



(12)发明专利申请

(10)申请公布号 CN 108909624 A

(43)申请公布日 2018. 11. 30

(21)申请号 201810452761.9

(22)申请日 2018.05.13

(71)申请人 西北工业大学

地址 710072 陕西省西安市友谊西路127号

(72)发明人 布树辉 冷鹏宇 韩鹏程

(74)专利代理机构 西北工业大学专利中心

61204

代理人 陈星

(51) Int. Cl.

B60R 1/00(2006.01)

B60W 30/08(2012.01)

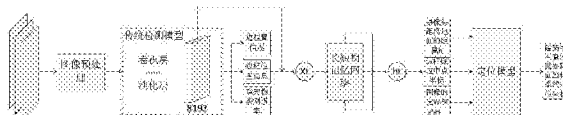
权利要求书2页 说明书11页 附图3页

(54)发明名称

一种基于单目视觉的实时障碍物检测和定位方法

(57)摘要

本发明提出一种基于单目视觉的实时障碍物检测和定位方法,只使用摄像头完成对障碍物的实时检测和定位(测距),将机器视觉和深度学习结合进行端对端的障碍物的检测,并定位出障碍物在真实世界坐标系下相对于车辆的位置。为了提高检测精度,本发明提出基于时空序列信息的长短期记忆网络,将通过检测模型获得的障碍物边框位置信息和检测模型中第一层全连接层输出的M维向量组合的向量,再将这个向量输入到长短期记忆网络进行训练,网络既考虑了障碍物周边的空间信息,也考虑了过去帧的鲁棒视觉特征。通过对时空序列的学习,对障碍物检测定位获取更好的效果。



1. 一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:包括以下步骤:

步骤1:数据采集与处理:在车辆上安装摄像头并拍摄车辆前方若干张照片;对于每一张包含有障碍物的照片,人工标定出照片中障碍物,并根据人工标定结果,给每一张照片设定检测标签和定位标签;再将附有检测标签和定位标签的若干张照片作为模型的训练数据集;

步骤2:搭建并训练障碍物检测、长短期记忆网络和定位的联合模型:搭建网络模型,构造损失函数优化模型,将处理后的训练数据集输入至网络,结合BP算法更新网络权重参数联合训练,得到障碍物检测、长短期记忆网络和定位模型;

步骤3:障碍物的检测和定位:将摄像头固定在车辆之前的位置,将摄像头拍摄的视频输入到步骤2训练好的模型中,预测图像中障碍物的类别、障碍物边框、置信度,以及障碍物边框在真实世界坐标系下的位置信息。

2. 根据权利要求1所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:步骤1中数据采集与处理的具体步骤为:

步骤1.1:将单目摄像头安装在车辆上,确定摄像头距地面的高度 h 以及摄像头光轴轴线与竖直方向的俯仰角 γ_0 ;

步骤1.2:在不同的天气情况下,车辆在有障碍物的路面正常行驶,通过摄像头拍摄视频;从视频中获取若干张包含有障碍物的图片;

步骤1.3:人工用矩形边框对拍摄得到的包含有障碍物的图片中的障碍物的位置进行标定,得到边框的中心在图片坐标系中的坐标以及边框的宽和高;预先设定 C 类障碍物类别,采用one-hot编码方法,对障碍物的类别信息使用一个 C 维向量 \vec{c} 进行描述:当边框内的物体属于第 i 类障碍物时,类别向量 \vec{c} 的第 i 维为1,其余为0;

步骤1.4:对每张图片设定检测标签:将图片分为 $S*S$ 网格,对于每个网格建立一个数组 $(\vec{x}, \vec{y}, \vec{w}, \vec{h}, \vec{P}, \vec{c})$;

对于步骤1.3中矩形边框中心所在的网格,其数组取值为: \vec{x}, \vec{y} 分别为矩形边框中心与网格中心在图片坐标系 x 方向和 y 方向距离的差值的归一化结果, \vec{w}, \vec{h} 为矩形边框宽和高的归一化结果, $\vec{P} = [p]$, p 是置信度,代表网格含有物体的概率及定位准确度, \vec{c} 根据步骤1.3确定;对于其余网格,数组取值全为0;

所有 $S*S$ 网格的数组组成检测标签向量用于检测模型训练;

步骤1.5:对于步骤1.3中的边框坐标进行归一化处理,并与矩形边框宽和高的归一化结果组合,得到训练长短期记忆网络模型的边框标签 $(\vec{p}, \vec{q}, \vec{w}, \vec{h})$;

步骤1.6:记录步骤1.3中矩形边框底边中点在图像坐标系下的坐标 (a, b) ;再以摄像头光轴延长线交地面于 O 点定为路面坐标系中心原点,车辆前进方向为 Y 轴方向,路面平面为 XY 平面,记录与矩形边框底边中点对应的路面位置坐标 (\vec{A}, \vec{B}) ,作为训练数据集的定位标签。

3. 根据权利要求1所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:步骤2中搭建并训练障碍物检测、长短期记忆网络和定位的联合模型的具体过程为:

步骤2.1:采用卷积神经网络搭建障碍物检测模型:网络由16层卷积层,5层池化层,2层全连接层构成;模型输入是完整图像像素数据,输出结果是维度为 $S*S*(B*5+C)$ 的向量, B 为每个网格负责检测物体的个数;输出结果向量的含义为:每个网格检测出 B 个矩形边框,每个矩形边框位置信息包括了 $\vec{x}, \vec{y}, \vec{w}, \vec{h}$,每个矩形边框还具有置信度 P 信息,每个网格对应预测类别向量 c ,包括 C 个类别概率值,其中最大的为网格的类别预测结果;

根据检测模型结果得到障碍物边框信息,设定阈值 T ,只保留 $P \geq T$ 的边框,再根据非极大抑制原理来消除冗余的边框,最后获取余下边框的位置信息;

步骤2.2:搭建长短期记忆网络模型,分为输入层,隐藏层,输出层;输入层的输入分为:由当前图像经过检测模型得到的边框位置信息与检测模型中第一层全连接层输出的 M 维向量组合得到的向量,以及上一幅图像经长短期记忆网络模型的输出结果;输出层的输出是当前图像中障碍物边框新的位置信息;

步骤2.3:采用全连接神经网络搭建障碍物定位模型:定位模型采用三层网络,网络输入层节点为5,隐层节点数为20,输出层节点为2;其中5个输入分别是长短期记忆网络模型输出的向量中得到的障碍物边框底边中点坐标 (a, b) ,图像的宽 W 和高 H ,摄像头距地面的高度 h ;2个输出是障碍物边框在真实世界路面坐标系的位置坐标 (A, B) 。

4.根据权利要求1所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:训练步骤2中网络模型的具体过程为:

步骤3.1:初始化网络中的神经元权重参数;

步骤3.2:将采集得到的图片进行预处理,使所有训练样本的尺寸归一化为 $32S*32S*3$,对所有图像像素进行归一化,使图像的像素值在0到1之间;并对图片的曝光,色调,饱和,镜像做随机变化来扩大样本数量;将训练样本送入模型进行训练;

步骤3.3:通过优化损失函数,结合BP算法更新网络权重参数,迭代至与标签的误差降低到预设的阈值,结束训练。

一种基于单目视觉的实时障碍物检测和定位方法

技术领域

[0001] 本发明属于图像处理和机器视觉领域,涉及一种基于单目视觉的前方障碍物检测和定位方法,该方法是基于深度学习和时空序列信息对障碍物实时检测,并在真实世界坐标系下定位障碍物相对于车辆的位置。

背景技术

[0002] 21世纪车辆将进入一个崭新的信息时代,主要发展趋势体现在车辆的智能化。智能化车辆将不仅在高速公路上行驶,车载信息系统还将在信息高速公路上奔驰。

[0003] 车辆在人们带来方便的同时,随之而来的车辆安全问题也显而易见,那就是随着车辆的增多,交通事故的频繁发生,所导致的人员伤亡和财产损失数目惊人。根据交通部统计,2010年,全国因车辆交通事故死亡人数达到6.5万人。

[0004] 对于公路交通事故的分析表明,超速,占道,疲劳,酒驾等驾驶行为是导致交通事故高发的主要原因,而交通事故中以碰撞为主,超过65%的车辆碰撞是车辆追尾,其余较多是侧面相撞,所以,车辆自动避障技术的研究变得极其重要。据研究调查:如果在驾驶员提前0.5秒意识到有事故危险,则可减少大约60%的交通事故;如果能提前1秒,则可减少90%的交通事故;而提前1.5秒以上,驾驶员完全有能力去避免交通事故的发生。这就说明了如果能提前检测出车辆所处的环境中的障碍物或者前方车辆的位置,在遇到危险时能给出预警并自动采取恰当的措施(比如自动制动,自动转向等),将极大改善驾驶员在复杂条件下对车辆的控制能力,对减少甚至避免交通事故的发生有着非常重要的意义。因此,研究开发出具有提前预警功能并能给出相关控制操作的车辆自动避障系统对进一步提升车辆安全有着非常重要的意义,已经成为了智能车领域的研究热点和重要课题。

[0005] 而障碍物定位(测距)作为智能车自动避障的关键技术之一,不可避免的成为其中的一个研究热点,如果能正确的检测出前方障碍物的类别并定位其位置,基于此便能在综合分析后,通过对各种危险情况的及早发现,及时提醒驾驶员,让驾驶员有充分的时间做出合理的处理,或者是启动智能车的自动避障系统,尽可能的减少交通事故的发生。因此,障碍物定位(测距)技术的研究具有很高的工程价值,对车辆自动避障系统有很关键的意义。

[0006] 目前国内外的障碍物定位(测距)主要研究现状是利用激光雷达,摄像头,毫米波雷达这三类传感器感知前方物体,又或者是融合摄像头和雷达技术,相结合了摄像头的识别能力和雷达的测距能力。比如,美国高级波导公司研发的红外激光全方位回转车辆防撞系统则可以进行360度全方位的障碍物扫描,能够识别出障碍物的方位,距离和碰撞时间等参数;美国EATON公司研发的EATON VORAD-300采用24.725GHz单脉冲雷达,可同时探测到主车正前方120m,左右偏移80m范围内的24个目标车辆,当检测有车辆距离小于预设的警戒距离值时,该系统发出警报。近年来,德国大陆特威斯公司开发出采用毫米波雷达和红外传感相结合技术,可探测出前方150m范围内的目标车辆与主车之间的车距和相对速度。

[0007] 车载摄像头在自动驾驶车辆中应用广泛且价格相对低廉,是最基本最常见的传感器。如果能提出一种基于摄像头实时而准确地检测出障碍物并且定位(测距)的算法,这项

技术就能让自动驾驶功能以相对便宜的价格添加到车辆上。仅配备摄像头就可以获得一个不错的入门级自动驾驶,这使得普通人也能负担得起,并配备在自己的车辆中,这在未来市场空间是非常可观的。

[0008] 所以本发明提出只采用摄像头,将机器视觉和深度学习结合,把检测和定位问题转换为回归问题,充分利用时空序列信息丰富的特点,联合训练出一个可以对障碍物的实时检测和定位(测距)的模型,做出高效准确的碰撞预警。

发明内容

[0009] 本发明提出一种基于单目视觉的实时障碍物检测和定位方法,本发明的特点在于只使用摄像头完成对障碍物的实时检测和定位(测距),即将机器视觉和深度学习结合进行端对端的障碍物的检测,并定位出障碍物在真实世界坐标系下相对于车辆的位置。为了提高检测精度,本发明提出基于时空序列信息的长短期记忆网络,将通过检测模型获得的障碍物边框位置信息和检测模型中第一层全连接层输出的M维向量组合的向量,再将这个向量输入到长短期记忆网络进行训练,网络既考虑了障碍物周边的空间信息,也考虑了过去帧的鲁棒视觉特征。通过对时空序列的学习,对障碍物检测定位获取更好的效果。

[0010] 本发明的技术方案为:

[0011] 所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:包括以下步骤:

[0012] 步骤1:数据采集与处理:在车辆上安装摄像头并拍摄车辆前方若干张照片;对于每一张包含有障碍物的照片,人工标定出照片中障碍物,并根据人工标定结果,给每一张照片设定检测标签和定位标签;再将附有检测标签和定位标签的若干张照片作为模型的训练数据集;

[0013] 步骤2:搭建并训练障碍物检测、长短期记忆网络和定位的联合模型:搭建网络模型,构造损失函数优化模型,将处理后的训练数据集输入至网络,结合BP算法更新网络权重参数联合训练,得到障碍物检测、长短期记忆网络和定位模型;

[0014] 步骤3:障碍物的检测和定位:将摄像头固定在车辆之前的位置,将摄像头拍摄的视频输入到步骤2训练好的模型中,预测图像中障碍物的类别、障碍物边框、置信度,以及障碍物边框在真实世界坐标系下的位置信息。

[0015] 进一步的优选方案,所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:步骤1中数据采集与处理的具体步骤为:

[0016] 步骤1.1:将单目摄像头安装在车辆上,确定摄像头距地面的高度 h 以及摄像头光轴轴线与竖直方向的俯仰角 γ_0 ;

[0017] 步骤1.2:在不同的天气情况下,车辆在有障碍物的路面正常行驶,通过摄像头拍摄视频;从视频中获取若干张包含有障碍物的图片;

[0018] 步骤1.3:人工用矩形边框对拍摄得到的包含有障碍物的图片中的障碍物的位置进行标定,得到边框的中心在图片坐标系中的坐标以及边框的宽和高;预先设定C类障碍物类别,采用one-hot编码方法,对障碍物的类别信息使用一个C维向量 \vec{e} 进行描述:当边框内的物体属于第 i 类障碍物时,类别向量 \vec{e} 的第 i 维为1,其余为0;

[0019] 步骤1.4:对每张图片设定检测标签:将图片分为 $S*S$ 网格,对于每个网格建立一个

数组 (x, y, w, h, P, c) ;

[0020] 对于步骤1.3中矩形边框中心所在的网格,其数组取值为: x, y 分别为矩形边框中心与网格中心在图片坐标系x方向和y方向距离的差值的归一化结果, w, h 为矩形边框宽和高的归一化结果, $P=1, P$ 是置信度,代表网格含有物体的概率及定位准确度, c 根据步骤1.3确定;对于其余网格,数组取值全为0;

[0021] 所有 $S*S$ 网格的数组组成检测标签向量用于检测模型训练;

[0022] 步骤1.5:对于步骤1.3中的边框坐标进行归一化处理,并与矩形边框宽和高的归一化结果组合,得到训练长短期记忆网络模型的边框标签 (p, q, w, h) ;

[0023] 步骤1.6:记录步骤1.3中矩形边框底边中点在图像坐标系下的坐标 (a, b) ;再以摄像头光轴延长线交地面于0点定为路面坐标系中心原点,车辆前进方向为Y轴方向,路面平面为XY平面,记录与矩形边框底边中点对应的路面位置坐标 (A, B) ,作为训练数据集的定位标签。

[0024] 进一步的优选方案,所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:步骤2中搭建并训练障碍物检测、长短期记忆网络和定位的联合模型的具体过程为:

[0025] 步骤2.1:采用卷积神经网络搭建障碍物检测模型:网络由16层卷积层,5层池化层,2层全连接层构成;模型输入是完整图像像素数据,输出结果是维度为 $S*S*(B*5+C)$ 的向量, B 为每个网格负责检测物体的个数;输出结果向量的含义为:每个网格检测出 B 个矩形边框,每个矩形边框位置信息包括了 x, y, w, h ,每个矩形边框还具有置信度 P 信息,每个网格对应预测类别向量 c ,包括 C 个类别概率值,其中最大的为网格的类别预测结果;

[0026] 根据检测模型结果得到障碍物边框信息,设定阈值 T ,只保留 $P \geq T$ 的边框,再根据非极大抑制原理来消除冗余的边框,最后获取余下边框的位置信息;

[0027] 步骤2.2:搭建长短期记忆网络模型,分为输入层,隐藏层,输出层;输入层的输入分为:由当前图像经过检测模型得到的边框位置信息与检测模型中第一层全连接层输出的 M 维向量组合得到的向量,以及上一幅图像经长短期记忆网络模型的输出结果;输出层的输出是当前图像中障碍物边框新的位置信息;

[0028] 步骤2.3:采用全连接神经网络搭建障碍物定位模型:定位模型采用三层网络,网络输入层节点为5,隐层节点数为20,输出层节点为2;其中5个输入分别是长短期记忆网络模型输出的向量中得到的障碍物边框底边中点坐标 (a, b) ,图像的宽 W 和高 H ,摄像头距地面的高度 h ;2个输出是障碍物边框在真实世界路面坐标系的位置坐标 (A, B) 。

[0029] 进一步的优选方案,所述一种基于单目视觉的实时障碍物检测和定位方法,其特征在于:训练步骤2中网络模型的具体过程为:

[0030] 步骤3.1:初始化网络中的神经元权重参数;

[0031] 步骤3.2:将采集得到的图片进行预处理,使所有训练样本的尺寸归一化为 $32S*32S*3$,对所有图像像素进行归一化,使图像的像素值在0到1之间;并对图片的曝光,色调,饱和,镜像做随机变化来扩大样本数量;将训练样本送入模型进行训练;

[0032] 步骤3.3:通过优化损失函数,结合BP算法更新网络权重参数,迭代至与标签的误差降低到预设的阈值,结束训练。

[0033] 有益效果

[0034] 本发明提出的方法与诸多现有方法相比,障碍物检测能力都在现有的检测方法处于中上等水平,本方法在检测障碍物时能够获取更多环境信息并且加入了长短期记忆网络,通过对时空序列信息的学习,获取更高的检测能力。而且在检测速率上,与现有的检测方法相比,有了很大提高,在GPU上运行能达到30帧每秒,满足实时性的要求。

[0035] 而且在对障碍物的定位时,是采用提出的基于端到端神经网络的障碍物定位模型。这种方法不需要知道相机参数,配合障碍物检测模型和长短期记忆网络模型能够快速准确地定位出障碍物在真实世界坐标系下相对于汽车的位置。因此汽车行驶中,使用本发明的方法,能够实时检测定位出障碍物位置,及时提醒驾驶员,让驾驶员有充分的时间做出合理的处理,或者是启动智能车的自动避障系统,从而提高汽车安全。该方法设计合理,能够在不同情况下使用,适用性广泛。

[0036] 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

[0037] 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:

[0038] 图1:本发明方法流程图

[0039] 图2:长短期记忆网络模型检测过程

[0040] 图3:车载摄像头图像投影

[0041] 图4:障碍物定位的边框中点标定图

[0042] 图5:定位模型网络结构

[0043] 图6:障碍物的检测和定位

[0044] 图7:长短期记忆网络模型构成

具体实施方式

[0045] 下面详细描述本发明的实施例,所述实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0046] 本发明的流程图如图1所示,首先将车载摄像机采集的视频图像尺寸归一化到 $32S*32S*3$ (S 为划分的网格数),再对所有图像像素进行归一化,使图像的像素值在0到1之间,然后输入至训练好的检测模型和长短期记忆网络中,实时检测出汽车前方障碍物,并定位障碍物在真实坐标系下的相对于汽车的位置。下面结合附图对本发明技术方案的具体实施过程作进一步说明:

[0047] 本发明的一些术语解释:

[0048] 术语1:CNN:卷积神经网络(Convolutional Neural Network)是一种多层的神经网络,类似与生物神经网络,能够模拟人类大脑学习过程,具有很强的容错,自学习和并行处理能力。卷积神经网络是一种多层前馈网络,每层由多个二维平面组成,每个平面由多个

神经元组成。网络输入为二维视觉模式,作为网络中间层的卷积层(Convolutional Layer, C)和抽样层(Subsampling, S)交替出现,类似与生物神经网络中两种重要的感知单元:简单单元和复杂单元。1) 卷积层是特征抽取层。每个卷积层中包括多个卷积神经元(C元),每个神经元只和前一层网络对应位置的局部感受野相连接,提取特征,具体体现在该神经元与前一层局部感受野的连接权重上。相比较一般的神经网络,卷积神经网络的局部连接方式大大减少网络参数,此外网络又采用权值共享的方式,进一步减少网络参数,使得网络训练更加有效。2) 抽样层是特征映射层,每个抽样层包含多个抽样神经元(S元),S元只和前一层网络对应位置的局部感受野相连接并且其连接权重固定为特定值。通过卷积层和抽样层的交替提取特征,使得网络性能更好,更鲁棒。

[0049] 术语2:BP算法:误差反向传播(Error Back Propagation, BP)算法,神经网络采用BP算法进行训练, BP算法分为信号的正向传播和误差的反向传播。信号的正向传播是将数据输入到网络中得到预测结果的过程,反向传播是将预测结果和期望值的差值作为误差,从后向前更新各层神经元权重的过程。正向传播和反向传播不断多次进行,直到误差满足要求为止。

[0050] 本发明的障碍物识别定位方法的具体处理步骤如下所述:

[0051] 步骤1:采集处理数据图像,包括如下步骤:

[0052] 1.1) 将单目摄像头安装在车上,确定摄像头距地面的高度 h 以及摄像头光轴轴线与竖直方向的夹角,即俯仰角 γ_0 。

[0053] 1.2) 在不同的天气情况(主要体现的光照条件),汽车在有障碍物的路面正常行驶,通过摄像头拍摄视频(包含背景);从视频中获取若干张包含有障碍物的图片。

[0054] 1.3) 人工用矩形边框(边框与图片边缘对应平行)对拍摄得到的包含有障碍物的图片中的障碍物的位置进行标定(矩形边框完全包围图片中的障碍物),得到边框的中心在图片坐标系(以图片中心为原点,向上为Y轴正方向,向右为X轴正方向)中的坐标 (x_1, y_1) 和边框的宽 w_1 和高 h_1 ;预先设定障碍物类别为C类,并采用one-hot编码方法,对障碍物的类别信息使用一个C维向量 \vec{c} 进行描述。比如当边框内的物体属于第i类障碍物时,类别向量 \vec{c} 的第i维为1,其余为0。

[0055] 1.4) 给每一张照片设定检测标签:将拍摄到的图片分为 $S \times S$ 网格,对于每个网格建立一个数组 (x, y, w, h, P, \vec{c}) ;对于步骤1.3中矩形边框中心所在的网格,其数组取值为: \vec{x}, \vec{y} 分别为矩形边框中心与网格中心在x方向和y方向距离的差值的归一化结果(归一化过程中的基准为网格边长), \vec{w}, \vec{h} 为矩形边框宽和高的归一化结果(归一化过程中的基准分别为图片的宽和高), $\vec{P} = 1$ (\vec{P} 是置信度, $\vec{P} \in (0, 1)$,代表网格含有物体的概率及定位准确度), \vec{c} 根据步骤1.3确定;对于其余网格,数组取值全为0;所有 $S \times S$ 网格的数组组成检测标签向量用于检测模型训练。

[0056] 1.5) 对于1.3中矩形边框的位置信息 (x_1, y_1, w_1, h_1) 均作归一化处理(归一化过程中的基准分别为图片的宽和高),得到 $(\vec{p}, \vec{q}, \vec{w}, \vec{h})$ 。作为训练长短期记忆网络模型的边框标签。

[0057] 1.6) 再增加数据集的障碍物定位信息,把前面标定各个边框的底边中点的图像平

面坐标记录下来 (a, b), 作为边框 (障碍物) 在图片坐标系 (以图片中心为原点) 下的定位坐标; 如图3所示, 然后再以摄像头光轴作延长线交地面于O点定为路面坐标系中心原点 (也是视野梯形的对角线交点), 车辆前进方向为Y轴方向, 并取路面平面为XY平面。再把与图片中的边框的底边中点相对应的路面位置坐标也记录下来 (A, B) , 作为边框 (障碍物) 在真实世界路面坐标系的位置坐标, 即为模型的训练数据集的定位标签。

[0058] 步骤2: 训练模型, 包括如下步骤:

[0059] 本发明模型分为障碍物检测模型, 长短期记忆网络和定位模型。

[0060] 设计一个可以进行端对端训练的模型, 进行联合训练。将步骤1中制作好的的图像数据集作为训练集送入模型进行训练, 最终得到模型不仅能够预测障碍物类别, 边框位置信息和置信度; 还能够预测障碍物 (边框) 在真实世界下定位坐标。

[0061] 2.1障碍物检测模型:

[0062] 搭建障碍物检测网络架构, 本发明障碍物检测模型主要采用如下的多层神经网络, 各层的网络定义如下表所列:

[0063]

	网络类型	卷积核	尺寸/步长	输出特征尺寸
--	------	-----	-------	--------

[0064]

1	Convolutional	64	3x3/ 1	32S*32S*64
2	Convolutional	64	3x3/ 1	32S*32S*64
3	Maxpool		2x2/ 2	16S*16S*64
4	Convolutional	128	3x3	16S*16S*128
5	Convolutional	128	3x3	16S*16S*128
6	Maxpool		2x2/ 2	8S*8S*128
7	Convolutional	256	3x3	8S*8S*256
8	Convolutional	256	3x3	8S*8S*256
9	Convolutional	256	3x3	8S*8S*256
10	Maxpool		2x2/ 2	4S*4S*256
11	Convolutional	512	3x3	4S*4S*512
12	Convolutional	512	3x3	4S*4S*512
13	Convolutional	512	3x3	4S*4S*512
14	Maxpool		2x2/ 2	2S*2S*256
15	Convolutional	512	3x3	2S*2S*512
16	Convolutional	512	3x3	2S*2S*512
17	Convolutional	512	3x3	2S*2S*512
18	Maxpool		2x2/ 2	S*S*256
19	Convolutional	1024	3x3	S*S*1024
20	Convolutional	1024	3x3	S*S*1024
21	Convolutional	1024	3x3	S*S*1024
22	Fully Connection			M
23	Fully Connection			S*S*(B*5+C)

[0065] 网络由16层卷积层(Convolutional Layer),5层池化层(Maxpool Layer),2层全连接层(Fully Connection Layer)构成。其中只使用3*3卷积核,提升模型计算速度。

[0066] 在每一层卷积层都增加Batch Normalization,采用Leaky激活函数。最后一层预测类别的概率,边框的位置信息和置信度,输出向量大小为S*S*(B*5+C),S为划分的网格

数,B为每个网格负责检测物体的个数,C为类别个数。该维度主要包括含义是:1.每个网格会检测出B个矩形边框,每个矩形边框位置信息包括了 (x,y,w,h) ,分别是矩形边框中心与网格中心在x方向和y方向距离的差值的归一化结果(归一化基准为网格边长)和矩形边框宽和高的归一化结果(归一化基准分别为图片的宽和高);2.每个矩形边框还有一个置信度P,代表该处是否有物体及定位准确度;3.每个网格会预测类别向量c,包括C个类别概率值,其中最大的即为网格的类别预测结果。

[0067] 由检测模型结果得到障碍物边框信息,设定阈值T,只保留 $P \geq T$ 的边框;再根据非极大抑制原理来消除冗余的边框,具体做法是1.将检测结果中所有的框按置信度排序,选中最高分及其对应的框;2.遍历其余的边框,如果和当前最高分边框的重叠面积大于一定阈值t,我们就将框剔除;3.从未处理的框中继续选一个得分最高的,重复上述过程。这样我们就能得到边框。再获取这部分边框的位置信息,将其中边框坐标 (x,y) 变换为矩形边框中心与图片坐标系中心坐标的归一化结果 (p,q) (归一化基准分别为图片的宽和高),最终得到边框位置信息 (p,q,w,h) 。

[0068] 2.2长短期记忆网络(LSTM)模型:

[0069] 长短期记忆网络是循环神经网络的一种,循环神经网络一大优势就是处理具有某种序列的数据。搭建长短期记忆网络模型,从历史视觉语义也获取帮助,网络既考虑了障碍物周边的空间信息,也考虑了过去帧的鲁棒视觉特征,充分利用长短期记忆网络模型对上述有时空序列的视觉特征向量学习得到更佳的障碍物边框的位置信息,从而提高了障碍物识别和定位能力。

[0070] 模型构成如图7所示,主要由忘记门,输入门,输出门组成。

[0071] 忘记门决定长短期记忆网络丢弃什么信息,该门的决定丢弃信息由以下函数确定:

$$[0072] \quad f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

[0073] 其中, σ 表示Sigmoid函数。 W_f 为遗忘门权值, b_f 为遗忘门偏置项, x_t 表示t时刻的输入值, h_{t-1} 表示t-1时刻输出值, f_t 表示遗忘系数。

[0074] 输入门决定记忆细胞cells中更新的信息,该门的更新信息由以下函数确定:

$$[0075] \quad i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$

$$[0076] \quad \bar{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c)$$

[0077] 其中, W_i, W_c 分别表示输入门权重和记忆细胞权重, b_i, b_c 分别表示输入门偏置项和记忆细胞偏置项, i_t 表示输入系数, \bar{C}_t 表示新的输入值向量。

[0078] 遗忘门和输入门对记忆细胞cells的状态进行更新,状态更新函数由以下函数确定:

$$[0079] \quad C_t = f_t * C_{t-1} + i_t * \bar{C}_t$$

[0080] 其中, C_{t-1} 表示t-1时刻的记忆细胞cells状态值, C_t 表示t时刻记忆细胞cells状态值。

[0081] 输出门决定了长短期记忆网络的输出值,其由以下函数决定:

$$[0082] \quad o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$[0083] \quad h_t = o_t * \tanh(C_t)$$

[0084] 其中, w_o 表示输出门权重, b_o 表示输出门偏置项, o_t 表示输出系数, C_t 表示 t 时刻记忆细胞 $cells$ 状态值, h_t 表示在 t 时刻长短期记忆网络的输出。

[0085] LSTM 的核心就是记忆细胞的存在, 本质上扮演了状态信息积累器的角色。过去记忆细胞 C_{t-1} 通过“忘记门 f_t ”来决定哪些信息可以留下, 再通过“输入门 i_t ”来决定加入哪些新的信息构成了新的记忆细胞 C_t 。最后通过“输出门 o_t ”决定输出哪些信息 h_t 。本发明利用长短期记忆网络序列化学习的特点, 来深入挖掘前后帧的关系, 有效地结合时空信息, 提高障碍物检测定位正确率。

[0086] 搭建障碍物检测的长短期记忆网络, 如图2所示, 分为输入层, 隐藏层, 输出层。其中隐藏层的节点个数设为256个, 以保证较好的运算效率。输入层其中一个输入 x_t 是当前帧的经障碍物检测模型得到的边框位置信息 (p_t, q_t, w_t, h_t) 和检测模型中第一层全连接层输出的 M 维向量组合的向量, 另一个输入 h_{t-1} 是上一帧图像经长短期记忆网络输出结果 $(p_{t-1}, q_{t-1}, w_{t-1}, h_{t-1})$; 输出层 h_t 是当前帧图像的障碍物边框新的位置信息 (p_t', q_t', w_t', h_t') 。

[0087] 2.3障碍物的定位模型:

[0088] 搭建障碍物定位模型: 使用全连接神经网络。定位模型采用三层网络, 网络输入层节点为5, 隐层节点数为20, 输出层节点为2。其中5个输入分别是长短期记忆网络模型输出的向量中得到的有效边框(障碍物)的底边中点坐标 (a, b) , 图像的宽 W 和高 H , 摄像头距地面的高度 h 。2个输出是预测边框(障碍物)在真实世界路面坐标系的位置坐标 (A, B) 。在神经元的激活函数选择 sigmoid 函数。

[0089] 2.4联合训练模型

[0090] (1) 用一些很小的随机数初始化网络中神经元权重参数。

[0091] (2) 将采集得到的图片进行预处理, 使所有训练样本的尺寸归一化为 $32S \times 32S \times 3$ (S 为划分的网格数), 再对所有图像像素进行归一化, 使图像的像素值在0到1之间。并对图片的曝光, 色调, 饱和, 镜像等做随机变化来扩大样本数量。将训练样本送入进行模型训练。

[0092] (3) 通过优化损失函数, 结合BP算法更新网络权重参数, 迭代10000~9000000次, 直到误差降低到预设的阈值就结束网络的训练。损失函数由5部分组成,

$$[0093] \quad L1 = \sum_{i=0}^{S^2} (P_i - \bar{P}_i)^2$$

$$[0094] \quad L2 = \sum_{i=0}^{S^2} \sum_{c \in \text{classes}} (p_i(c) - \bar{p}_i(c))^2$$

$$[0095] \quad L3 = \sum_{i=0}^{S^2} \sum_{j=0}^B [(x_{ij} - \bar{x}_i)^2 + (y_{ij} - \bar{y}_i)^2 + (w_{ij} - \bar{w}_i)^2 + (h_{ij} - \bar{h}_i)^2]$$

$$[0096] \quad L4 = \frac{1}{n} \sum_{k=1}^n [(p_k - \bar{p}_k)^2 + (q_k - \bar{q}_k)^2 + (w_k - \bar{w}_k)^2 + (h_k - \bar{h}_k)^2]$$

$$[0097] \quad L5 = \frac{1}{n} \sum_{k=1}^n [(A_k - \bar{A}_k)^2 + (B_k - \bar{B}_k)^2]$$

$$[0098] \quad L = \frac{\alpha * L1 + \beta * L2 + \lambda * L3 + \theta * L4 + \delta * L5}{\alpha + \beta + \lambda + \theta + \delta}$$

[0099] 其中S是图片划分网格的个数;B是每个网格预测边框的个数;L1,L2,L3分别是检测模型中边框的置信度损失,边框障碍物类别概率损失和位置信息损失,L4是长短期记忆网络模型中边框的位置信息损失,L5是定位模型中边框(障碍物)在真实世界下位置信息损失; $\alpha, \beta, \lambda, \theta, \delta$ 的分别为L1,L2,L3,L4,L5的权重系数;L1中的 P_i 是第i个网格预测的边框的置信度;L2中的 $p_i(c)$ 是第i个网格预测的边框中物体属于障碍物c的类别概率;L3中的 $x_{ij}, y_{ij}, w_{ij}, h_{ij}$ 分别是第i个网格预测的边框j的中心坐标以及边框的宽和高;L4中的 p_k, q_k, w_k, h_k 分别是第k个边框的中心坐标以及边框的宽和高,L5中的 A_k, B_k 是第k个边框在真实世界下路面坐标系的定位坐标。 $\bar{P}_i, \bar{x}_i, \bar{y}_i, \bar{w}_i, \bar{h}_i, \bar{p}_i(c)$ 则分别为第i个网格的边框标签的置信度,中心坐标,宽和高和类别概率的期望值; $\bar{p}_k, \bar{q}_k, \bar{w}_k, \bar{h}_k, \bar{A}_k, \bar{B}_k$ 分别为第k个边框的中心坐标,宽和高和定位坐标的期望值。最终的损失函数L是通过5个部分损失函数加权的方式获得的。

[0100] 为了计算各个隐藏层中的误差,我们可以通过BP算法的反向传播来间接计算得到。反向传播回来的误差可以看做是每个神经元的基的灵敏度(就是误差对基的变化率),定义如下:

$$[0101] \quad \frac{\partial L}{\partial b} = \frac{\partial L}{\partial u} * \frac{\partial u}{\partial b} = \delta$$

[0102] 其中 $\frac{\partial L}{\partial b}$ 为误差对当前层的单个神经元u的偏导数,又 $\frac{\partial u}{\partial b} = 1$, 故

$$[0103] \quad \delta = \frac{\partial L}{\partial u}$$

[0104] 所以能直接根据下式求出各层级的灵敏度:

$$[0105] \quad \delta^l = (W^{l+1})^T \delta^{l+1} \circ f'(u^l)$$

[0106] 其中, W^{l+1} 为第l+1层的权值系数,得到各层级的灵敏度后,根据下式对权值进行更新。

$$[0107] \quad \frac{\partial E}{\partial W^l} = x^{l-1} (\delta^l)^T$$

$$[0108] \quad \Delta W^l = -\eta * \frac{\partial E}{\partial W^l}$$

[0109] 其中,对于第l层,误差对于该层权值的导数 $\frac{\partial E}{\partial W^l}$ 是该层的输入 x^{l-1} (等于上一层的输出)与该层的灵敏度 δ^l (该层每个神经元的组合成一个向量的形式)的叉乘。然后得到的偏导数乘以学习率 η ,就是该层的神经元的权值的更新。

[0110] (4)得到模型。

[0111] 步骤3:障碍物的检测和定位。

[0112] 将摄像头固定在汽车之前的位置,将摄像头拍摄的视频直接输入到已经训练好的模型中,就能直接端对端预测出每一帧图像中障碍物的类别,障碍物边框,置信度,以及障碍物(边框)在真实世界坐标系下位置信息。又已知路面坐标系中心原点O点到摄像头的距

离为 d ,所以可以得到矩形边框相对与汽车前端的位置 $(A, B-d)$,即障碍物相对于汽车的位置信息。模型在GPU下运行,检测速度可达到30帧每秒,满足实时要求。

[0113] 本发明有以下优点:

[0114] 1.方法简单,容易实施。本发明只采用摄像头,将机器视觉和深度学习结合进行端对端的障碍物的检测,联合训练出一个可以对障碍物的实时检测和定位(测距)的模型,能达到高效准确的碰撞预警。计算速度快,且不需要高精度的雷达设备,方法简单实用。

[0115] 2.检测和定位能力强,本发明在设计过程中,将检测和定位问题转换为回归问题,充分利用时空序列信息丰富的特点,提高了检测和定位的准确率。

[0116] 3.本发明能够准确有效的检测出不同环境下障碍物并准确定位,满足实时性要求。

[0117] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在不脱离本发明的原理和宗旨的情况下在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

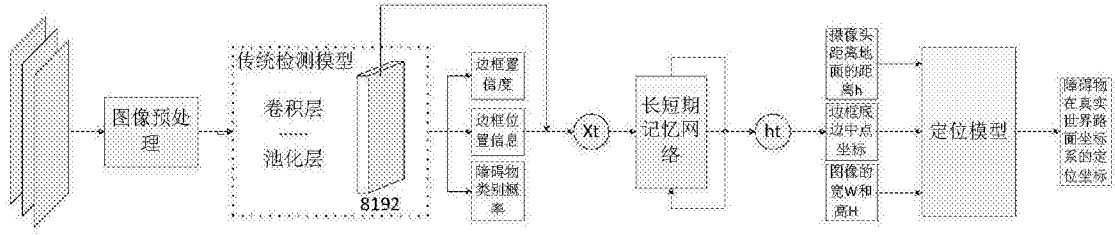


图1

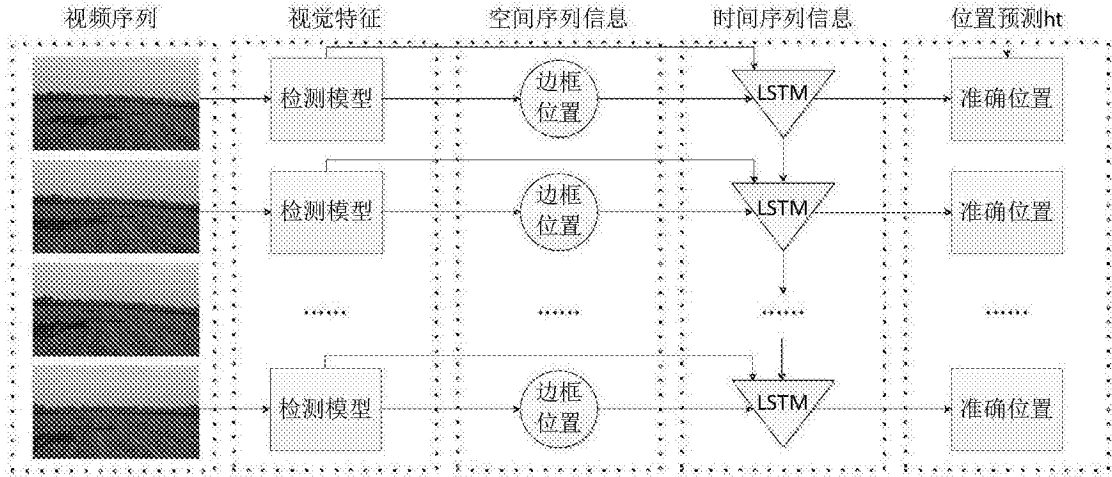


图2

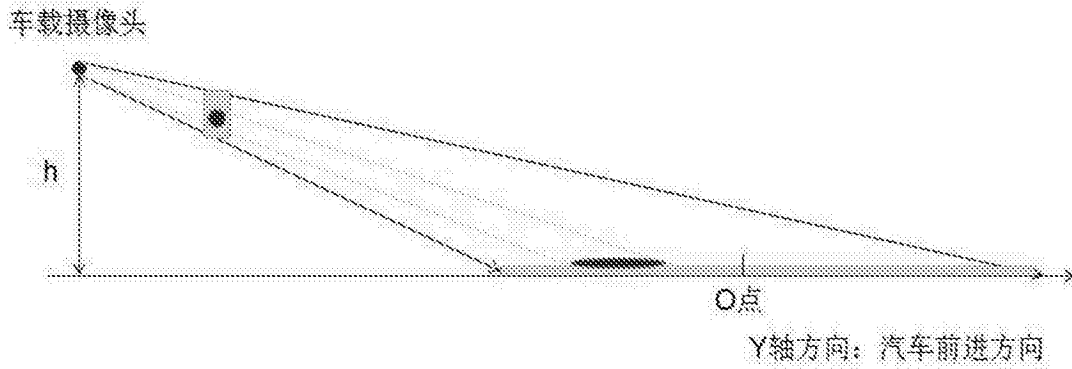


图3

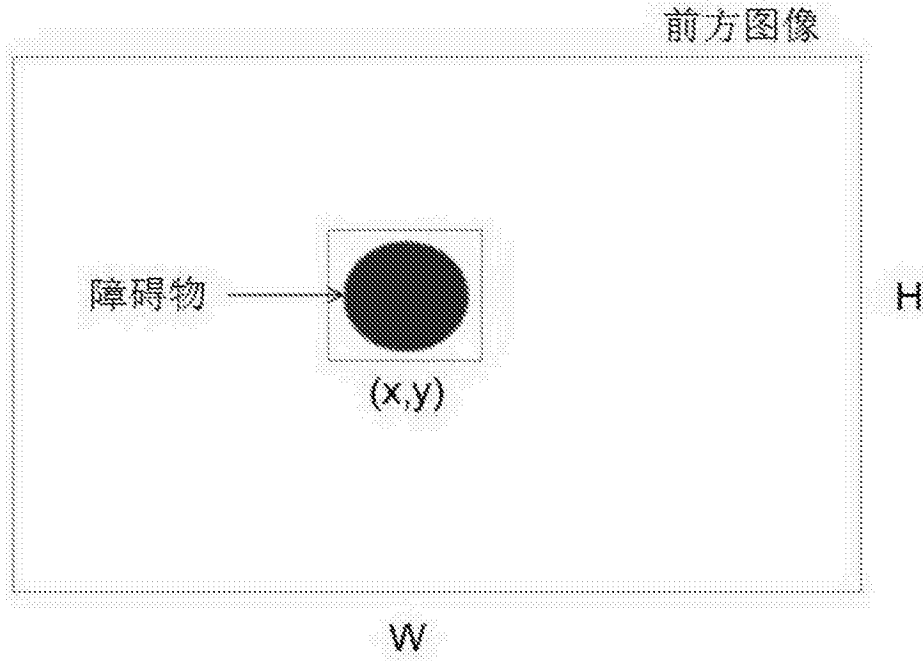


图4

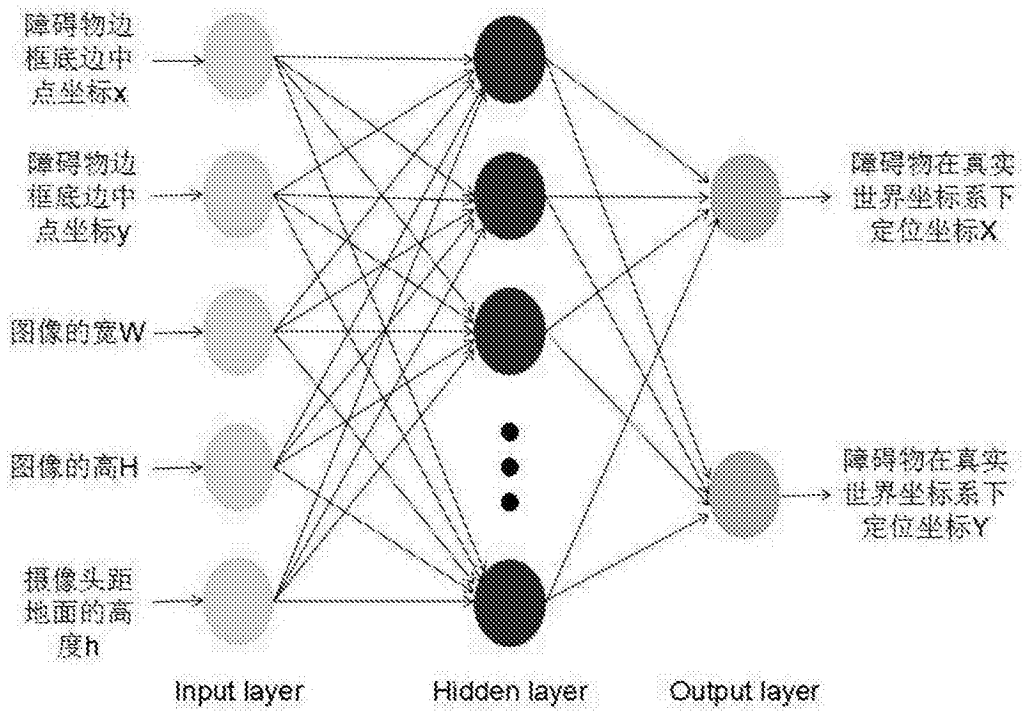


图5



图6

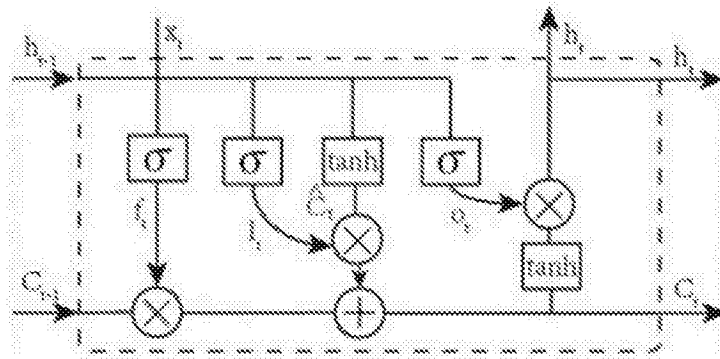


图7