像的特征信息进行视角转换处理;不同的多层感知机对应不同的第一视角图像。

[0019] 在本申请再一实施方式中,所述神经网络系统的训练过程包括:将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统中的编码器;经由所述编码器进行特征提取,以获得至少一第一视角图像样本的特征信息,由所述神经网络系统中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理,获得至少一第一视角图像样本对应的三维场景的俯视角特征信息,并由所述神经网络系统中的解码器对所述至少一第一视角图像样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析;根据所述俯视语义分析的结果与所述三维场景的语义俯视图样本之间的损失,对所述神经网络系统中的网络参数进行调整。

[0020] 在本申请再一实施方式中,所述第一场景域包括:真实三维场景;所述第二场景域包括:虚拟三维场景。

[0021] 根据本申请实施方式其中另一方面,提供一种基于三维场景分析的控制方法,包括:基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像;采用上述三维场景分析方法,获得所述三维场景的俯视语义分析结果;根据所述俯视语义分析结果,对所述智能设备进行控制。

[0022] 在本申请一实施方式中,所述根据所述俯视语义分析结果,对所述智能设备进行控制,包括下述至少之一:根据所述俯视语义分析结果,进行路径规划,并根据所述路径规划的结果,对所述智能设备进行控制;根据所述俯视语义分析结果,进行导航规划,并根据所述导航规划的结果,对所述智能设备进行控制;根据所述俯视语义分析结果,进行避障规划,并根据所述避障规划的结果,对所述智能设备进行控制;根据所述俯视语义分析结果进行所述智能设备的泊车控制;根据所述俯视语义分析结果,进行目标抓取规划,并根据所述目标抓取规划的结果,对所述智能设备进行控制。

[0023] 根据本申请实施方式其中再一方面,提供一种三维场景分析装置,包括:获取图像模块,用于获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像,所述第一视角与俯视角成一定角度;特征提取模块,用于分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;视角转换模块,用于对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息;俯视语义分析模块,用于根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果。

[0024] 在本申请一实施方式中,所述至少一第一视角包括多个第一视角;所述多个第一视角中不同第一视角对应所述三维场景的采集区域至少部分不同。

[0025] 在本申请又一实施方式中,多个第一视角图像可形成所述三维场景的非全景图

像;或者,多个第一视角图像可形成所述三维场景的全景图像。

[0026] 在本申请再一实施方式中,所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像、从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图、以及从第一视角针对三维场景的至少局部区域采集而得的深度图中的至少一个。

[0027] 在本申请再一实施方式中,在所述第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图的情况下,所述获取图像模块包括:第一子模块,用于对从第一视角针对三维场景的至少局部区域采集而得的二维图像进行语义分割处理;第二子模块,用于根据所述语义分割处理的结果,获取所述二维图像的语义图。

[0028] 在本申请再一实施方式中,所述第二子模块进一步用于:基于所述语义分割处理的结果,生成所述二维图像的语义图;或者,基于所述语义分割处理的结果,生成待处理语义图,其中,所述待处理语义图采用第一语义表示,所述第一语义对应第一场景域预定的语义;根据预设的第一语义与第二语义的映射关系,将所述待处理语义图从第一语义表示转换为第二语义表示,所述转换后的语义图被作为所述二维图像的语义图;其中,所述第二语义对应第二场景域预定的语义。

[0029] 在本申请再一实施方式中,所述特征提取模块进一步用于:分别对所述三维场景的至少一第一视角图像进行编码处理,根据编码处理的结果获取至少一第一视角图像的特征信息;所述俯视语义分析模块进一步用于:对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取所述三维场景的俯视语义分析结果。

[0030] 在本申请再一实施方式中,所述第一视角图像的特征信息包括:第一视角图像的特征图;所述第一视角图像对应的三维场景的俯视角特征信息包括:第一视角图像对应的三维场景的俯视角特征图;所述视角转换模块进一步用于:将至少一第一视角图像的特征图转换为至少一第一视角图像的特征向量;将所述至少一第一视角图像的特征向量分别转换为至少一第一视角图像对应的三维场景的俯视角特征向量;将至少一第一视角图像的俯视角特征向量分别转换为特征图,获得至少一第一视角图像对应的三维场景的俯视角特征图。

[0031] 在本申请再一实施方式中,所述第一视角图像对应的三维场景的俯视角特征图与所述第一视角图像的特征图的大小相同。

[0032] 在本申请再一实施方式中,所述俯视语义分析结果包括:语义俯视图;所述俯视语义分析模块进一步用于:对所述至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

[0033] 在本申请再一实施方式中,在所述至少一第一视角图像为多个第一视角图像的情况下,所述俯视语义分析模块进一步用于:将多个第一视角图像各自对应的三维场景的俯视角特征信息叠加,获得三维场景的综合俯视角特征信息;对所述综合俯视角特征信息进行解码处理,根据解码处理的结果获取三维场景的语义俯视图。

[0034] 在本申请再一实施方式中,所述特征提取模块、视角转换模块以及俯视语义分析模块利用神经网络系统实现。

[0035] 在本申请再一实施方式中,所述神经网络系统是利用从第一视角针对属于第二场

景域的三维场景的至少局部区域采集而得的二维图像的语义图样本训练获得的。

[0036] 在本申请再一实施方式中,所述神经网络系统包括:编码器、视角转换器以及解码器;所述编码器用于,对所述至少一第一视角图像进行编码处理,并输出编码处理获得的至少一第一视角图像的特征信息;所述视角转换器用于,对所述至少一第一视角图像的特征信息进行视角转换处理,并输出至少一第一视角图像对应的三维场景的俯视角特征信息;所述解码器用于,根据所述视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息,获得并输出所述三维场景的俯视语义分析结果。

[0037] 在本申请再一实施方式中,所述视角转换器包括:至少一个多层感知机;在所述至少一第一视角图像为多个第一视角图像的情况下,经所述视角转换器中的多个多层感知机分别对其对应的第一视角图像的特征信息进行视角转换处理;不同的多层感知机对应不同的第一视角图像。

[0038] 在本申请再一实施方式中,所述装置还包括:样本输入模块,用于将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统中的编码器;经由所述编码器进行特征提取,以获得至少一第一视角图像样本的特征信息,由所述神经网络系统中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理,获得至少一第一视角图像样本对应的三维场景的俯视角特征信息,并由所述神经网络系统中的解码器对所述至少一第一视角图像样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析;调整模块,用于根据所述俯视语义分析的结果与所述三维场景的语义俯视图样本之间的损失,对所述神经网络系统中的网络参数进行调整。

[0039] 在本申请再一实施方式中,所述第一场景域包括:真实三维场景;所述第二场景域包括:虚拟三维场景。

[0040] 根据本申请实施方式其中再一方面,提供一种基于三维场景分析的控制装置,所述装置包括:获取模块,用于基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像;特征提取模块,用于分别对所述至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息;视角转换模块,用于对所述至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息;俯视语义分析模块,用于根据所述至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果;控制模块,用于根据所述俯视语义分析结果,对所述智能设备进行控制。

[0041] 在本申请再一实施方式中,所述控制模块包括下述至少之一:第一控制模块,用于根据所述俯视语义分析结果,进行路径规划,并根据所述路径规划的结果,对所述智能设备进行控制;第二控制模块,用于根据所述俯视语义分析结果,进行导航规划,并根据所述导航规划的结果,对所述智能设备进行控制;第三控制模块,用于根据所述俯视语义分析结果,进行避障规划,并根据所述避障规划的结果,对所述智能设备进行控制;第四控制模块,用于根据所述俯视语义分析结果进行所述智能设备的泊车控制;第五控制模块,用于根据所述俯视语义分析结果,进行目标抓取规划,并根据所述目标抓取规划的结果,对所述智能设备进行控制。

[0042] 根据本申请实施方式再一方面,提供一种电子设备,包括:存储器,用于存储计算机程序;处理器,用于执行所述存储器中存储的计算机程序,且所述计算机程序被执行时,

实现本申请任一方法实施方式。

[0043] 根据本申请实施方式再一个方面,提供一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时,实现本申请任一方法实施方式。

[0044] 根据本申请实施方式的再一个方面,提供一种计算机程序,包括计算机指令,当所述计算机指令在设备的处理器中运行时,实现本申请任一方法实施方式。

[0045] 基于本申请提供的一种三维场景分析方法及装置、基于三维场景分析的控制方法及装置、电子设备、计算机可读存储介质以及计算机程序,通过对三维场景的至少一第一视角图像进行特征提取和视角转换处理,可以获得至少一第一视角图像各自对应的三维场景的俯视角特征信息,从而通过对三维场景的俯视角特征信息的处理(如解码处理等),可以形成三维场景的俯视语义分析结果。由于三维场景的俯视语义分析结果可以在一定程度上体现出地面上任一位置处前后左右的情况,因此,可以利用三维场景的俯视语义分析结果来分析三维场景。由上述描述可知,本申请提供的技术方案可以利用二维信息,来分析三维场景,有利于提高三维场景分析的灵活性以及多样性。

[0046] 下面通过附图和实施方式,对本申请的技术方案做进一步的详细描述。

附图说明

[0047] 构成说明书的一部分的附图描述了本申请的实施方式,并且连同描述一起用于解释本申请的原理。

[0048] 参照附图,根据下面的详细描述,可以更加清楚地理解本申请,其中:

[0049] 图1为本申请的三维场景分析方法一个实施方式的流程图;

[0050] 图2为本申请的室内环境的三维场景和四个第一视角图像的示意图;

[0051] 图3为本申请的二维图像、二维图像的语义图以及最终预测出的语义俯视图;

[0052] 图4为本申请的三维场景分析方法另一个实施方式的流程图;

[0053] 图5为本申请的一神经网络系统的结构示意图;

[0054] 图6为本申请的又一神经网络系统的结构示意图;

[0055] 图7为本申请的三维场景分析方法一个实施方式的示意图;

[0056] 图8为本申请的神经网络系统的训练方法的一个实施方式的流程图;

[0057] 图9为本申请的基于三维场景分析的控制方法的一个实施方式的流程图;

[0058] 图10为本申请的三维场景分析装置一个实施方式的结构示意图;

[0059] 图11为本申请的基于三维场景分析的控制装置一个实施方式的结构示意图;

[0060] 图12为实现本申请实施方式的一示例性设备的框图。

具体实施方式

[0061] 现在将参照附图来详细描述本公开的各种示例性实施例。应注意到:除非另外具体说明,否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

[0062] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0063] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本公开

及其应用或使用的任何限制。

[0064] 对于相关领域普通技术人员已知的技术、方法以及设备可能不作详细讨论,但在适当情况下,所述技术、方法和设备应当被视为说明书的一部分。

[0065] 应当注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0066] 本公开实施例可以应用于终端设备、计算机系统及服务器等电子设备,其可与众多其它通用或者专用的计算系统环境或者配置一起操作。适于与终端设备、计算机系统以及服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子,包括但不限于:车载设备、个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0067] 终端设备、计算机系统以及服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑以及数据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

[0068] 示例性实施例

[0069] 图1为本申请的三维场景分析方法一个实施例的流程图。

[0070] S100、获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像。

[0071] 在一个可选示例中,本申请中的三维场景包括但不限于:真实世界中的三维场景或者虚拟世界中的三维场景。真实世界中的三维场景通常是指现实存在的环境中的场景。虚拟世界中的三维场景通常是指不存在于现实环境中,而存在于通过计算机等电子设备虚拟构造出的环境中的场景,例如,电子游戏中的三维场景;再例如,利用计算机等绘图工具绘制出的三维场景;又例如,利用模拟器生成或渲染而得的三维场景等。本申请中的三维场景可以包括但不限于:室内环境的三维场景、道路环境的三维场景、或者体育运动场环境的三维场景等。

[0072] 在一个可选示例中,本申请中的第一视角通常与俯视角成一定角度。即第一视角通常与俯视角不重合。本申请中的视角通常是指:针对三维场景的观测方向。例如,摄像装置在针对三维场景进行拍摄时,其光轴(即摄像装置的坐标系中的Z轴)所指向的方向。俯视角通常是指针对三维场景的竖直向下的观测方向。第一视角包括:与俯视角垂直的水平观测方向、与俯视角成锐角的观测方向(如斜向下的观测方向)以及与俯视角成钝角的观测方向(如斜向上的观测方向)。本申请中的至少一第一视角图像可以为一个第一视角图像,也可以为多个第一视角图像。多个第一视角中不同第一视角针对三维场景的采集区域(即取景区域)至少部分不同。

[0073] 可选的,多个第一视角图像可以形成三维场景的非全景图。可选的,多个第一视角图像可以形成三维场景的全景(如360度环形全景),例如,利用四张第一视角图像可以形成一三维场景的360度环形全景。一个可选例子,如图2所示,图2左侧是一个室内环境的三维场景,该三维场景中有一正方形,该正方形内有四个三角形,该正方形的中心点表示摄像装

置所在位置,每一个三角形的与靠近中心点的顶点相连的两条斜边均表示,摄像装置的拍摄视野。摄像装置通过面向前后左右四个观测方向进行拍摄,从而获得如图2右侧所示的四个从第一视角对三维场景进行采集而形成的二维图像,即前向二维图像、后向二维图像、左向二维图像和右向二维图像。这四个二维图像中的图像内容可以形成图2左侧的三维场景的360度环形全景。三维场景的360度环形全景为二维全景。

[0074] 由于三维场景的多个第一视角图像中的每一个第一视角图像均可以准确的反映出三维场景的局部区域的图像内容,因此,本申请利用可以形成三维场景的360度全景的多个(如四张)第一视角图像,来形成三维场景的俯视语义分析结果(如语义俯视图),有利于提高俯视语义分析结果的准确性。

[0075] 在一个可选示例中,本申请中的第一视角图像可以包括:从第一视角对三维场景的至少局部区域进行采集(如拍摄或者截屏等)而形成的二维图像(如RGB(红绿蓝)图像者基于红外的图像等)、从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图、以及从第一视角对三维场景的至少局部区域进行采集而形成的深度图中的至少一个。也就是说,本申请可以利用一个或多个二维图像,生成三维场景的俯视语义分析结果(如语义俯视图);本申请也可以利用一个或多个二维图像的语义图,生成三维场景的俯视语义分析结果;本申请还可以利用一个或者多个深度图,生成三维场景的俯视语义分析结果;另外,本申请可以利用二维图像、二维图像的语义图以及深度图中的两者或所有,生成三维场景的俯视语义分析结果。

[0076] 在一个可选示例中,在第一视角图像包括:从第一视角针对三维场景的至少局部区域进行采集(如拍摄或者截屏等)而形成的二维图像的语义图的情况下,本申请可以通过下述步骤1和步骤2来获得语义图:

[0077] 步骤1、对以第一视角对三维场景的至少局部区域进行采集而获得的二维图像,进行语义分割处理。例如,将二维图像输入用于实现语义分割的神经网络中,由该神经网络对输入的二维图像进行语义分割处理,从而获得该神经网络输出的语义分割处理的结果。例如,该神经网络针对二维图像中的每一个像素,分别输出多个类别置信度,其中的每一个类别置信度均可以认为是该像素属于该类别的概率值。

[0078] 可选的,上述步骤1中的类别可以包括但不限于:机动车辆、非机动车辆、行人、交通信号灯、路标牌、标志线、路灯杆、树、房屋、桌子、椅子、电视、柜子以及猫狗等中的至少一个。本申请中的类别通常与用于实现语义分割的神经网络的训练任务相关,如该神经网络的训练任务包括:基于道路环境中的N(N为大于1的整数)种类别的对象(如机动车辆、非机动车辆、行人、交通信号灯、路标牌、标志线、路灯杆、树……、以及背景等),对二维图像进行语义分割处理;那么,成功训练后的神经网络,可以对针对输入的二维图像中的每一个像素,分别输出N个置信度,以表示一个像素分别属于这N种类别的概率值。

[0079] 可选的,上述步骤1中的用于实现语义分割的神经网络可以包括但不限于:FCN(Fully Convolutional Networks,全卷积神经网络)、SegNET(一种由剑桥提出的对图像进行语义分割的深度神经网络)以及PSPnet(Pyramid Scene Parsing Network,金字塔场景分析神经网络)等。

[0080] 步骤2、根据语义分割处理的结果,获得从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图。

[0081] 在一个可选例子中,本申请可以基于神经网络输出的语义分割处理的结果,直接生成从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图。例如,针对语义分割处理的结果中的每一个像素,分别确定该像素的所有置信度中的最高置信度,并将最高置信度所对应的类别作为该像素所属的类别,从而获得二维图像的语义图。本申请中的一个类别表示一个语义。

[0082] 在另一个可选例子中,本申请可以先基于神经网络输出的语义分割处理的结果,生成一语义图,并将该语义图作为待处理语义图,然后,针对该待处理语义图进行语义映射处理,从而获得从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图。

[0083] 例如,先针对语义分割处理的结果中的每一个像素,分别确定该像素的所有置信度中的最高置信度,并将最高置信度所对应的类别作为该像素所属的类别,从而可以获得一语义图,本申请将该语义图作为待处理语义图;之后,根据预先设置的第一语义与第二语义的映射关系,将待处理语义图从第一语义表示转换为第二语义表示,如将待处理语义图中的第一语义分别转换为第二语义,从而获得转换处理后的语义图,本申请将该转换处理后的语义图作为从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图。第一语义对应第一场景域预定的语义。第二语义对应第二场景域预定的语义。第一场景域和第二场景域可以是两个不同场景域,例如,第一场景域为真实场景域,第二场景域为虚拟场景域等;又例如,第一场景域为训练场地域,所述第二场景域为开放道路域;等等。

[0084] 可选的,第一语义转换为第二语义的一个例子可以为:将“男人”这一第一语义转换为“行人”这一第二语义,将“栏杆”这一第一语义转换为“障碍物”这一第二语义。如果将用于实现语义分割的神经网络在执行语义分割处理时所涉及到的所有类别作为第一类别集合,则本申请中的第一语义属于第一类别集合。而如果将本申请中的生成三维场景的语义俯视图时所涉及到的所有类别作为第二类别集合,则本申请中的第二语义属于第二类别集合。

[0085] 本申请通过执行语义转换处理,使用于实现语义分割的神经网络所涉及的语义不必与本申请中的生成三维场景的语义俯视图所涉及的语义完全相同,本申请可以容忍两者之间存在一定程度的差异,从而有利于减小对用于实现语义分割的神经网络的限制,进而有利于提高本申请的实现灵活性。

[0086] 在一个可选示例中,如图3所示,左侧中间位置处为一三维场景,图3左上角和左下角位置处的图像分别为针对该三维场景从第一视角进行采集而形成的两个二维图像(即两个第一视角图像),第一视角图像的数量不只两个,为省略起见,图3仅示出了两个第一视角图像。通过对各第一视角图像分别进行语义分割处理,可以获得各二维图像的语义图(如图3中的各二维图像的右侧所示,语义图的数量与第一视角图像的数量相同,同样的,为省略起见,图3仅示出了两个语义图)。之后,本申请可以对图3所示的语义图进行特征提取处理,以获得每一个语义图的特征信息。

[0087] S110、分别对至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息。

[0088] 在一个可选示例中,本申请可以利用编码器(Encoder)来实现特征信息的提取。例如,将三维场景的每一个第一视角图像分别输入至编码器中,由编码器对输入的各第一视

角图像分别进行编码处理,从而根据编码器输出的编码处理的结果获得各第一视角图像的特征信息(如第一视角图像的特征图,Feature Map)。本申请中的编码器是共享参数的编码器。也就是说,对于同一三维场景的多个第一视角图像而言,本申请可以利用一个编码器对该三维场景的所有第一视角图像进行并行的编码处理。

[0089] 在一个可选示例中,本申请中的编码器可以包括但不限于:基于CNN(Convolutional Neural Networks,卷积神经网络)的编码器。可选的,本申请中的编码器可以为ResNet(Residual Neural Network,残差神经网络),例如,ResNet-18等。该编码器也可以为现有的用于实现语义分割的神经网络中的编码器等。

[0090] 在一个可选示例中,本申请中的编码器是利用属于第二场景域的三维场景的第一视角图像样本,训练而成的。例如,编码器是利用虚拟三维场景的第一视角图像样本,训练而成的。编码器的训练过程可以参见下述实施例中的描述。在此不再详细说明。

[0091] S120、对至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息。

[0092] 在一个可选示例中,本申请中的第一视角图像对应的三维场景的俯视角特征信息可以为:第一视角图像对应的三维场景的俯视角特征图。本申请可以先将每一个第一视角图像的特征图分别转换为第一视角图像的特征向量(例如,在保持第一视角图像的特征图的通道(Channel)数量不发生变化的情况下,拉伸该特征图的宽和高,从而形成第一视角图像的特征向量),然后,再将每一个第一视角图像的特征向量分别转换为第一视角图像对应的三维场景的俯视角特征向量,之后,通过将每一个第一视角图像的俯视角特征向量分别转换为特征图,可以获得每一个第一视角图像各自对应的三维场景的俯视角特征图。

[0093] 需要特别说明的是,由第一视角图像的特征图所形成的第一视角图像的特征向量的数量通常与特征图的通道数量相关,而一个特征向量所包含的元素的数量通常与特征图的宽和高相关,例如,第一视角图像的特征向量的数量可以等于特征图的通道数量,而一个特征向量所包含的元素的数量可以等于宽和高的乘积。

[0094] 在一个可选示例中,本申请可以利用MLP(Multi-Layer Perceptron,多层感知机)来实现视角转换处理。MLP的输入可以为:包含有多个元素的向量数组的形式,MLP的输出同样可以为:包含有多个元素的向量数组的形式,且输入的向量数组和输出的向量数组所包含的元素的数量相同。利用MLP实现视角转换处理的过程可以为:将每一个第一视角图像的特征图分别转换为第一视角图像的特征向量,特征向量的数量为多个,第一视角图像的所有特征向量形成一个特征向量组,特征向量组中的一个特征向量为一个向量数组,然后,将每一个第一视角图像的特征向量(如特征向量组中的所有特征向量)分别输入MLP中,由MLP针对输入的每一个第一视角图像的特征向量分别进行视角转换处理,并为每一个第一视角图像的特征向量生成三维场景的俯视角特征向量,并输出。本申请所形成的第一视角图像的俯视角特征向量通常为多个,第一视角图像的所有俯视角特征向量可以认为是一个俯视角特征向量组,俯视角特征向量组中的一个俯视角特征向量为一个包含有多个元素的向量数组。

[0095] 需要特别说明的是,虽然一个第一视角图像通常是三维场景中的局部区域的图像内容,但是,本申请中的MLP所输出的俯视角特征向量,是三维场景的全景的俯视角特征向量。由此可知,在同一三维场景对应多个第一视角图像的情况下,由于MLP是针对输入的每

一个第一视角图像的特征向量,分别进行视角转换处理,而一个第一视角图像的特征向量并不包含三维场景中的除该第一视角图像所涉及到的区域之外的其他区域的图像内容特征(如并不包含三维场景的全景特征),可是,MLP所生成并输出的是三维场景的俯视角特征向量,因此,MLP输出的俯视角特征向量中的对应除第一视角图像所涉及到的区域之外的其他区域的部分,其准确性,往往会低于俯视角特征向量中的对应第一视角图像所涉及到的区域的部分的准确性。

[0096] 在一个可选示例中,本申请中的MLP是参数不共享的MLP。也就是说,对于同一三维场景的多个第一视角图像而言,本申请需要利用多个MLP对同一三维场景对应的多个第一视角图像的特征向量进行视角转换处理,例如,设置四个MLP,第一个MLP对应图2中的前向二维图像的特征向量,第二个MLP对应图2中的后向二维图像的特征向量,第三个MLP对应图2中的左向二维图像的特征向量,第四个MLP对应图2中的右向二维图像的特征向量;在存在多个MLP的情况下,多个MLP的参数通常不相同。通过利用多个MLP对多个二维图像的特征向量进行视角转换处理,有利于提高MLP的训练效率,并有利于提高俯视角特征信息的准确性。

[0097] 本申请通过将第一视角图像的特征图转换为特征向量,并将第一视角图像的特征向量提供给MLP,使MLP在进行视角转换处理的过程中,可以在考虑整个第一视角图像的特征信息的情况下,生成该三维场景的俯视角特征向量,从而有利于避免由于MLP的感受野局限于第一视角图像中的局部区域,而导致的生成三维场景的俯视角特征向量不准确的现象。

[0098] 在一个可选示例中,本申请中的MLP是利用虚拟三维场景的第一视角图像样本,训练而成的。MLP的训练过程可以参见下述实施例中的描述。在此不再详细说明。

[0099] S130、根据至少一第一视角图像对应的三维场景的俯视角特征信息,获得三维场景的俯视语义分析结果。

[0100] 在一个可选示例中,本申请可以对各第一视角图像分别对应的三维场景的俯视角特征信息进行解码处理,从而根据解码处理的结果,获得三维场景的俯视语义分析结果。该三维场景的俯视语义分析结果通常为语义俯视图。本申请所获得三维场景的语义俯视图通常用于三维场景分析。三维场景分析可以是指:获知三维场景所包括的对象、各对象所处的方位以及各对象的尺寸大小等。

[0101] 在一个可选示例中,在至少一第一视角图像为一个的情况下,本申请可以针对该第一视角图像对应的三维场景的俯视角特征图进行解码处理,从而根据解码处理的结果,获得三维场景的语义俯视图。

[0102] 在另一个可选示例中,在至少一第一视角图像为多个第一视角图像的情况下,本申请可以先将多个第一视角图像各自对应的三维场景的俯视角特征图进行叠加,并将叠加的结果作为三维场景的综合俯视角特征图,之后,对获得的综合俯视角特征图进行解码处理,从而根据解码处理的结果,获得三维场景的语义俯视图。本申请通过对俯视角特征图进行叠加,有利于使综合俯视角特征图能够更准确的反映三维场景,从而有利于提高三维场景的俯视语义分析结果的准确性。

[0103] 在一个可选示例中,本申请可以利用解码器对各第一视角图像分别对应的三维场景的俯视角特征信息进行解码处理。例如,在至少一第一视角图像为一个的情况下,本申请

可以将第一视角图像对应的三维场景的俯视角特征图输入解码器,由解码器对输入的俯视角特征图进行解码处理,从而获得解码器生成并输出的三维场景的语义俯视图。再例如,在至少一第一视角图像为多个第一视角图像的情况下,本申请可以先将多个第一视角图像各自对应的三维场景的俯视角特征图进行叠加(如相加),将叠加的结果作为三维场景的综合俯视角特征图,然后,将综合俯视角特征图输入解码器中,由解码器对输入的综合俯视角特征图进行解码处理,从而本申请可以根据解码器输出的解码处理的结果,获得三维场景的俯视语义分析结果(如语义俯视图)。

[0104] 在一个可选示例中,本申请中的解码器可以包括但不限于:基于CNN的解码器。可选的,本申请中的解码器可以采用金字塔池化模块(Pyramid Pooling Module)等。该解码器也可以为现有的用于实现语义分割的神经网络中的解码器等。本申请中的解码器是利用属于第二场景域的三维场景的第一视角图像样本,训练而成的。例如,解码器是利用虚拟三维场景的第一视角图像样本,训练而成的。解码器的训练过程可以参见下述实施例中的描述。在此不再详细说明。

[0105] 续前例,本申请针对图3所示的多个语义图进行特征提取处理、视角转换处理以及语义俯视图生成处理之后,最终获得的三维场景的语义俯视图,如图3中的右侧所示。

[0106] 本申请的三维场景分析方法可以由一神经网络系统来实现,该神经网络系统可以称为视角解析神经网络(View Parsing Network,VPN)。利用神经网络系统实现三维场景分析方法的一个实施例如图4所示。

[0107] S400、将三维场景的至少一第一视角图像输入至神经网络系统。

[0108] 在一个可选示例中,本申请的神经网络系统包括:编码器(encoder,如基于CNN的编码器等)、视角转换器(View transformer,如MLP等)以及解码器(decoder,如基于CNN的解码器)。

[0109] 可选的,本申请的神经网络系统中视角转换器可以为一个,用于(如图5所示)。

[0110] 可选的,本申请的神经网络中视角转换器的数量可以为多个,且多个视角转换器的网络参数不共享。

[0111] 神经网络所包含的视角转换器的数量与三维场景的第一视角图像的数量相关。例如,三维场景的第一视角图像的数量为1则神经网络系统可以包含1个视角转换器。再例如,三维场景的第一视角图像的数量为2,则神经网络系统可以包含2个网络参数不共享的视角转换器。再例如,三维场景的第一视角图像的数量为3,则神经网络系统可以包含3个网络参数不共享的视角转换器。再例如,如图7左侧所示的三维场景,该三维场景的第一视角图像的数量为4(即图7中的前向二维图像、后向二维图像、左向二维图像和右向二维图像),则神经网络系统可以包含4个网络参数不共享的视角转换器。再例如,三维场景的第一视角图像的数量为8,则神经网络系统可以包括8个网络参数不共享的视角转换器。本申请中的一个视角转换器仅对应一个拍摄方向的第一视角图像。第一视角图像的数量通常会最终形成的三维场景的语义俯视图的准确性产生影响,例如,第一视角图像的数量越多,则最终形成的三维场景的语义俯视图往往会越准确。如果对形成三维场景的语义俯视图的准确性和到实时性进行综合考量,第一视角图像的数量可以设置为4。

[0112] 可选的,本申请中的拍摄方向可以认为是:观测(observation)方向。例如,以三维场景中的一点为拍摄位置,并针对该拍摄位置预先确定至少一方向(如前后左右四个方

向),在摄像装置的光轴指向预先确定的方向进行拍摄时,该光轴所执行的方向即为拍摄方向。

[0113] 可选的,本申请的神经网络系统可以包括:多个编码器、多个视角转换器以及多个解码器,且多个编码器的网络参数共享,多个解码器的网络参数也共享。神经网络系统所包含的编码器、视角转换器以及解码器的数量与三维场景的第一视角图像的数量相关。例如,三维场景的第一视角图像的数量为4,则神经网络系统可以包含4个网络参数共享的编码器、4个网络参数不共享的视角转换器、以及4个网络参数共享的解码器。

[0114] 在一个可选示例中,神经网络系统中的编码器和解码器的结构可以采用:现有的用于实现语义分割的神经网络中的编码器和解码器的结构。由于用于实现语义分割的神经网络是由编码器和解码器组成的,因此,本申请可以通过将视角转换器插入在用于实现语义分割的神经网络中的编码器和解码器之间,从而形成本申请的神经网络系统。由此可知,本申请的神经网络系统易于实现,即有利于提高本申请的神经解析神经网络的实现灵活性。另外,由于用于实现语义分割的神经网络的性能在不断的提升,因此,本申请的神经网络系统的性能可以得到一定的保证。

[0115] 在一个可选示例中,输入至神经网络系统的第一视角图像可以为:从第一视角对三维场景的至少局部区域进行采集而形成的二维图像(如图7所示的前向二维图像、后向二维图像、左向二维图像和右向二维图像)、从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图(如图3中的左上第2个图和左下第2个图所示的两个语义图)、或者从第一视角对三维场景的至少局部区域进行采集而形成的深度图等。

[0116] 在输入至神经网络系统的第一视角图像为:从第一视角对三维场景的至少局部区域进行采集而形成的二维图像的语义图的情况下,本申请可以先获得从第一视角对三维场景的至少局部区域进行采集而形成的二维图像(如RGB图像等),然后将二维图像输入用于实现语义分割的神经网络(如FCN、SegNET或PSPnet)中,从而可以基于用于实现语义分割的神经网络输出的语义分割处理的结果,获得该二维图像的语义图。

[0117] 如果不需要对上述获得的语义图进行语义映射处理,则本申请可以将该二维图像的语义图直接输入至神经网络系统中。

[0118] 如果需要对上述获得的语义图进行语义映射处理,则本申请可以先根据预先设置的第一语义与第二语义的映射关系,将利用第一语义表示的语义图转换为利用第二语义表示的语义图,也就是说,将语义图中的第一语义分别转换为第二语义,然后,将语义转换后的语义图输入神经网络系统中。如果将用于实现语义分割的神经网络在语义分割处理所涉及到的所有类别作为第一类别集合,则本申请中的第一语义属于第一类别集合。而如果将本申请中的神经网络系统在生成三维场景的语义俯视图时所涉及到的所有类别作为第二类别集合,则本申请中的第二语义属于第二类别集合。

[0119] 本申请通过执行语义转换处理,使用于实现语义分割的神经网络所涉及的语义不必与本申请中的神经网络系统所涉及的语义完全相同,也就是说,本申请可以容忍用于实现语义分割的神经网络与神经网络系统之间存在一定程度的差异,从而有利于减小对用于实现语义分割的神经网络的限制,换个角度而言,本申请有利于提高神经网络系统的可适用性能,从而有利于提高本申请的可实施性能。

[0120] S410、经由该神经网络系统对输入的第一视角图像进行特征提取处理,对特征提

取处理获得的第一视角图像的特征信息进行视角转换处理,并对视角转换处理获得的第一视角图像对应的三维场景的俯视角特征信息进行俯视语义分析处理,获得并输出三维场景的俯视语义分析结果。

[0121] 在一个可选示例中,本申请可以利用神经网络系统中的编码器来实现特征信息的提取。如图7所示,三维场景的四个第一视角图像被分别输入至神经网络系统中,由神经网络系统中的编码器对输入的各第一视角图像分别进行编码处理,从而编码器为每一个视角图像生成并输出第一视角图像的特征图(如图7中的标注有前、后、左、右的四个长方体)。

[0122] 在一个可选示例中,本申请可以利用神经网络系统中的各视角转换器(如多个MLP)来实现视角转换处理。例如:

[0123] 首先,本申请将编码器输出的每一个第一视角图像的特征图(例如,图7中位于视角转换器左侧的标注有“W”(W表示宽)、“C”(C表示Channel,通道)和“H”(H表示高)的长方体)分别转换为第一视角图像的特征向量(例如,在保持第一视角图像的特征图的通道维度不发生变化的情况下,拉伸该特征图的宽和高,从而形成第一视角图像的特征向量,图7中位于视角转换器左侧的标注有“C”和“HW”的长方体);本申请所形成的第一视角图像的特征向量通常为多个,第一视角图像的所有特征向量可以认为是一个特征向量组。可选的,针对一个第一视角图像而言,假定编码器输出的该第一视角图像的特征图的大小为 $13(W) \times 13(H) \times 512(C)$,则本申请可以将该特征图转换为512个特征向量,这512个特征向量形成一个特征向量组,且每一个特征向量均包括169(13×13)个元素。

[0124] 其次,将每一个第一视角图像的特征向量组中的特征向量分别输入至相应的视角转换器(图7中仅示出了一个视角转换器,实际应包括四个视角转换器)中,由每一个视角转换器分别对其接收到的特征向量进行视角转换处理,从而生成每一个第一视角图像的俯视角特征向量(例如,图7中位于视角转换器右侧的标注有“C”和“HW”的长方体);本申请形成的第一视角图像的俯视角特征向量通常为多个,第一视角图像的所有俯视角特征向量可以认为是一个包含有多个元素的俯视角特征向量组。续前例,针对一个第一视角图像而言,本申请的视角转换器输出512个俯视角特征向量,这512个俯视角特征向量形成一个俯视角特征向量组,且每一个俯视角特征向量均包括169(13×13)个元素。

[0125] 最后,本申请通过将视角转换器输出的每一个第一视角图像的所有俯视角特征向量(即第一视角图像的俯视角特征向量组)分别转换为特征图(例如,图7中位于视角转换器右侧的标注有“W”、“C”和“H”的长方体),从而形成每一个第一视角图像各自对应的三维场景的俯视角特征图(图7中仅示出了一个俯视角特征图,实际应包括四个俯视角特征图)。编码器输出的第一视角图像的特征图的大小与输入解码器的俯视角特征图的大小是相同的,本申请的视角转换器并不会导致特征图的大小发生变化。续前例,针对一个第一视角图像而言,本申请的形成的三维场景的俯视角特征图的大小为 $13(W) \times 13(H) \times 512(C)$ 。

[0126] 本申请通过将第一视角图像的特征图转换为特征向量,并将第一视角图像的特征向量提供给视角转换器,使视角转换器在进行视角转换处理的过程中,可以在考虑整个第一视角图像的特征信息的情况下,生成该三维场景的俯视角特征向量,从而有利于避免由于视角转换器的感受野局限于第一视角图像中的局部区域,而导致的生成三维场景的俯视角特征向量不准确的现象。

[0127] 需要特别说明的是,虽然一个第一视角图像通常是三维场景中的局部区域的图像

内容,但是,本申请中的视角转换器所输出的俯视角特征向量,是三维场景的全景的俯视角特征向量。由此可知,在同一三维场景对应多个第一视角图像的情况下,由于视角转换器是针对输入的每一个第一视角图像的特征向量,分别进行视角转换处理,而一个第一视角图像的特征向量并不包含三维场景中的除该第一视角图像所涉及到的区域之外的其他区域的图像内容特征(如并不包含三维场景的全景特征),可是,视角转换器所生成并输出的是三维场景的俯视角特征向量,因此,视角转换器输出的俯视角特征向量中的对应除第一视角图像所涉及到的区域之外的其他区域的部分,其准确性,往往会低于俯视角特征向量中的对应第一视角图像所涉及到的区域的部分的准确性。

[0128] 在一个可选示例中,本申请可以利用神经网络系统中的解码器来实现俯视语义分析处理,如生成语义俯视图的处理。例如,在神经网络系统只包括一个视角转换器的情况下,本申请可以将该视角转换器输出的三维场景的俯视角特征向量转换为三维场景的俯视角特征图后,将三维场景的俯视角特征图输入至解码器中,由解码器针对该三维场景的俯视角特征图进行解码处理,并输出解码处理的结果。再例如,在神经网络系统包括多个视角转换器的情况下,本申请可以先将多个视角转换器输出的三维场景的俯视角特征向量分别转换为三维场景的俯视角特征图,然后,将各俯视角特征图进行叠加(如相加等),并将叠加后形成的三维场景的俯视角特征图输入至解码器(如图7所示的解码器,图7中未示出四个俯视角特征图相加的过程中,由解码器针对该叠加后的三维场景的俯视角特征图进行解码处理,并输出解码处理的结果。该解码处理的结果即为俯视语义分析结果。

[0129] S420、根据神经网络系统输出的俯视语义分析结果,获得三维场景的语义俯视图。

[0130] 在一个可选示例中,神经网络系统输出的俯视语义分析结果可以包括:所形成的三维场景的俯视图中的每一个像素的类别置信度(类别置信度的数量与类别的数量相关),其中的每一个类别置信度均可以认为是该像素属于该类别的概率值。本申请中的类别通常与神经网络系统的训练任务相关,例如,神经网络系统的训练任务包括:对包含有N(N为大于1的整数)种类别的对象(如机动车辆、非机动车辆、行人、交通信号灯、路标牌、标志线、路灯杆、树……、以及背景等)的三维道路环境进行俯视语义分析;那么,成功训练后的神经网络,可以对针对输入的多个第一视角图像,形成一俯视图,且该俯视图中的每一个像素均具有N个置信度,以表示该像素分别属于这N种类别的概率值。

[0131] 在一个可选例子中,本申请可以针对俯视图中的每一个像素,分别确定该像素的所有置信度中的最高置信度,并将最高置信度所对应的类别作为该像素所属的类别,从而本申请可以形成一语义俯视图(如图7最右侧所示的图)。

[0132] 本申请的神经网络系统通常是利用虚拟三维场景的至少一第一视角图像样本训练获得的。下面结合图8所示的流程,对本申请的神经网络系统的训练过程进行说明。

[0133] S800、将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统。

[0134] 在一个可选示例中,本申请中的第二场景域可以为虚拟场景域,属于第二场景域的三维场景可以为虚拟三维场景,虚拟三维场景可以为游戏中的三维场景等。由于虚拟三维场景中的每一个像素均具有类别标注,因此,本申请可以避免样本逐像素的标注,所导致的工作量大、周期长以及成本高等问题,从而本申请可以低成本并快速的获得大量的用于训练的样本。

[0135] 在一个可选示例中,本申请可以在虚拟三维场景中设置拍摄位置,并基于该拍摄位置和第一视角,形成一个或多个第一视角图像样本。第一视角图像样本可以为从第一视角对三维场景(如虚拟三维场景)中的至少局部区域进行采集而形成的二维图像样本、从第一视角对三维场景(如虚拟三维场景)中的至少局部区域进行采集而形成的二维图像样本的语义图样本、或者从第一视角对三维场景(如虚拟三维场景)中的至少局部区域进行采集而形成的深度图样本。

[0136] 在一个可选示例中,在第一视角图像为:从第一视角对虚拟三维场景中的至少局部区域进行采集而形成的二维图像样本的语义图样本的情况下,由于虚拟三维场景中的每一个像素均具有类别标注,因此,本申请的二维图像样本中的每一个像素也具有类别标注,从而本申请可以在不借助用于实现语义分割的神经网络的情况下,非常方便的获得用于训练神经网络系统的语义图样本。

[0137] S810、经由待训练的神经网络系统中的编码器进行特征提取,以获得至少一第一视角图像样本的特征信息,由视角神经网络中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理,以获得至少一第一视角图像样本对应的三维场景的俯视角特征信息,并由视角神经网络中的解码器对至少一第一视角图像样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析处理。

[0138] 在一个可选示例中,本申请的神经网络系统中的编码器对输入的第一视角图像样本进行特征提取处理,从而生成各第一视角图像样本的特征图。每一个第一视角图像样本的特征图在被转换为第一视角图像样本的特征向量后,被分别输入至相应的视角转换器(如四个第一视角图像样本的特征向量被分别输入至四个视角转换器)中。每一个视角转换器分别对其接收到的特征向量进行视角转换处理,从而生成第一视角图像样本的俯视角特征向量,之后,本申请将视角转换器输出的每一个第一视角图像样本的俯视角特征向量分别转换为特征图,从而形成各第一视角图像样本各自对应的三维场景的俯视角特征图。编码器输出的第一视角图像样本的特征图的大小与输入解码器的俯视角特征图的大小是相同的。

[0139] 在一个可选示例中,在待训练的神经网络系统包括一个视角转换器的情况下,本申请可以将该视角转换器输出的三维场景(如虚拟三维场景)的俯视角特征向量转换为三维场景的俯视角特征图后,将三维场景的俯视角特征图输入至解码器中,由解码器针对该三维场景的俯视角特征图进行解码处理,并输出解码处理的结果。

[0140] 在一个可选示例中,在待训练的神经网络系统包括多个视角转换器的情况下,本申请可以先将多个视角转换器输出的三维场景的俯视角特征向量分别转换为三维场景的俯视角特征图,然后,将各俯视角特征图进行叠加(如相加),并将叠加(如相加)后形成的三维场景的俯视角特征图输入至解码器中,由解码器针对该叠加后的三维场景的俯视角特征图进行解码处理,并输出解码处理的结果。

[0141] S820、根据神经网络系统输出的俯视语义分析处理的结果与三维场景的语义俯视图样本之间的损失,对神经网络系统中的网络参数进行调整。

[0142] 在一个可选示例中,待训练的神经网络系统输出的俯视语义分析处理的结果可以包括:所形成的三维场景(如虚拟三维场景)的俯视图中的每一个像素的类别置信度(类别置信度的数量与类别的数量相关),其中的每一个类别置信度均可以认为是该像素属于该

类别的概率值。

[0143] 在一个可选示例中,本申请可以针对俯视图中的每一个像素,分别确定该像素的所有置信度中的最高置信度,并将最高置信度所对应的类别作为该像素所属的类别,从而本申请可以形成一语义俯视图。

[0144] 在一个可选示例中,本申请可以利用虚拟三维场景生成语义俯视图,由于虚拟三维场景中的每一个像素均具有类别标注,因此,本申请可以方便的获得虚拟三维场景的俯视图样本中的每一个像素的类别,本申请可以通过虚拟三维场景中的像素所具有的类别标注获得像素的语义,从而可以方便的形成语义俯视图样本。本申请可以利用NLL损失函数来计算基于视角解析神经形成的俯视语义分析结果(如语义俯视图)与三维场景的语义俯视图样本之间的损失,并根据该损失对待训练的神经网络系统中的网络参数进行调整,如对神经网络系统中的编码器、视角转换器以及解码器的卷积核的权值等进行调整。

[0145] 在一个可选示例中,在针对神经网络系统的训练达到预定迭代条件时,本次训练过程结束。本申请中的预定迭代条件可以包括:基于神经网络系统输出的俯视语义分析处理结果(如语义俯视图)与语义俯视图样本之间的差异满足预定差异要求。在该差异满足要求的情况下,本次对神经网络系统成功训练完成。本申请中的预定迭代条件也可以包括:对神经解析神经网络进行训练,所使用的第一视角图像样本的数量达到预定数量要求等。在使用的第一视角图像样本的数量达到预定数量要求,然而,基于神经网络系统输出的俯视语义分析处理结果(如语义俯视图)与语义俯视图样本之间的差异并未满足预定差异要求情况下,本次对神经网络系统并未训练成功。

[0146] 下面结合图9所示的流程,对基于三维场景分析的控制方法进行说明。

[0147] S900、基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像。

[0148] 可选的,该智能设备包括但不限于:车辆中的智能驾驶系统、机器人、机器臂或者智能导航设备等。该摄像装置可以包括:多个摄像头,以便于形成三维场景的多个第一视角图像。

[0149] S910、分别对至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息。该步骤的具体内容可以参见上述实施例中针对S110的描述。

[0150] S920、对至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息。该步骤的具体内容可以参见上述实施例中针对S120的描述。

[0151] S930、根据至少一第一视角图像对应的三维场景的俯视角特征信息,获得三维场景的俯视语义分析结果。该步骤的具体内容可以参见上述实施例中针对S130的描述。

[0152] S940、根据俯视语义分析结果,对智能设备进行控制。

[0153] 可选的,本申请可以根据俯视语义分析结果(如语义俯视图),进行路径规划,并根据路径规划的结果,对智能设备进行控制。例如,本申请可以根据语义俯视图中的各像素的类别,判断出三维场景中的可行走区域以及不可行走区域,从而可以根据可行走区域进行路径规划。例如,运用启发式算法等进行路径规划。

[0154] 可选的,本申请可以根据俯视语义分析结果,进行导航规划,并根据导航规划的结果,对智能设备进行控制。例如,本申请可以根据语义俯视图中的各像素的类别,判断出三维场景中的可行走区域、不可行走区域以及目标对象所在的位置,从而可以根据可行走区

域以及目标对象所在的位置进行导航规划。

[0155] 可选的,本申请可以根据俯视语义分析结果,进行避障规划,并根据避障规划的结果,对智能设备进行控制。例如,本申请可以根据语义俯视图中的各像素的类别,判断出三维场景中的障碍物所在的位置,从而可以根据障碍物所在的位置进行避障规划。

[0156] 可选的,本申请可以根据俯视语义分析结果对智能设备进行泊车控制。例如,本申请可以根据语义俯视图中的各像素的类别,判断出三维场景中的目标车位以及其他车辆所在的位置,从而可以根据目标车位以及其他车辆所在的位置进行泊车控制。

[0157] 可选的,本申请可以根据俯视语义分析结果,进行目标抓取规划,并根据目标抓取规划的结果,对智能设备进行控制。如本申请可以根据语义俯视图中的各像素的类别,判断出三维场景中的目标对象所在的位置,从而可以根据目标对象所在的位置进行抓取规划。

[0158] 可选的,本申请对智能设备的控制包括但不限于:控制智能设备的行动、发出报警等。

[0159] 图10为本申请的三维场景分析装置一个实施例的结构示意图。如图10所示,该实施例的装置包括:获取图像模块1000、特征提取模块1010、视角转换模块1020以及俯视语义分析模块1030。

[0160] 获取图像模块1000用于获取从至少一第一视角针对三维场景采集而得的至少一第一视角图像。第一视角与俯视角成一定角度。

[0161] 可选的,至少一第一视角包括多个第一视角;多个第一视角中不同第一视角对应三维场景的采集区域至少部分不同。可选的,多个第一视角图像可形成三维场景的非全景图像。可选的,多个第一视角图像可形成三维场景的全景图像。可选的,第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像、从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图、以及从第一视角针对三维场景的至少局部区域采集而得的深度图中的至少一个。

[0162] 可选的,在第一视角图像包括:从第一视角针对三维场景的至少局部区域采集而得的二维图像的语义图的情况下,获取图像模块1000包括:第一子模块和第二子模块。其中的第一子模块用于将从第一视角针对三维场景的至少局部区域采集而得的二维图像,进行语义分割处理。其中的第二子模块用于根据语义分割处理的结果,获得二维图像的语义图。

[0163] 可选的,在一些应用场景中,第二子模块可以基于语义分割处理的结果,生成二维图像的语义图。在另一些应用场景中,第二子模块可以基于语义分割处理的结果,生成待处理语义图,并根据预设的第一语义与第二语义的映射关系,将待处理语义图从第一语义表示转换为第二语义表示,转换后的语义图被作为二维图像的语义图。其中,待处理语义图采用第一语义表示,第一语义对应第一场景域预定的语义。其中,第二语义对应第二场景域预定的语义。

[0164] 特征提取模块1010用于分别对至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息。可选的,特征提取模块1010可以对三维场景的至少一第一视角图像进行编码处理,以根据编码处理的结果获得至少一第一视角图像的特征信息。其中的第一视角图像的特征信息可以包括:第一视角图像的特征图。

[0165] 视角转换模块1020用于对至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息。其中的第一视角图像对应的三

维场景的俯视角特征信息可以包括：第一视角图像对应的三维场景的俯视角特征图。

[0166] 可选的，视角转换模块1020可以将至少一第一视角图像的特征图转换为至少一第一视角图像的特征向量，并将至少一第一视角图像的特征向量分别转换为至少一第一视角图像对应的三维场景的俯视角特征向量，之后，视角转换模块1020将至少一第一视角图像的特征向量分别转换为特征图，获得至少一第一视角图像对应的三维场景的俯视角特征图。其中，第一视角图像对应的三维场景的俯视角特征图与第一视角图像的特征图的大小相同。

[0167] 俯视语义分析模块1030用于根据至少一第一视角图像对应的三维场景的俯视角特征信息，获得三维场景的俯视语义分析结果。可选的，俯视语义分析模块1030可以对至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理，以根据解码处理的结果获得三维场景的俯视语义分析结果。

[0168] 可选的，俯视语义分析结果包括：语义俯视图。俯视语义分析模块1030可以对至少一第一视角图像对应的三维场景的俯视角特征信息进行解码处理，以根据解码处理的结果获得三维场景的语义俯视图。

[0169] 可选的，在至少一第一视角图像为多个第一视角图像的情况下，俯视语义分析模块1030可以将多个第一视角图像各自对应的三维场景的俯视角特征信息叠加(如相加)，获得三维场景的综合俯视角特征信息；然后，俯视语义分析模块1030对综合俯视角特征信息进行解码处理，以根据解码处理的结果获得三维场景的语义俯视图。

[0170] 可选的，三维场景分析装置中的特征提取模块1010、视角转换模块1020以及俯视语义分析模块1030可以利用神经网络系统实现。该装置还包括：样本输入模块和调整模块。

[0171] 神经网络系统用于对至少一第一视角图像进行特征提取，对特征提取获得的至少一第一视角图像的特征信息进行视角转换处理，并根据视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息，获得并输出三维场景的俯视语义分析结果。本申请中的神经网络系统是利用基于第一视角针对属于第二场景域的三维场景拍摄而形成的二维图像的语义图样本训练获得的。本申请中的神经网络系统可以称为视角解析神经网络。

[0172] 可选的，该神经网络系统可以包括：编码器、视角转换器以及解码器。编码器用于对至少一第一视角图像进行编码处理，并输出编码处理获得的至少一第一视角图像的特征信息。视角转换器用于对至少一第一视角图像的特征信息进行视角转换处理，并输出至少一第一视角图像对应的三维场景的俯视角特征信息。解码器用于根据视角转换处理获得的至少一第一视角图像对应的三维场景的俯视角特征信息，获得并输出三维场景的俯视语义分析结果。其中的视角转换器可以包括：至少一个多层感知机。在至少一第一视角图像为多个第一视角图像的情况下，经视角转换器中的多个多层感知机分别对其对应的第一视角图像的特征信息进行视角转换处理；不同的多层感知机对应不同的第一视角图像。

[0173] 样本输入模块用于在神经网络系统的训练过程中，将属于第二场景域的三维场景的至少一第一视角图像样本输入待训练的神经网络系统中的编码器。经由编码器进行特征提取处理，以获得至少一第一视角图像样本的特征信息，由神经网络系统中的视角转换器对至少一第一视角图像样本的特征信息进行视角转换处理，获得至少一第一视角图像样本对应的三维场景的俯视角特征信息，并由神经网络系统中的解码器对至少一第一视角图像

样本对应的三维场景的俯视角特征信息,进行三维场景的俯视语义分析。

[0174] 调整模块用于根据俯视语义分析的结果与三维场景的语义俯视图样本之间的损失,对神经网络系统中的网络参数进行调整。

[0175] 可选的,本申请中的第一场景域包括:真实三维场景。第二场景域包括:虚拟三维场景。

[0176] 图11为本申请的基于三维场景分析的控制装置一个实施例的结构示意图。如图11所示,该实施例的装置包括:获取模块1100、特征提取模块1010、视角转换模块1020、俯视语义分析模块1030以及控制模块1110。

[0177] 获取模块1100用于基于智能设备上设置的摄像装置,获得三维场景的至少一第一视角图像。

[0178] 特征提取模块1010用于分别对至少一第一视角图像进行特征提取,获得至少一第一视角图像的特征信息。

[0179] 视角转换模块1020用于对至少一第一视角图像的特征信息进行视角转换处理,获得至少一第一视角图像对应的三维场景的俯视角特征信息。

[0180] 俯视语义分析模块1030用于根据至少一第一视角图像对应的三维场景的俯视角特征信息,获得所述三维场景的俯视语义分析结果。

[0181] 控制模块1110用于根据俯视语义分析结果,对智能设备进行控制。

[0182] 可选的,控制模块1110可以包括:第一控制模块、第二控制模块、第三控制模块、第四控制模块以及第五控制模块中的至少一个。其中的第一控制模块用于根据俯视语义分析结果,进行路径规划,并根据路径规划的结果,对所述智能设备进行控制。其中的第二控制模块用于根据俯视语义分析结果,进行导航规划,并根据导航规划的结果,对所述智能设备进行控制。其中的第三控制模块用于根据俯视语义分析结果,进行避障规划,并根据避障规划的结果,对智能设备进行控制。其中的第四控制模块用于根据俯视语义分析结果,对智能设备进行泊车控制。第五控制模块用于根据俯视语义分析结果,进行目标抓取规划,并根据目标抓取规划的结果,对智能设备进行控制。

[0183] 示例性设备

[0184] 图12示出了适于实现本申请的示例性设备1200,设备1200可以是汽车中配置的控制装置/电子系统、移动终端(例如,智能移动电话等)、个人计算机(PC,例如,台式计算机或者笔记型计算机等)、平板电脑以及服务器等。图12中,设备1200包括一个或者多个处理器、通信部等,所述一个或者多个处理器可以为:一个或者多个中央处理单元(CPU)1201,和/或,一个或者多个的加速单元(如GPU)1213等,处理器可以根据存储在只读存储器(ROM)1202中的可执行指令或者从存储部分1208加载到随机访问存储器(RAM)1203中的可执行指令而执行各种适当的动作和处理。通信部1212可以包括但不限于网卡,所述网卡可以包括但不限于IB(Infiniband)网卡。处理器可与只读存储器1202和/或随机访问存储器1203中通信以执行可执行指令,通过总线1204与通信部1212相连、并经通信部1212与其他目标设备通信,从而完成本申请中的相应步骤。

[0185] 上述各指令所执行的操作可以参见上述方法实施例中的相关描述,在此不再详细说明。此外,在RAM 1203中,还可以存储有装置操作所需的各种程序以及数据。CPU1201、ROM1202以及RAM1203通过总线1204彼此相连。

[0186] 在有RAM1203的情况下,ROM1202为可选模块。RAM1203存储可执行指令,或在运行时向ROM1202中写入可执行指令,可执行指令使中央处理单元1201执行上述三维场景分析方法或者基于三维场景分析的控制方法所包括的步骤。输入/输出(I/O)接口1205也连接至总线1204。通信部1212可以集成设置,也可以设置为具有多个子模块(例如,多个IB网卡),并分别与总线连接。

[0187] 以下部件连接至I/O接口1205:包括键盘、鼠标等的输入部分1206;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分1207;包括硬盘等的存储部分1208;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分1209。通信部分1209经由诸如因特网的网络执行通信处理。驱动器1210也根据需要连接至I/O接口1205。可拆卸介质1211,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器1210上,以便于从其上读出的计算机程序根据需要被安装在存储部分1208中。

[0188] 需要特别说明的是,如图12所示的架构仅为一种可选实现方式,在具体实践过程中,可根据实际需要对上述图12的部件数量和类型进行选择、删减、增加或替换;在不同功能部件设置上,也可采用分离设置或集成设置等实现方式,例如,加速单元(如GPU)1213和CPU1201可分离设置,再如,可将加速单元1213集成在CPU1201上,通信部1212可分离设置,也可集成设置在CPU1201或加速单元1213上等。这些可替换的实施方式均落入本申请的保护范围。

[0189] 特别地,根据本申请的实施方式,下文参考流程图描述的过程可以被实现为计算机软件程序,例如,本申请实施方式包括一种计算机程序产品,其包含有形地包含在机器可读介质上的计算机程序,计算机程序包含用于执行流程图所示的步骤的程序代码,程序代码可包括对应执行本申请提供的方法中的步骤对应的指令。

[0190] 在这样的实施方式中,该计算机程序可以通过通信部分1209从网络上被下载及安装,和/或从可拆卸介质1211被安装。在该计算机程序被中央处理单元(CPU)1201执行时,执行本申请中记载的实现上述相应步骤的指令。

[0191] 在一个或多个可选实施方式中,本公开实施例还提供了一种计算机程序产品,用于存储计算机可读指令,所述指令被执行时使得计算机执行上述任意实施例中所述的三维场景分析方法或者基于三维场景分析的控制方法。

[0192] 该计算机程序产品可以具体通过硬件、软件或其结合的方式实现。在一个可选例子中,所述计算机程序产品具体体现为计算机存储介质,在另一个可选例子中,所述计算机程序产品具体体现为软件产品,例如软件开发包(Software Development Kit,SDK)等等。

[0193] 在一个或多个可选实施方式中,本公开实施例还提供了另一种三维场景分析方法、基于三维场景分析的控制方法及其对应的装置和电子设备、计算机存储介质、计算机程序以及计算机程序产品,其中的三维场景分析方法或者基于三维场景分析的控制方法包括:第一装置向第二装置发送三维场景分析指示或者基于三维场景分析的控制指示,该指示使得第二装置执行上述任一可能的实施例中的三维场景分析方法或者基于三维场景分析的控制方;第一装置接收第二装置发送的三维场景分析结果或者基于三维场景分析的控制结果。

[0194] 在一些实施例中,该三维场景分析指示或者基于三维场景分析的控制指示可以具体为调用指令,第一装置可以通过调用的方式指示第二装置执行三维场景分析操作或者基

于三维场景分析的控制操作,相应地,响应于接收到调用指令,第二装置可以执行上述三维场景分析方法或者基于三维场景分析的控制方法中的任意实施例中的步骤和/或流程。

[0195] 应理解,本公开实施例中的“第一”、“第二”等术语仅仅是为了区分,而不应理解成对本公开实施例的限定。还应理解,在本公开中,“多个”可以指两个或两个以上,“至少一个”可以指一个、两个或两个以上。还应理解,对于本公开中提及的任一部件、数据或结构,在没有明确限定或者在前后文给出相反启示的情况下,一般可以理解为一个或多个。还应理解,本公开对各个实施例的描述着重强调各个实施例之间的不同之处,其相同或相似之处可以相互参考,为了简洁,不再一一赘述。

[0196] 可能以许多方式来实现本申请的方法和装置、电子设备以及计算机可读存储介质。例如,可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本申请的方法和装置、电子设备以及计算机可读存储介质。用于方法的步骤的上述顺序仅是为了进行说明,本申请的方法的步骤不限于以上具体描述的顺序,除非以其它方式特别说明。此外,在一些实施方式中,还可将本申请实施为记录在记录介质中的程序,这些程序包括用于实现根据本申请的方法的机器可读指令。因而,本申请还覆盖存储用于执行根据本申请的方法的程序的记录介质。

[0197] 本申请的描述,是为了示例和描述起见而给出的,而并不是无遗漏的或者将本申请限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言,是显然的。选择和描述实施方式是为了更好说明本申请的原理以及实际应用,并且使本领域的普通技术人员能够理解本申请实施例可以从而设计适于特定用途的带有各种修改的各种实施方式。

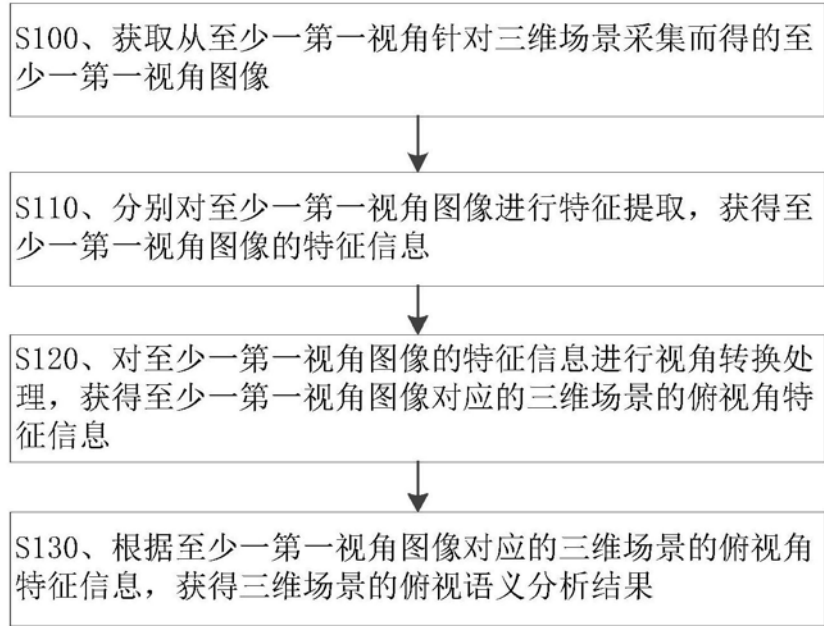


图1

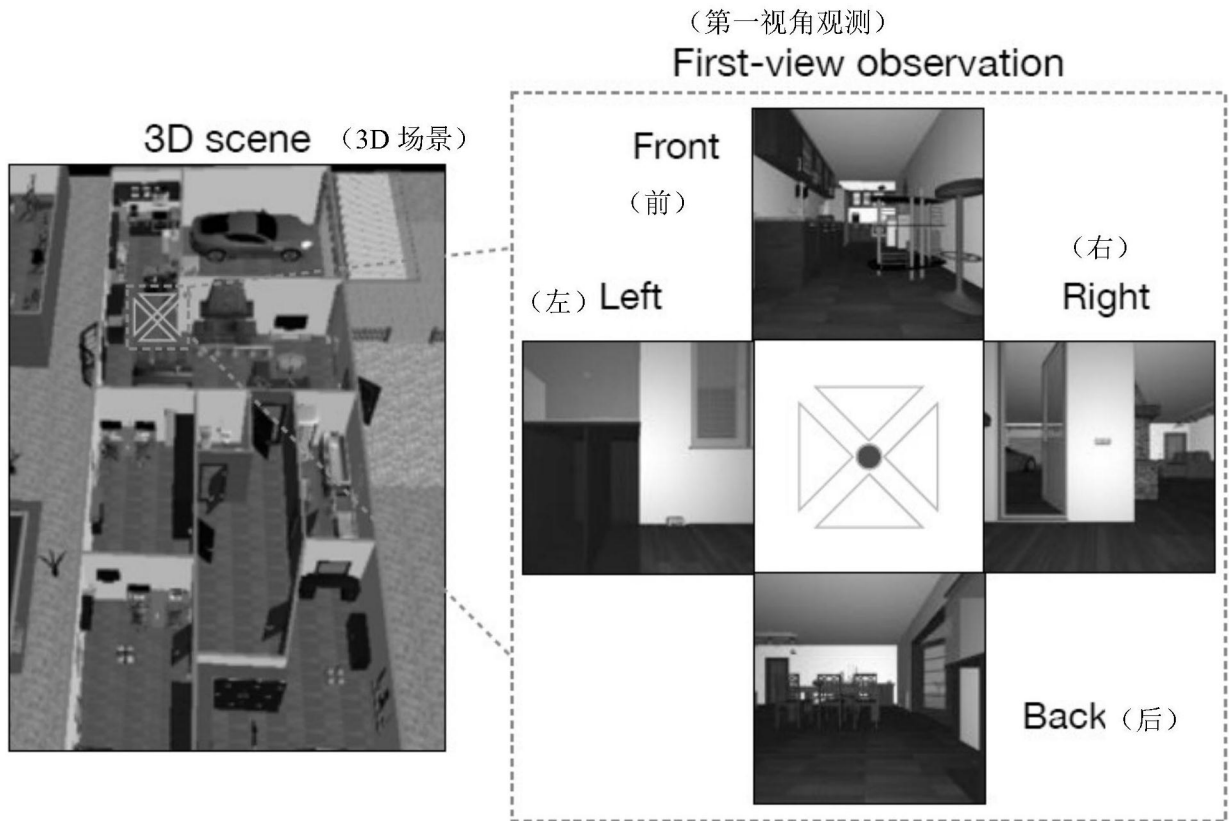


图2

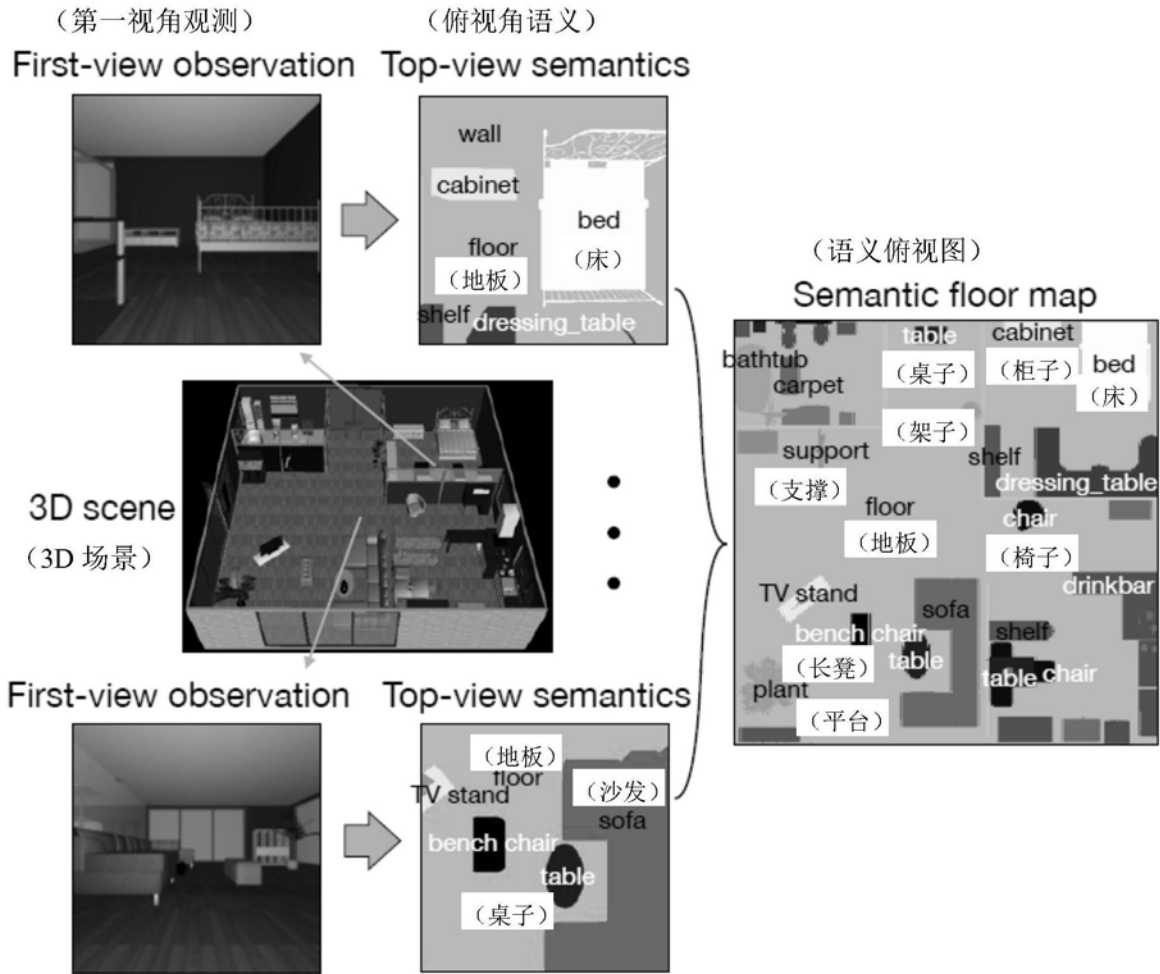


图3

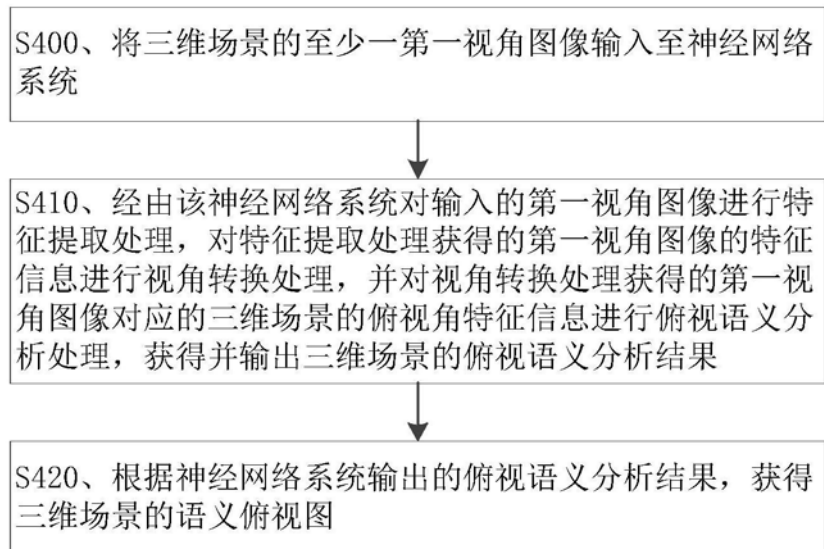


图4

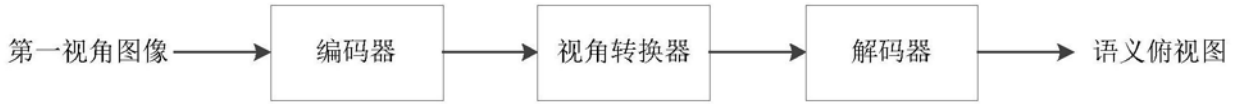


图5

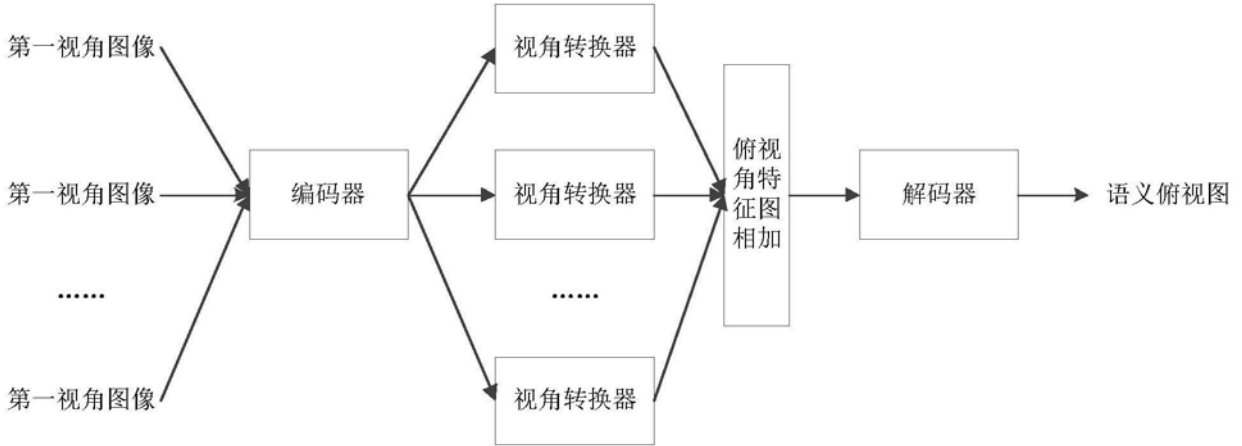


图6

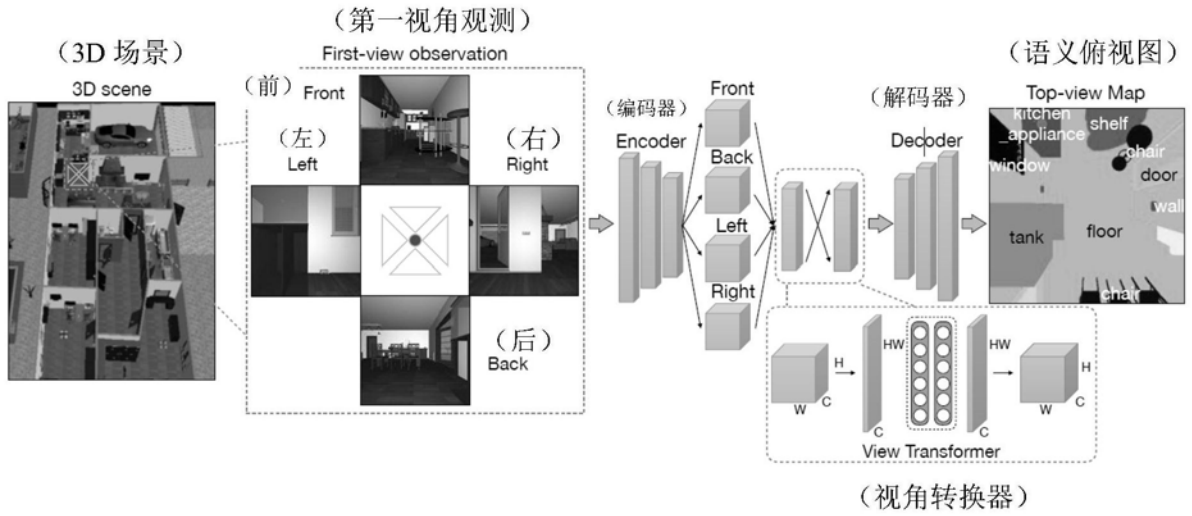


图7

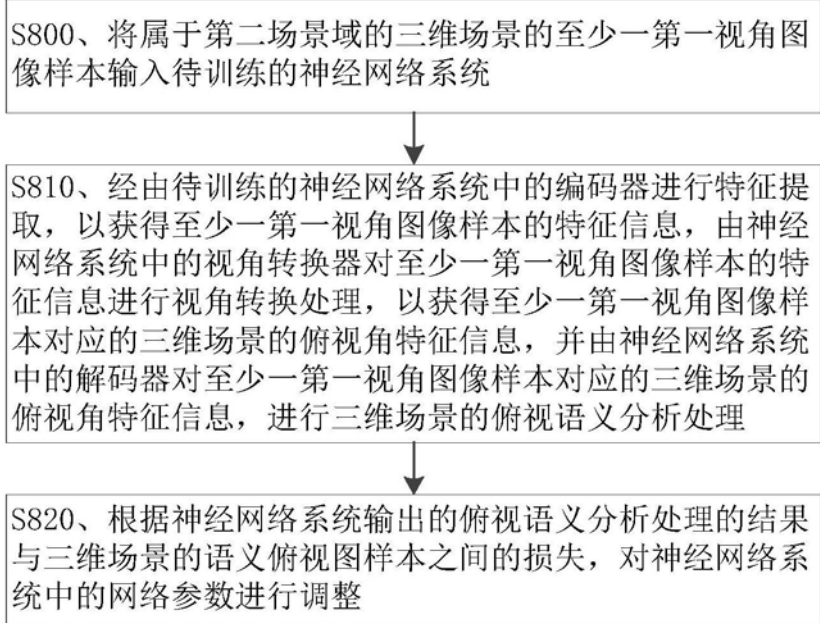


图8

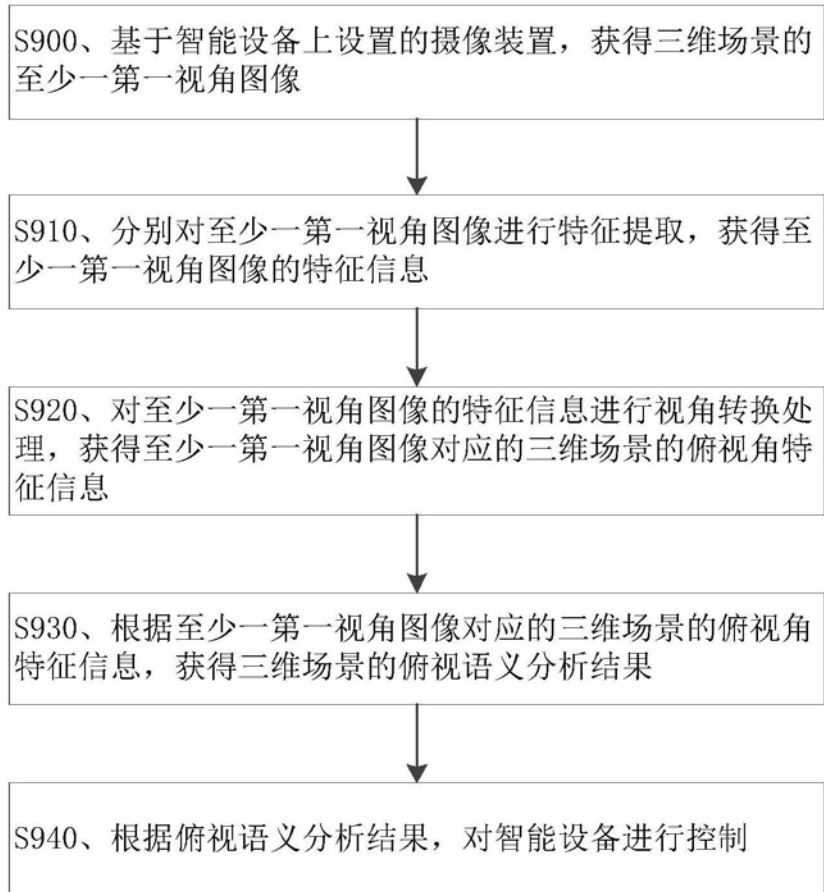


图9



图10



图11

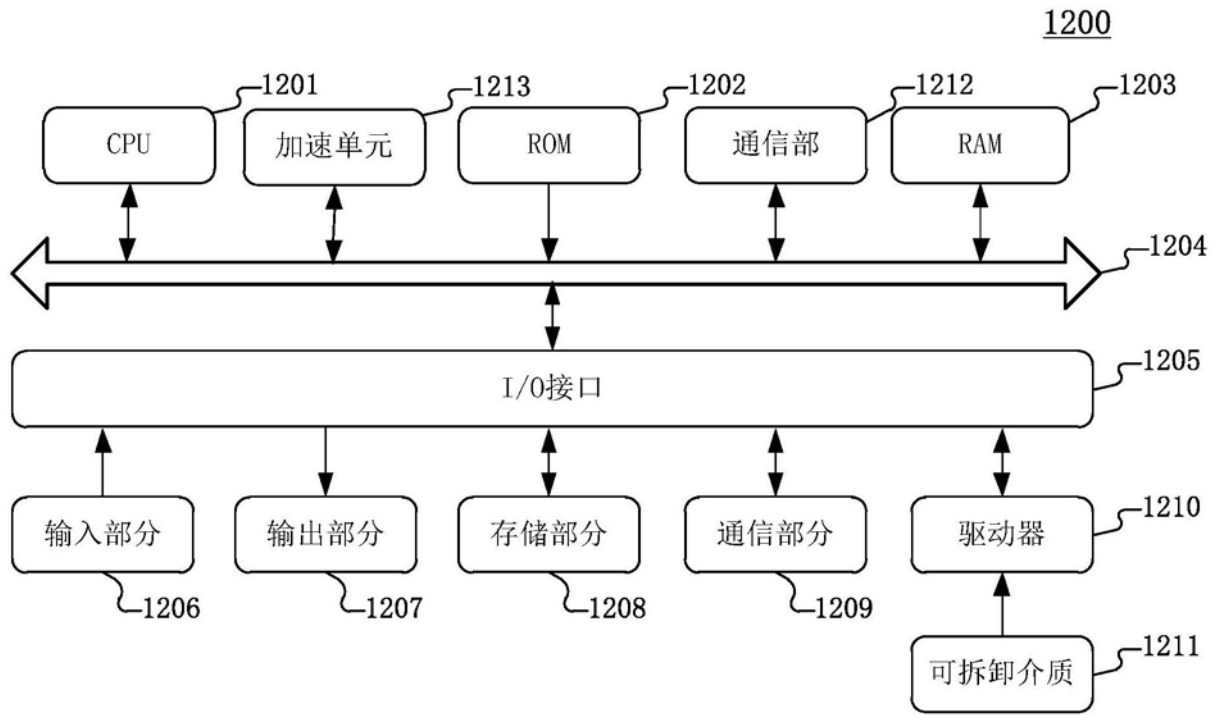


图12