



(12) 发明专利申请

(10) 申请公布号 CN 113984037 A

(43) 申请公布日 2022. 01. 28

(21) 申请号 202111157543.0

(22) 申请日 2021.09.30

(71) 申请人 电子科技大学长三角研究院(湖州)

地址 313099 浙江省湖州市西塞山路819号  
南太湖科技创新综合体B2幢8层

申请人 南京视道信息技术有限公司

(72) 发明人 范峥荣 方黎勇 杨方健 刘伟

李昊 周启南

(74) 专利代理机构 重庆天成卓越专利代理事务

所(普通合伙) 50240

代理人 王宏松

(51) Int. Cl.

G01C 21/00 (2006.01)

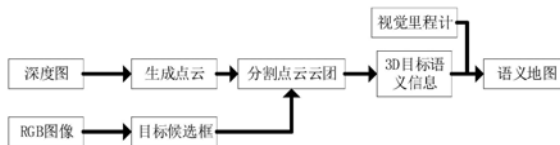
权利要求书1页 说明书6页 附图8页

(54) 发明名称

一种基于任意方向目标候选框的语义地图构建方法

(57) 摘要

本发明提出了一种基于任意方向目标候选框的语义地图构建方法,包括:S1,采用相机获取深度图像,再由深度图像获得原始点云数据;S2,采用相机获得RGB彩色图像;S3,针对RGB彩色图像,设计任意方向目标检测器,对所获得图像进行检测,获得目标的任意方向候选框;S4,将S3得到的任意方向候选框投影到步骤S1中的点云中,进行点云的框取,并进行后处理,获得较为规整的点云团;并完成框取点云与目标类别的数据关联,得到3D目标的语义信息;S5,利用视觉里程计信息将点云坐标转换为世界坐标系下的坐标并进行存储,即可保存点云语义地图。本发明能理解和利用环境中物体级别的高维语义信息进行定位、建图和路径规划,拥有对环境的全局理解能力。



1. 一种基于任意方向目标候选框的语义地图构建方法,其特征在於,包括以下步骤:

S1,采用相机一获取深度图像,再由深度图像获得原始点云数据;

S2,采用相机二获得RGB彩色图像;

S3,针对RGB彩色图像,设计基于改进YOLO神经网络的任意方向目标检测器,对所获得图像进行检测,获得目标的任意方向候选框;

S4,将S3得到的任意方向候选框投影到步骤S1中的点云中,进行点云的框取,并将框取的点云进行后处理,获得较为规整的点云团,即使用候选框信息进行点云分割;并完成框取点云与目标类别的数据关联,得到3D目标的语义信息;

S5,完成单帧点云分割的基础上,利用视觉里程计信息将点云坐标转换为世界坐标系下的坐标并进行存储,即可保存点云语义地图。

2. 根据权利要求1所述的一种基于任意方向目标候选框的语义地图构建方法,其特征在於,所述S1中的相机一采用RGB-D相机;所述S2中的相机二采用RGB-D相机或彩色相机。

3. 根据权利要求1所述的一种基于任意方向目标候选框的语义地图构建方法,其特征在於,所述S3中的YOLO神经网络包括主干网络,主干网络为一个CSP模块和4个CBL模块;

还包括:将顶层粗粒度特征和经过CSP网络模块的细粒度特征通过route层融合,最后经过卷积层扩充通道,连接到Yolo层进行特征解码;通过减小骨干网络的深度、进行粗粒度的特征融合进而减小网络的参数量和计算量。

4. 根据权利要求1所述的一种基于任意方向目标候选框的语义地图构建方法,其特征在於,所述S3中的目标的任意方向候选框包括:

使用标准化候选框和五参数法,将目标的groundtruth和候选框通过任意方向各自的角度变换成为轴对齐的框,再平移到原点处。

5. 根据权利要求4所述的一种基于任意方向目标候选框的语义地图构建方法,其特征在於,还包括:

在损失函数中增加预测角度的损失,对角度信息进行回归。

6. 根据权利要求1所述的一种基于任意方向目标候选框的语义地图构建方法,其特征在於,所述S4中的后处理包括:

通过相机内参矩阵将像素坐标转换为相机坐标系下的坐标,转换公式如下:

$$\begin{cases} x_c = (u - pp_x) * z / f_x \\ y_c = (v - pp_y) * z / f_y \end{cases}$$

其中, $x_c$ 为物体在相机坐标系下的实际横坐标;

$u$ 为像素的横坐标;

$pp_x$ 为深度相机在X轴方向的主点偏移;

$z$ 为通过深度图获取的物体的深度距离信息;

$f_x$ 为深度相机在X轴方向的焦距;

$y_c$ 为物体在相机坐标系下的实际纵坐标;

$v$ 为像素的纵坐标;

$pp_y$ 为深度相机在Y轴方向的主点偏移;

$f_y$ 为深度相机在Y轴方向的焦距。

## 一种基于任意方向目标候选框的语义地图构建方法

### 技术领域

[0001] 本发明涉及机器人智能感知领域,尤其涉及基于任意方向目标候选框的语义地图构建方法。

### 背景技术

[0002] 工业技术发展迅猛,移动机器人在生活中已经是屡见不鲜,例如送餐机器人、物流机器人、移动服务类机器人。这些机器人均实现了基本的建图、定位、导航等任务,但是这些机器人在大场景、跨区域、弱特征、高动态的场景下真的能够完成更复杂的任务吗?地面移动机器人经常面临复杂的、非结构化的场景。因此其核心技术涉及环境感知和建模能力、目标识别及自主任务规划能力、路径规划和决策能力。其中,环境感知和建模就如同人的眼睛获取外界图像一样,对机器人而言是至关重要的。目前的移动机器人普遍无法做到理解和利用环境中物体级别的高维语义信息,只能利用环境中的点、线、面等低维信息进行定位、建图和路径规划,缺乏对环境的全局理解能力。

### 发明内容

[0003] 本发明旨在至少解决现有技术中存在的技术问题,特别创新地提出了一种基于任意方向目标候选框的语义地图构建方法。

[0004] 为了实现本发明的上述目的,本发明提供了一种基于任意方向目标候选框的语义地图构建方法,包括以下步骤:

[0005] S1,采用相机一获取深度图像,再由深度图像获得原始点云数据;

[0006] S2,采用相机二获得RGB彩色图像;相机一采用RGB-D相机;相机二采用RGB-D相机或彩色相机,当其相机一和相机二均采用RGB-D相机时,此时仅有一个相机,即共用同一相机。

[0007] S3,针对RGB彩色图像,设计基于改进YOLO神经网络的任意方向目标检测器,对所获得图像进行检测,获得目标的任意方向候选框;

[0008] S4,将S3得到的任意方向候选框投影到步骤S1中的点云中,进行点云的框取,并将框取的点云进行后处理,获得较为规整的点云团,即使用候选框信息进行点云分割;并完成框取点云与目标类别的数据关联,得到3D目标的语义信息;

[0009] 这样即可完成单帧点云的语义提取、点云分割和数据关联;

[0010] S5,完成单帧点云分割的基础上,利用视觉里程计信息将点云坐标转换为世界坐标系下的坐标并进行存储,即可保存点云语义地图。

[0011] 进一步地,所述S3中的YOLO神经网络包括主干网络,主干网络为一个CSP模块和4个CBL模块;

[0012] 还包括:将顶层粗粒度特征和经过CSP网络模块的细粒度特征通过route层融合,最后经过卷积层扩充通道,连接到Yo1o层进行特征解码;通过减小骨干网络的深度、进行粗细粒度的特征融合进而减小网络的参数量和计算量。

[0013] 进一步地,所述S3中的目标的任意方向候选框包括:

[0014] 使用标准化候选框和五参数法,将目标的groundtruth和候选框通过任意方向各自的角度变换成为轴对齐的框,再平移到原点处。

[0015] 这时两个矩形框都变成了轴对齐框,容易计算其交并比IOU。

[0016] 进一步地,还包括:

[0017] 在损失函数中增加预测角度的损失,对角度信息进行回归。

[0018] 进一步地,所述S4中的后处理包括:

[0019] 通过相机内参矩阵将像素坐标转换为相机坐标系下的坐标,转换公式如下:

$$[0020] \quad \begin{cases} x_c = (u - pp_x) * z / f_x \\ y_c = (v - pp_y) * z / f_y \end{cases}$$

[0021] 其中 $x_c, y_c$ 分别为物体在相机坐标系下的实际横坐标和纵坐标, $u, v$ 分别为像素的横坐标与纵坐标, $f_x, f_y$ 分别为深度相机在X轴、Y轴方向的焦距, $pp_x, pp_y$ 分别为深度相机在X轴、Y轴方向的主点偏移, $z$ 为通过深度图获取的物体的深度距离信息。

[0022] 本发明还公开了一种基于任意方向目标候选框的语义地图构建系统,包括机器人及与所述机器人无线通讯的云平台,机器人包括底座,在底座底部设置有用于行走用的行走装置,在底座顶部设置有用于升降的升降装置,在升降装置上设置有竖向旋转平台,在竖向旋转平台上固定安装有相机一或/和相机二,相机一与相机二的镜头方向相同;

[0023] 还包括设置在竖向旋转平台内的PCB电路板,在PCB电路板上设置有控制器和无线数据模块,控制器的无线数据连接端与无线数据模块的无线数据连接端相连;

[0024] 第一相机数据端与控制器的第一相机数据端相连,第二相机数据端与控制器的第二相机数据端相连,行走装置的行走控制端与控制器的行走控制端相连,升降装置的升降控制端与控制器的升降控制端相连,竖向旋转平台的旋转控制端与控制器的旋转控制端相连;

[0025] 机器人根据云平台发送的移动指令进行移动,将相机一和相机二通过上下升降以及旋转拍摄的图像数据传输至云平台。

[0026] 在本发明的一种优选实施方式中,当其相机一和相机二共用一个相机时,包括机器人及与所述机器人无线通讯的云平台,机器人包括底座,在底座底部设置有用于行走用的行走装置,在底座顶部设置有用于升降的升降装置,在升降装置上设置有竖向旋转平台,在竖向旋转平台上固定安装有相机;

[0027] 还包括设置在竖向旋转平台内的PCB电路板,在PCB电路板上设置有控制器和无线数据模块,控制器的无线数据连接端与无线数据模块的无线数据连接端相连;

[0028] 机器人根据云平台发送的移动指令进行移动,将相机一通过上下升降以及旋转拍摄的图像数据传输至云平台。将拍摄的图像数据传输至云平台的方法为:

[0029] S1-1,将待上传至云平台的图像数据进行压缩处理;降低其传输时长;得到其压缩图像包;对得到的压缩图像包进行解压缩处理,得到其解压图像数据,判断其解压图像数据与待上传至云平台的图像数据是否一致;

[0030] 若解压图像数据与待上传至云平台的图像数据一致,则压缩图像包为待上传图像包,执行步骤S1-2;

[0031] 若解压图像数据与待上传至云平台的图像数据不一致,则此时解压图像数据文件数量多于待上传至云平台的图像数据,将压缩图像包中多余的文件剔除后,得到其待上传图像包;

[0032] S1-2,对待上传图像包进行计算上传码,其上传码的计算方法为:

[0033] Upload code=MD5 function[upload Image package,N],

[0034] 其中,MD5 function[,]表示哈希函数,优选采用MD5算法;

[0035] upload Image package表示待上传图像包;

[0036] N表示执行哈希函数迭代次数;N为大于或者等于1的正整数;优选N取4。

[0037] 当N取1时,执行一次MD5算法算法,

[0038] 
$$\text{Upload code}=\text{MD5 function}[\text{upload Image package},1],$$
  

$$=U_1,$$

[0039]  $U_1=\text{MD5 function}[\text{upload Image package}],$

[0040] 当N取2时,执行两次MD5算法算法,

[0041] 
$$\text{Upload code}=\text{MD5 function}[\text{upload Image package},2],$$
  

$$=U_2,$$

[0042] 即是, $U_1=\text{MD5 function}[\text{upload Image package}]$

[0043]  $U_2=\text{MD5 function}[U_1];$

[0044] 当N取3时,执行三次MD5算法算法,

[0045] 
$$\text{Upload code}=\text{MD5 function}[\text{upload Image package},3],$$
  

$$=U_3,$$

[0046] 即是, $U_1=\text{MD5 function}[\text{upload Image package}]$

[0047]  $U_2=\text{MD5 function}[U_1];$

[0048]  $U_3=\text{MD5 function}[U_2];$

[0049] 当N取4时,执行四次MD5算法算法,

[0050] 
$$\text{Upload code}=\text{MD5 function}[\text{upload Image package},3],$$
  

$$=U_4,$$

[0051] 即是, $U_1=\text{MD5 function}[\text{upload Image package}]$

[0052]  $U_2=\text{MD5 function}[U_1];$

[0053]  $U_3=\text{MD5 function}[U_2];$

[0054]  $U_4=\text{MD5 function}[U_3];$

[0055] 当N取5时,执行五次MD5算法算法,

[0056] 
$$\text{Upload code}=\text{MD5 function}[\text{upload Image package},3],$$
  

$$=U_5,$$

[0057] 即是, $U_1=\text{MD5 function}[\text{upload Image package}]$

[0058]  $U_2=\text{MD5 function}[U_1];$

- [0059]  $U_3 = \text{MD5 function}[U_2]$ ;
- [0060]  $U_4 = \text{MD5 function}[U_3]$ ;
- [0061]  $U_5 = \text{MD5 function}[U_4]$ ;
- [0062] ……。
- [0063] Upload code表示上传码;
- [0064] S1-3,将待上传图像包和待上传图像包对应的上传码上传至云平台,云平台接收到上传图像包和上传码后,分别对其重命名为接收图像包和接收码,对接收图像包进行对比码进行计算:
- [0065]  $\text{Receive code} = \text{MD5 function}[\text{Receive image package}, N]$ ,
- [0066] 其中,Receive image package表示接收图像包;
- [0067] Receive code表示对比码;
- [0068] 若对比码与接收码一致,则对接收图像包进行解压缩;
- [0069] 若对比码与接收码不一致,则返回步骤S1-2。若3次传输仍然不正确,传输其余图像数据。
- [0070] 综上所述,由于采用了上述技术方案,本发明的有益效果是:
- [0071] 能理解和利用环境中物体级别的高维语义信息进行定位、建图和路径规划,拥有对环境的全局理解能力。
- [0072] 本发明的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

## 附图说明

- [0073] 本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解,其中:
- [0074] 图1是本发明的逻辑结构示意图。
- [0075] 图2是本发明改进的YOLO目标识别神经网络结构图示意图。
- [0076] 图3是本发明目标的任意方向候选框生成示意图。
- [0077] 图4是本发明标准化候选框计算IOU方式的示意图。
- [0078] 图5是本发明映射后的点云数据示意图。
- [0079] 图6是本发明与固定方向点云分割的对比示意图。
- [0080] 图7是本发明开放场景下的点云分割效果图。
- [0081] 其中图7(a)为点云分割效果图;图7(b)为目标检测原图示意图。
- [0082] 图8是本发明TUM数据集下的典型图片示意图。
- [0083] 图9是本发明rgb\_dataset\_freiburg1\_desk上的语义地图实验结果示意图。
- [0084] 其中图9(a),图9(b)为不同方向的视图;图9(c)为语义标签。
- [0085] 图10是本发明rgb\_dataset\_freiburg2\_desk上的语义地图实验结果。
- [0086] 其中图10(a),图10(b)为不同方向的视图;图10(c)为语义标签。

## 具体实施方式

- [0087] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出,其中自始至终

相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,仅用于解释本发明,而不能理解为对本发明的限制。

[0088] 为解决上述问题,本发明公开了一种基于任意方向目标检测候选框的语义地图建模方法。该方法在构建任意方向目标检测网络的基础上,生成任意方向的2D目标候选框,根据任意方向的候选框信息在三维点云中框取合适的点云,有效的滤除多余的点云信息,再将框取的点云进行滤波等后处理计算出点云3D包围框的信息,完成点云分割和数据关联,从而获得这些地图数据的语义地图。

[0089] 本发明提供了一种基于任意方向目标检测候选框的语义地图建模方法,如图1所示,包括:

[0090] 步骤一:采用RGB-D相机获取深度图像,获得原始点云数据。

[0091] 步骤二:采用RGB-D相机获得RGB彩色图像。

[0092] 步骤三:针对RGB彩色图像,设计基于改进YOLO神经网络的任意方向目标检测器,对所获得图像进行检测,获得目标的任意方向候选框,如图3所示。

[0093] 步骤四:把RGB图像中的任意方向目标检测器投影到步骤一中的点云中,进行点云的框取,并将框取的点云进行后处理,获得较为规整的点云团,完成数据关联。这样即可完成单帧点云的语义提取、点云分割和数据关联,如图4所示。

[0094] 步骤五:完成单帧点云分割的基础上,利用视觉里程计信息将点云坐标转换为世界坐标系下的坐标并进行存储,即可保存点云语义地图,如图5所示。

[0095] 其中,图1显示了该发明的整体技术框架和整体步骤。本章采用2D图像目标检测候选框为主进行点云分割的思路,首先使用RGB图像基于Yolo的目标检测器输出常规的轴对齐候选框,再使用深度图像生成点云数据,然后将候选框投影到3D点云中框取点云。对框取到的有序点云进行后处理,获得可靠的点云团,并完成框取点云与目标类别的数据关联,得到3D目标的语义信息。在完成单帧点云分割的基础上,利用视觉里程计信息将点云坐标转换为世界坐标系下的坐标并进行存储,即可保存点云语义地图。

[0096] 图2示出了基于YOLO网络的目标识别。该网络的主干网络为一个CSP模块和4个CBL模块。与原始的网络相比,改进后的特征提取器网络缩短了主干网的深度;减少了CSP网络模块的同时,将顶层粗粒度特征和经过CSP网络模块的细粒度特征通过route层融合,最后经过卷积层扩充通道,连接到Yolo层进行特征解码。通过减小骨干网络的深度、进行粗细粒度的特征融合进而减小网络的参数量和计算量。

[0097] 图3示出了目标的任意方向候选框生成。使用标准化候选框和五参数法,将目标的groundtruth和候选框通过任意方向各自的角度变换成为轴对齐的框,再平移到原点处。这时两个矩形框都变成了轴对齐框,容易计算其交并比IOU。如图4所示,为标准化任意方向框的示意图。其次考虑在损失函数中增加预测角度的损失,对角度信息进行回归。

[0098] 在图4中,黑色框为目标的groundtruth (groundtruth为输入数据的标注信息),红色虚线框为神经网络的推断输出。首先将目标的groundtruth通过旋转平移到坐标原点处,对神经网络训练时的输出候选框也经过旋转平移到坐标原点处。在坐标原点处两个候选框形成交叠区域或者不交叠。橘黄色为交叠区域,即可以方便计算出IOU,使用这个方法只考虑了矩形框是否重叠,丢失了角度的信息,所以要在损失函数中增加角度损失项。

[0099] 图5为单帧点云的语义提取、点云分割和数据关联。将候选框投影到3D点云即原始

点云数据中,使用候选框信息进行点云分割。本文中使用RGB-D相机,能够通过深度图获得像素点的深度信息,通过相机内参矩阵将像素坐标转换为相机坐标系下的坐标。转换公式如式4-1所示。

$$[0100] \quad \begin{cases} x_c = (u - pp_x) * z / f_x \\ y_c = (v - pp_y) * z / f_y \end{cases} \quad (4-1)$$

[0101] 式4-1中 $x_c, y_c$ 为物体在相机坐标系下的实际坐标; $(u, v)$ 为物体的像素坐标; $f_x, f_y$ 为相机的焦距,理论上两者相等; $(pp_x, pp_y)$ 为相机的主点偏移; $z$ 为通过深度图获取的物体的深度距离信息。其中 $x_c, y_c$ 分别为物体在相机坐标系下的实际横坐标和纵坐标, $u, v$ 分别为像素的横坐标与纵坐标, $f_x, f_y$ 分别为深度相机在X轴、Y轴方向的焦距; $(pp_x, pp_y)$ 分别为深度相机在X轴、Y轴方向的主点偏移。

[0102] 图8是在TUM公开数据集下构建的语义地图构建,如图9所示,TUM数据集包含RGB-D相机即深度相机采集的彩色图像和深度图像以及通过IMU采集的数据。

[0103] 该数据集包含一系列的办公场景和其他场景。数据集的数据由RGB-D相机和IMU模块采集,包含彩色和深度图像,及对应的IMU数据。在`rgbd_dataset_freiburg1_desk`和`rgbd_dataset_freiburg2_desk`数据上测试,测试结果图和标签如图9和图10所示。

[0104] 尽管已经示出和描述了本发明的实施例,本领域的普通技术人员可以理解:在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由权利要求及其等同物限定。



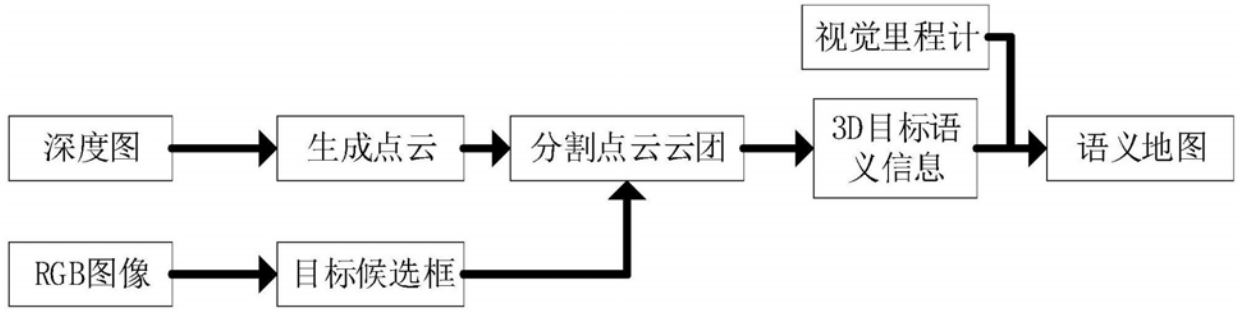


图1

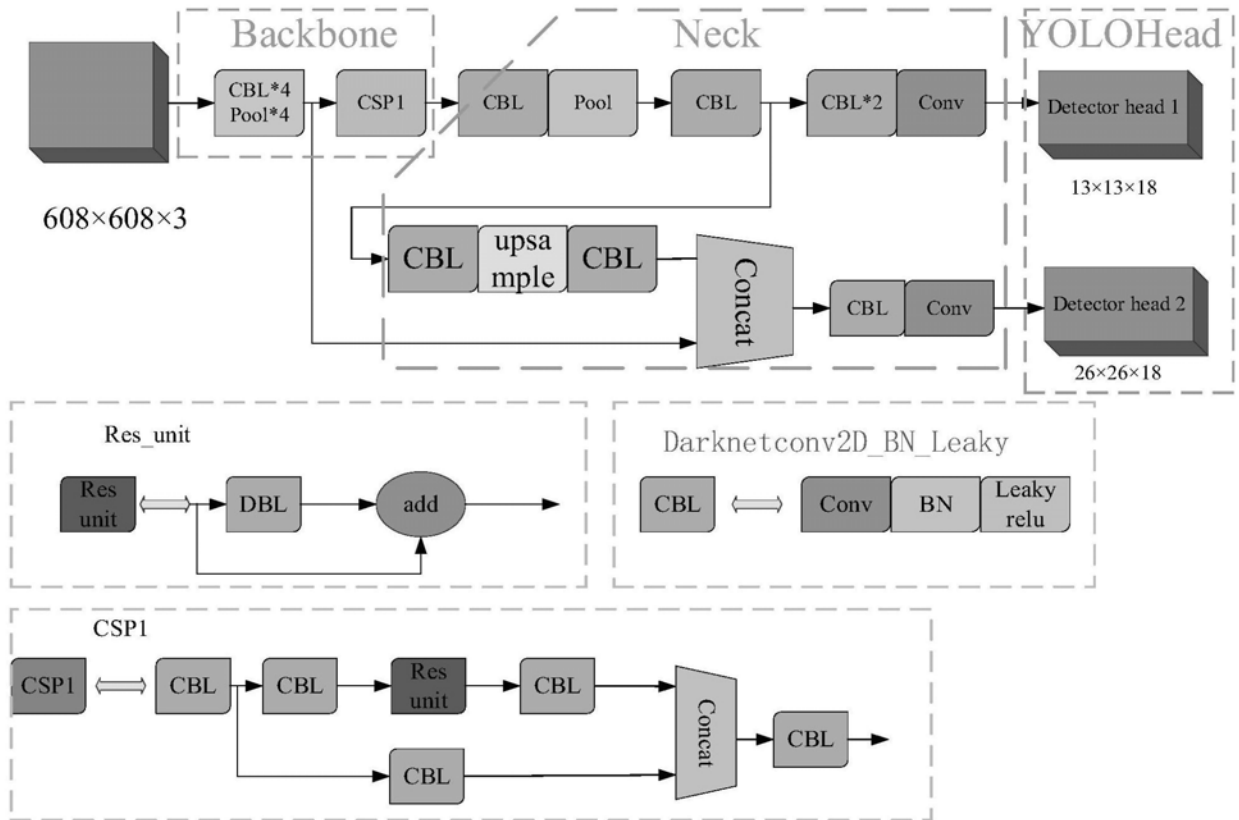


图2

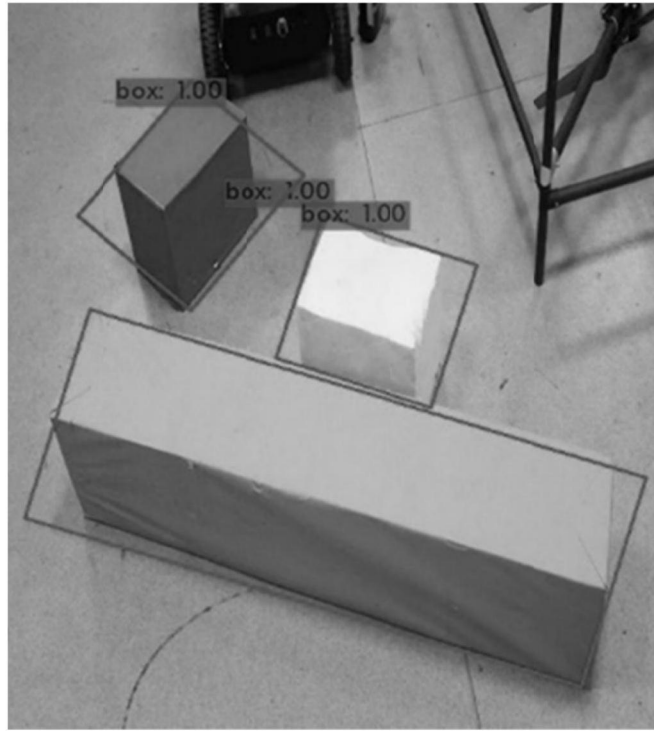


图3

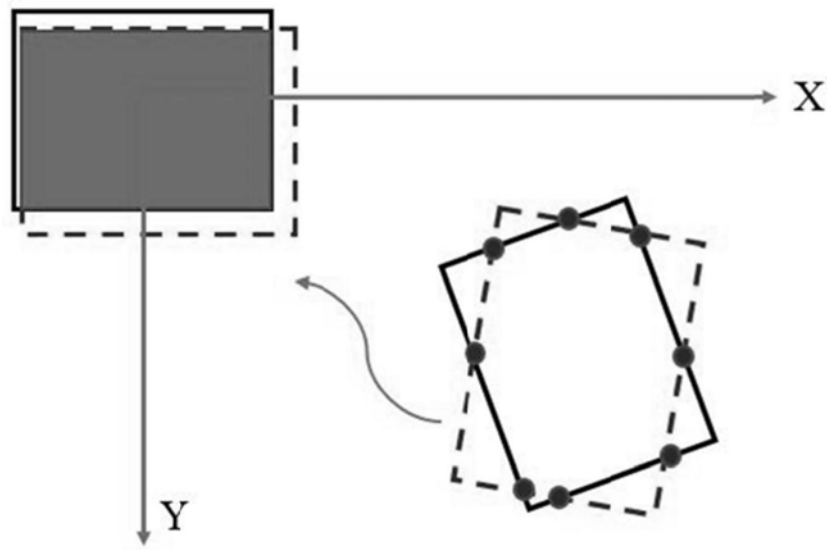


图4

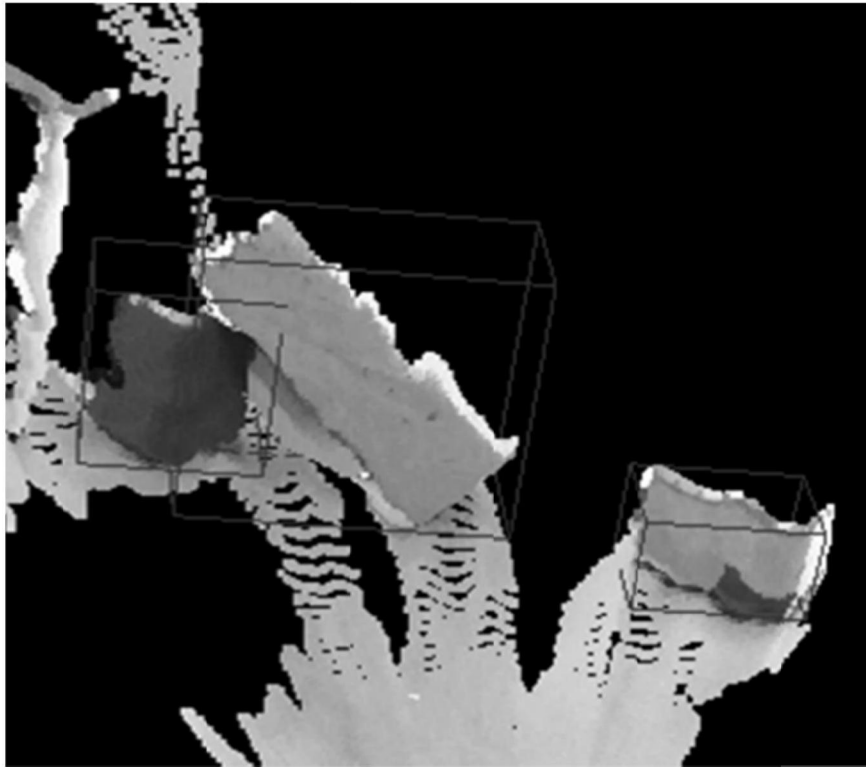


图5

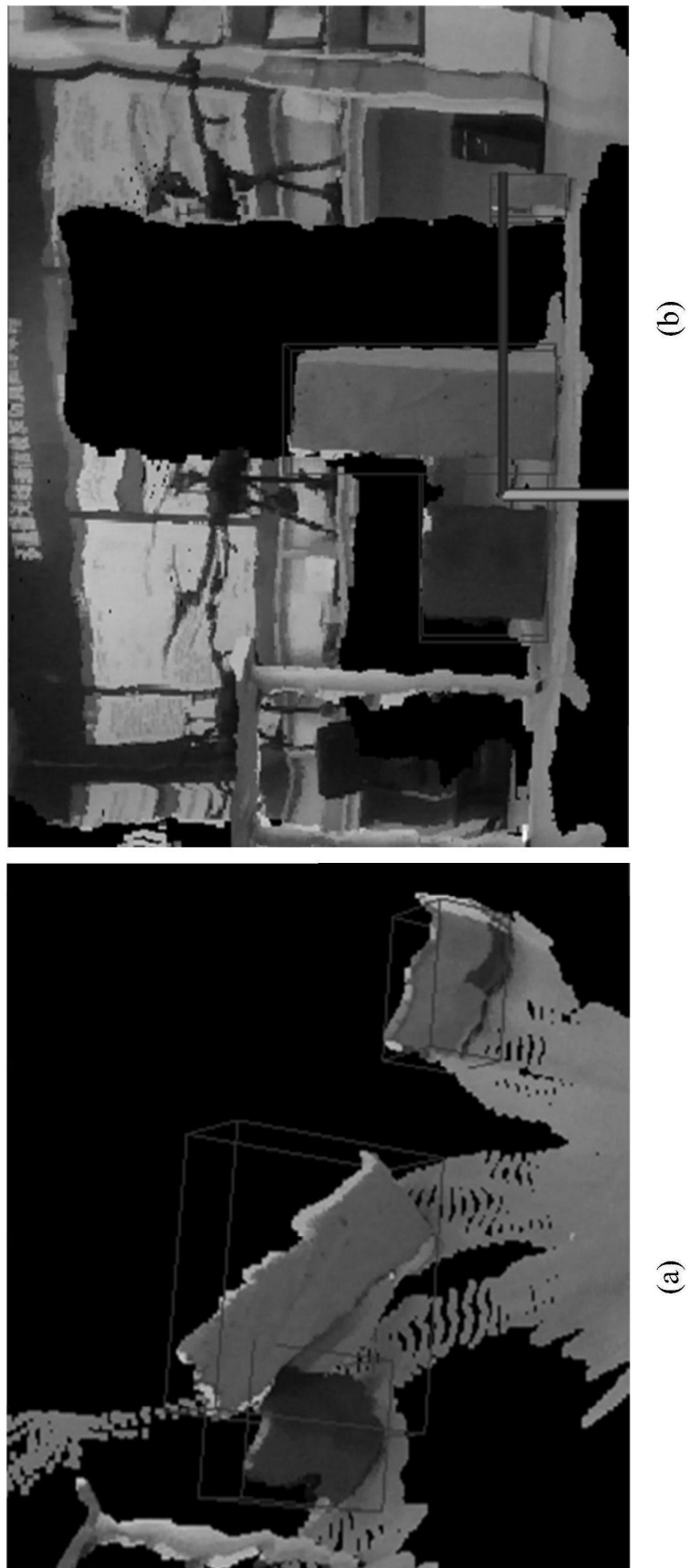


图6



图7

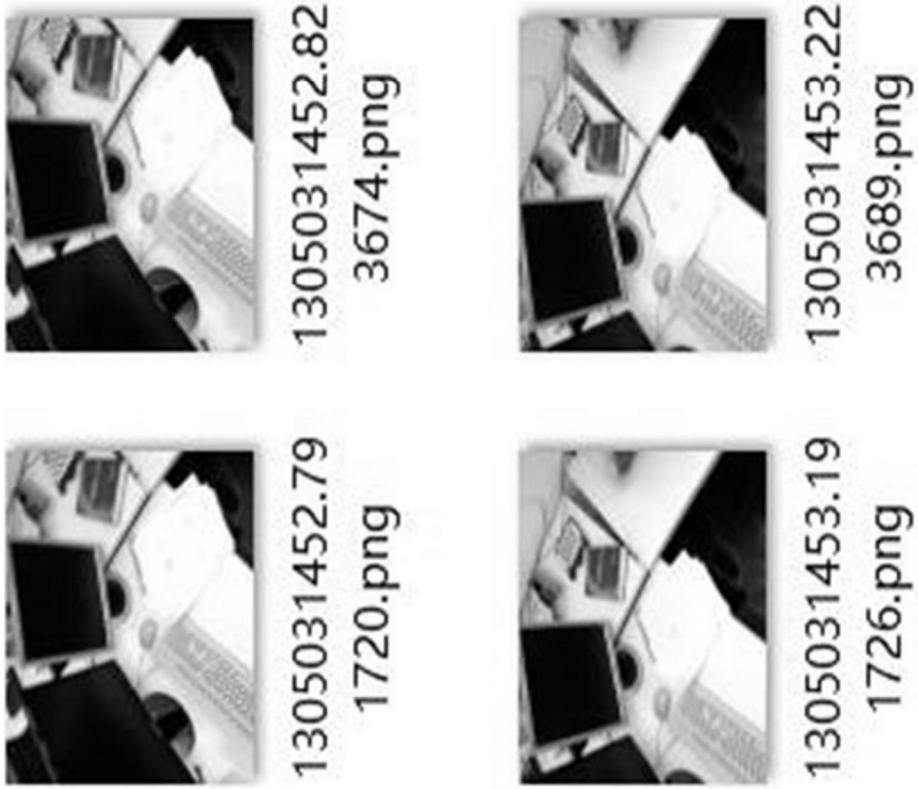


图8

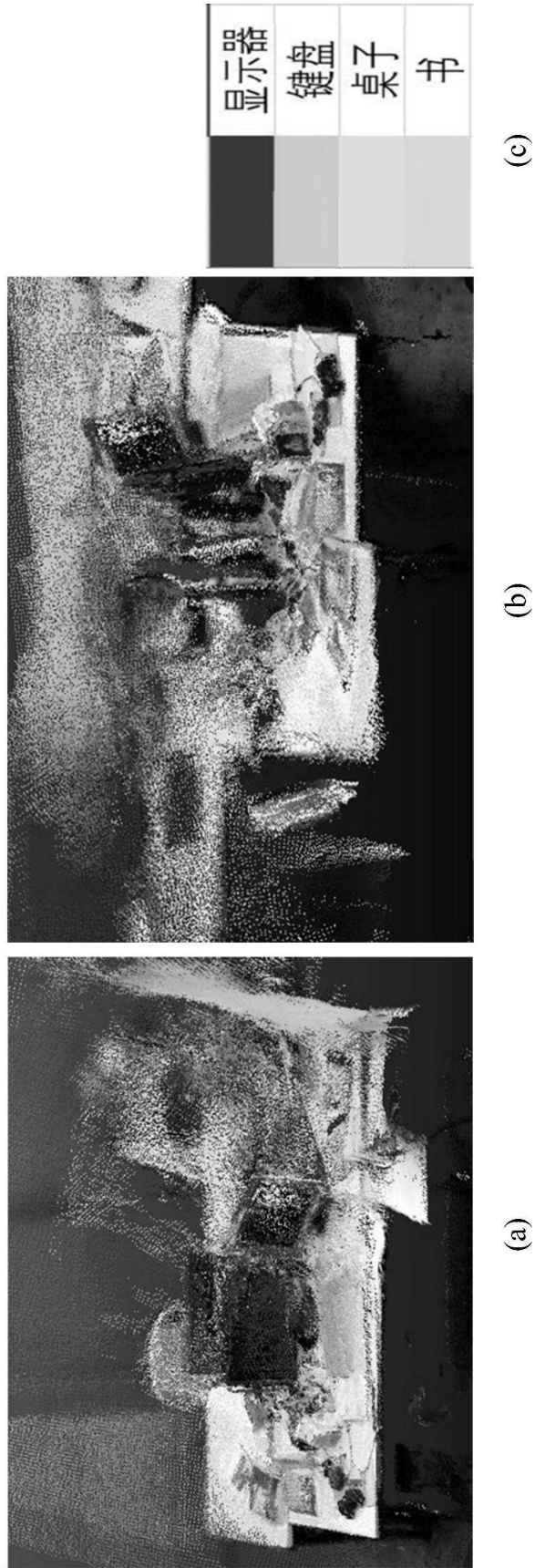


图9

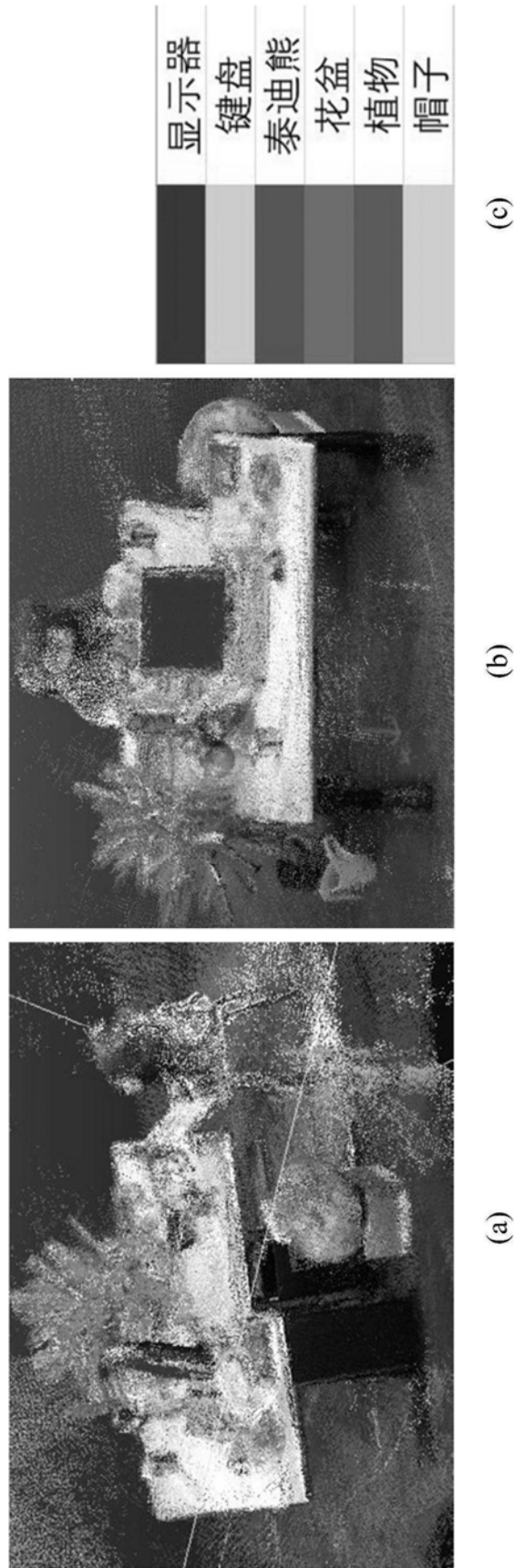


图10