



(12) 发明专利

(10) 授权公告号 CN 107741231 B

(45) 授权公告日 2020.11.27

(21) 申请号 201710940904.6

G06K 9/00 (2006.01)

(22) 申请日 2017.10.11

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 107741231 A

CN 105574552 A, 2016.05.11

CN 104795011 A, 2015.07.22

CN 101079210 A, 2007.11.28

(43) 申请公布日 2018.02.27

CN 105469429 A, 2016.04.06

(73) 专利权人 福州大学

CN 102609686 A, 2012.07.25

CN 103838933 A, 2014.06.04

地址 350116 福建省福州市闽侯县上街镇
大学城学园路2号福州大学新区

WO 0113323 A1, 2001.02.22

(72) 发明人 吴丽君 吴晓烽 陈志聪 文吉成
林培杰 程树英

US 2010157280 A1, 2010.06.24

US 9031317 B2, 2015.05.12

(74) 专利代理机构 福州元创专利商标代理有限公司 35100

审查员 陈丹华

代理人 蔡学俊

(51) Int. Cl.

G01C 21/20 (2006.01)

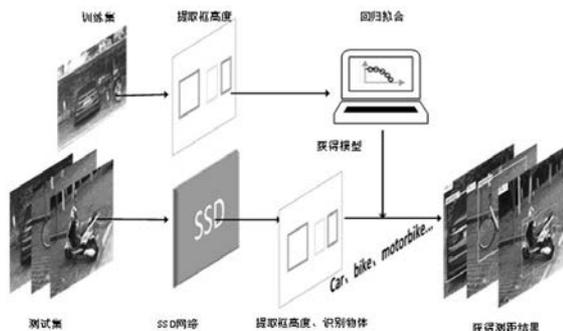
权利要求书2页 说明书6页 附图4页

(54) 发明名称

一种基于机器视觉的多运动目标快速测距方法

(57) 摘要

本发明涉及一种基于机器视觉的多运动目标快速测距方法。该方法首先对图像训练集中标注框 (Ground Truth Box) 与实际距离进行拟合, 并保存其测距模型, 接着利用实时检测算法SSD (Single Shot MultiBox Detector) 对特定目标所产生的检测框 (Bounding Box) 进行高度提取并输入到测距模型中, 从而实时地获取多个目标的测量距离。本发明所提出的多运动目标快速识别及测距方法, 能够高效、实时地对特定多个目标进行识别与测距。相较其他ETA测距系统更适合盲人, 更加实用, 其拥有较好的应用前景以及实际工程的应用价值。



1. 一种基于机器视觉的多运动目标快速测距方法,其特征在於:包括如下步骤,

S1、训练及测试图像数据集采集:对盲人出行影响较大的物体类别进行图像数据采集,将采集的数据分为三份,分别为训练集、测试集和侧面测试集,其中侧面测试集用来评估摄像头不同角度拍摄对测距的影响;

S2、目标识别:采用卷积神经网络进行目标识别,利用基础分类网络微调以提升准确率,利用不同层级的特征图进行检测以适应不同尺寸的目标物体,同时采用默认检测框机制来离散化输出框;

S3、训练图像数据集人工标注:对训练集以及测试集中每个图片中的每一类目标物进行人工标注,并记录其宽与高;

S4、数据拟合:已标记好的训练集中标注框的高度记为 X ,距离记为 Y ,显著值 $a=0.05$;根据其趋势采取指数模型: $y = \beta_0 e^{\beta_1 x}$,接着考虑到高度与距离量纲区别较大,为了消除其影响,首先将数据归一化到0-1范围;其次,为了提高模型的泛化能力,加入高斯噪声;然后,利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$;最后,经过若干次迭代得到测距模型;

所述步骤S4中,利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$ 的具体过程为:

利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$,对 β 随机生成初始

$$\text{值,并计算 } \mathbf{X}^* = \begin{bmatrix} e^{\beta_1^0 x_1} & \beta_0^0 e^{\beta_1^0 x_1} \\ \vdots & \vdots \\ e^{\beta_1^0 x_n} & \beta_0^0 e^{\beta_1^0 x_n} \end{bmatrix}_{n \times 2} \text{ 与 } \mathbf{Y}^* = \begin{bmatrix} y_1 - \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ y_2 - \beta_0^0 x_2 e^{\beta_1^0 x_2} \\ \vdots \\ y_n - \beta_0^0 x_n e^{\beta_1^0 x_n} \end{bmatrix};$$

S5、测距实现:将步骤S4得到的测距模型与卷积神经网络结合,所采用的卷积神经网络是SSD,即利用SSD对目标检测产生检测框并提取检测框高度输入测距模型,从而得出距离。

2. 根据权利要求1所述的一种基于机器视觉的多运动目标快速测距方法,其特征在於:所述步骤S4中,将数据归一化到0-1范围采用的是Min-Max方法,高斯噪声均值为0,方差为0.05。

3. 根据权利要求1所述的一种基于机器视觉的多运动目标快速测距方法,其特征在於:所述对盲人出行影响较大的物体类别为三类,分别为汽车、自行车、摩托车。

4. 根据权利要求3所述的一种基于机器视觉的多运动目标快速测距方法,其特征在於:所述步骤S1中训练集的采集为对三类物体从4米开始每间隔0.5米进行拍摄,同时记录拍摄处距离。

5. 根据权利要求3所述的一种基于机器视觉的多运动目标快速测距方法,其特征在於:所述步骤S1中测试集的采集为随机站在不同距离对三类物体进行拍摄,并记录拍摄处距离。

6. 根据权利要求1所述的一种基于机器视觉的多运动目标快速测距方法,其特征在于:所述步骤S2中,基础分类网络为VGG-16,在基础分类网络后额外添加了逐层递减的卷积层进行预测以及实现多尺度预测和共享权重;同时在网络训练阶段,将已标注好的图片中的标注框与特征图中的默认检测框进行匹配,符合设为正样本,不符合为负样本;并且对于每个默认检测框预测其定位误差与置信度,并将两者加权和为设为目标函数。

7. 根据权利要求1所述的一种基于机器视觉的多运动目标快速测距方法,其特征在于:所述步骤S5中,利用实时检测算法SSD对目标检测产生检测框之前,需对采集的图片进行尺度变化,变换为与实时检测算法SSD相匹配的大小。

一种基于机器视觉的多运动目标快速测距方法

技术领域

[0001] 本发明涉及一种基于机器视觉的多运动目标快速测距方法。

背景技术

[0002] 现如今,我国存在七千五百万视障人群,其生活以及出行都因其视力的障碍而受到极大困扰。而帮助盲人的电子行走辅助(ETA,Electronic TravelAids)系统的出现,为盲人带来了极大便利,其中障碍物检测与测距是ETA系统中的关键技术,检测识别准确性和实时性是衡量ETA系统实用性的重要标志,也是亟待解决的难题。

[0003] 在以往非计算机视觉领域,大多都采用激光、红外、超声波测距的方法。但其测距距离短,波束窄,无法识别障碍物,且只能通过频率区分多个目标,实用度略低。在计算机视觉领域,有ESSVI、NAVI、SoundView等视觉辅助系统,其中具有代表性是上海交大的SoundView,其从摄像头取得图像,然后通过耳机把视觉检测信息通过声音反馈给盲人。但是SoundView是采用给每个物体贴条形码进行检测识别的,并适用于现实应用场景。此外,现有的其它基于机器视觉的ETA系统,在测距时仍然要依赖于额外的测距仪来实现,并且无法同时对多个运动目标测距。针对现有ETA系统的不足,本发明从目标检测算法与回归测距结合这一新的角度,来解决ETA中核心的检测测距模块中的实时性与识别问题。

[0004] 随着深度学习的出现,神经网络因其具有较好的特征提取功能,而被广泛用在视觉领域,使得目标检测领域不断推陈出新,主要代表的算法有RCNN、faster-RCNN、SSD等,其中SSD拥有较高的检测精确度、召回率以及实时性。本发明提出一种基于SSD及指数函数拟合的多目标识别及测距算法,其主要思想是将SSD检测出的待测目标检测框高度传入基于回归拟合训练所得的测距结构,从而实时地获取多个目标的距离。

发明内容

[0005] 本发明的目的在于改进现有盲人行走辅助系统(ETA)的不足,提供一种基于机器视觉的多运动目标快速测距方法,该方法能够自动高效、实时地对特定多个目标进行识别与测距,对提升盲人的电子行走辅助有重要的作用。

[0006] 为实现上述目的,本发明的技术方案是:一种基于机器视觉的多运动目标快速测距方法,包括如下步骤,

[0007] S1、训练及测试图像数据集采集:对盲人出行影响较大的物体类别进行图像数据采集,将采集的数据分为三份,分别为训练集、测试集和侧面测试集,其中侧面测试集用来评估摄像头不同角度拍摄对测距的影响;

[0008] S2、目标识别:采用卷积神经网络进行目标识别,利用基础分类网络微调以提升准确率,利用不同层级的特征图进行检测以适应不同尺寸的目标物体,同时采用默认检测框机制来离散化输出框;

[0009] S3、训练图像数据集人工标注:对训练集以及测试集中每个图片中的每一类目标物进行人工标注,并记录其宽与高;

[0010] S4、数据拟合：已标记好的训练集中标注框的高度，记为X，距离记为Y，显著值 $a=0.05$ 。根据其趋势采取指数模型： $y = \beta_0 e^{\beta_1 x}$ ，接着考虑到高度与距离量纲区别较大，为了消除其影响，首先将数据归一化到0-1范围；其次，为了提高模型的泛化能力，加入高斯噪声；然后，利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$ ；最后，经过若干次迭代得到测距模型；

[0011] S5、测距实现：步骤S4得到的测距模型与卷积神经网络结合，具体地，本发明所采用的卷积神经网络是SSD，即利用SSD对目标检测产生的检测框并提取高度输入测距模型，从而得出距离。

[0012] 在本发明一实施例中，所述步骤S4中，将数据归一化到0-1范围采用的是Min-Max方法，高斯噪声均值为0，方差为0.05。

[0013] 在本发明一实施例中，所述步骤S4中，利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$ 的具体过程为：

[0014] 利用高斯-牛顿迭代法求解模型 $y = \beta_0 e^{\beta_1 x}$ 的未知数 $\beta = (\beta_0, \beta_1)^T$ ，对 β 随机生成初

始值，并计算 $X = \begin{bmatrix} e^{\beta_1^0 x_1} & \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ \vdots & \vdots \\ e^{\beta_1^0 x_n} & \beta_0^0 x_n e^{\beta_1^0 x_n} \end{bmatrix}_{n \times 2}$ 与 $Y^* = \begin{bmatrix} y_1 - \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ y_2 - \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ \vdots \\ y_n - \beta_0^0 x_1 e^{\beta_1^0 x_1} \end{bmatrix}$ 。

[0015] 在本发明一实施例中，所述对盲人出行影响较大的物体类别为三类，分别为汽车、自行车、摩托车。

[0016] 在本发明一实施例中，所述步骤S1中训练集的采集为对三类物体从4米开始每隔0.5米进行拍摄，同时记录拍摄处距离。

[0017] 在本发明一实施例中，所述步骤S1中测试集的采集为随机站在不同距离对三类物体进行拍摄，并记录拍摄处距离。

[0018] 在本发明一实施例中，所述步骤S2中，基础分类网络为VGG-16，在基础分类网络后额外添加了逐层递减的卷积层进行预测以及实现多尺度预测和共享权重；同时在网络训练阶段，将已标注好的图片中的标注框与特征图中的默认检测框进行匹配，符合设为正样本，不符合为负样本；并且对于每个默认检测框预测其定位误差与置信度，并将两者加权和为设为目标函数。

[0019] 在本发明一实施例中，所述步骤S5中，利用实时检测算法SSD对目标检测产生检测框之前，需对采集的图片进行尺度变化，变换为与实时检测算法SSD相匹配的大小。

[0020] 相较于现有技术，本发明具有以下有益效果：与传统辅盲测距算法相比，本发明提出的测距算法有如下几个优点：

[0021] (1) 测距速度快。SSD在VOC2007测试集中达到实时检测的效果(46FPS, 74.3MAP, GTX TITAN X)，与回归测距模块结合后，在本发明工作环境下亦能达到20FPS(GTX 1070显卡的环境下)。

[0022] (2) 能够多目标检测识别并测距。传统的视觉辅助系统只能检测出障碍物,而并不能识别出障碍物类别。很显然,“前方有车”比“前方有人”应让人警惕,因此能够识别障碍物十分关键。同时传统视觉检测只能对于某一像素点进行测距,而无法对于某一确切运动目标,甚至多个运动目标进行实时测距,很显然后者更适用于现实需求的ETA。

[0023] (3) 测距较远,精度适当。激光、红外、超声波等测距距离十分倚靠于其功率,因此测距距离十分有限,而采用视觉测距能够‘极目远视’,检测距离达到30米以上,并且可以定量地反馈给盲人距离信息,更适用于盲人。

附图说明

[0024] 图1为本发明测距方法整体框图。

[0025] 图2为本发明识别方法网络结构图

[0026] 图3为本发明训练数据示意图。

[0027] 图4为本发明测试数据示意图。

[0028] 图5为本发明侧面角度测试数据示意图。

[0029] 图6为本发明正面采集图像的标注示意图。

[0030] 图7为本发明侧面采集图像的标注示意图。

[0031] 图8为本发明回归方程训练数据拟合回带图

[0032] 图9为本发明回归方程测试数据拟合回带图

[0033] 图10为本发明在视频流中实时测距结果示意图。

具体实施方式

[0034] 下面结合附图,对本发明的技术方案进行具体说明。

[0035] 本发明的一种基于机器视觉的多运动目标快速测距方法,包括如下步骤,

[0036] S1、训练及测试图像数据集采集:对盲人出行影响较大的物体类别进行图像数据采集,将采集的数据分为三份,分别为训练集、测试集和侧面测试集,其中侧面测试集用来评估摄像头不同角度拍摄对测距的影响;

[0037] S2、目标识别:采用卷积神经网络进行目标识别,利用基础分类网络微调以提升准确率,利用不同层级的特征图进行检测以适应不同尺寸的目标物体,同时采用默认检测框机制来离散化输出框;

[0038] S3、训练图像数据集人工标注:对训练集以及测试集中每个图片中的每一类目标物进行人工标注,并记录其宽与高;

[0039] S4、数据拟合:已标记好的训练集中标注框的高度,记为X,距离记为Y,显著值 $\alpha = 0.05$ 。根据其趋势采取指数模型: $y = \beta_0 e^{\beta_1 x}$,接着考虑到高度与距离量纲区别较大,为了消除其影响,首先采用Min-Max方法将数据归一化到0-1范围;其次,为了提高模型的泛化能力,加入高斯噪声,高斯噪声均值为0,方差为0.05;然后,利用高斯-牛顿迭代法求解模型

$$y = \beta_0 e^{\beta_1 x} \text{ 的未知数 } \beta = (\beta_0, \beta_1)^T, \text{ 对 } \beta \text{ 随机生成初始值, 并计算 } X = \begin{bmatrix} e^{\beta_1^0 x_1} & \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ \vdots & \vdots \\ e^{\beta_1^0 x_n} & \beta_0^0 x_n e^{\beta_1^0 x_n} \end{bmatrix}_{n \times 2}$$

$$\text{与 } Y^* = \begin{bmatrix} y_1 - \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ y_2 - \beta_0^0 x_1 e^{\beta_1^0 x_1} \\ \vdots \\ y_n - \beta_0^0 x_1 e^{\beta_1^0 x_1} \end{bmatrix}; \text{ 最后, 经过若干次迭代得到测距模型;}$$

[0040] S5、测距实现：步骤S4得到的测距模型与卷积神经网络结合，具体地，本发明所采用的卷积神经网络是SSD，即利用SSD对目标检测产生的检测框并提取高度输入测距模型，从而得出距离。

[0041] 所述对盲人出行影响较大的物体类别为三类，分别为汽车、自行车、摩托车。

[0042] 所述步骤S1中训练集的采集为对三类物体从4米开始每间隔0.5米进行拍摄，同时记录拍摄处距离。所述步骤S1中测试集的采集为随机站在不同距离对三类物体进行拍摄，并记录拍摄处距离。

[0043] 所述步骤S2中，基础分类网络为VGG-16，在基础分类网络后额外添加了逐层递减的卷积层进行预测以及实现多尺度预测和共享权重；同时在网络训练阶段，将已标注好的图片中的标注框与特征图中的默认检测框进行匹配，符合设为正样本，不符合为负样本；并且对于每个默认检测框预测其定位误差与置信度，并将两者加权和为设为目标函数。

[0044] 所述步骤S4中，利用实时检测算法SSD对目标检测产生检测框之前，需对采集的图片进行尺度变化，变换为与实时检测算法SSD相匹配的大小。

[0045] 以下为本发明的具体实现实例。

[0046] 一种用于盲人视觉辅助的多运动目标快速识别及测距方法，采集训练及测试图像数据集并对每一类别进行人工标注，其次提取已标记好的训练集以及测试集中标注框的高度进行回归拟合获得测距模型，并且通过测试集验证模型有效性，同时通过测试集来验证本发明方法的鲁棒性，最后利用检测算法SSD获得目标的检测框，并将检测框高度输入测距模型，从而实现实时地获得多个运动目标的距离。具体原理框图参见图1。

[0047] 进一步的，选择基础分类网络为VGG-16，并在基础分类网络后添加了逐层递减的额外卷积层进行预测以及实现多尺度预测和共享权重。同时在网络训练阶段，将已标注好的图片中的标注框与特征图中的默认检测框进行匹配，符合设为正样本，不符合为负样本。并且对于每个默认检测框预测其定位误差与置信度，并将两者加权和为设为目标函数，具体识别网络结构图参见图2。

[0048] 进一步的，如图3、4所示，本发明对三类（汽车、自行车、摩托车）对盲人出行影响较大的类别进行图像数据采集，采集的数据分为三份，分别为训练集，测试集，侧面测试集。其中侧面测试集用来评估摄像头不同角度拍摄对测距的影响。训练集采集：对三类物体从4米开始每间隔0.5米进行拍摄，并记录现在所在距离，拍摄38组（4~22.5米），其距离分别为

6.5米、12米、22米,测试集的采集:随机站在不同距离对三类物体进行拍摄,并记录此时距离。本发明拍摄了22组测试集,其距离分别为16.2米、9.3米、22.9米。每类侧面角度的测试集采集:随机站在三类物体侧面不同距离位置进行拍摄,并且记录距离。侧面数据示例如图5所示,距离皆为5米。

[0049] 进一步的,对测距所采用的特征进行研究选择,对正面及侧面的图片分别进行人工标注,并记录其宽与高(单位为像素),其中图6为已标注好的正面图片一张,图7为各类已标注好的侧面图片一张(如图7(a)、7(b)、7(c)所示),且图6和图7的拍摄距离一样。因在相同距离时,不同拍摄视角会使得目标的标注框宽度影响较为严重,以汽车作为例子,如图6与图7(a)所示,汽车标注框的宽度受拍摄视角变化所产生的影响较为巨大(1331变化为3074),而其高度则大多保持不变(1038变化为1262),且误差在可接受范围之类,其他类别亦是如此。因此,本发明采取较为不受视角影响的物体检测框的高度作为测距模型输入参数从而进行测距,以提高测距算法的鲁棒性。

[0050] 进一步的,对训练集中每个图片中的每一类进行人工标注,对已标记好的训练集以及测试集中标注框的高度,记为X(pixel),距离记为Y(m),显著值。以下以汽车为例,旨在利用训练集拟合出回归方程,并用测试集验证。根据其趋势采取指数模型: $y = \beta_0 e^{\beta_1 x}$, 考虑到高度与距离量纲区别较大,为了消除其影响,首先将数据归一化到0-1范围。之后,为了提高模型的泛化能力,对X,Y分别加上均值为0,方差为0.05的高斯噪声,由模型进一步看到 β_1 的大小反应了车辆高度对距离的影响程度。拟合结果如图8、图9所示,拟合结论如下表所示。

[0051] 表1模型在训练集回归结论(针对线性化后的数据)

回归方程	指数类型 $y = 0.9695e^{-3.666x}$
统计量	$n=38$,自变量数 $m=1$, $SE=2.6089$, $SR=31.6045$, $ST=34.2134$, 统计量 $F=437.600769^{**}$, $F_{0.05}=4.121$
方程检验结论	因为统计量 $F > F_{0.05}$,所以认为方程线性相关性显著

[0053] 表2模型在测试集回归结论(针对线性化后的数据)

回归方程	指数类型 $y = 0.9695e^{-3.666x}$
测试集统计量	$n=20$ 自变量数 $m=1$, $SE=2.6713$, $SR=17.2539$, $ST=19.9252$, 统计量 $F^1=116.261820^{**}$, $F_{0.05}=4.41$
侧边测试集统计量	$n=20$ 自变量数 $m=1$, $SE=3.5318$, $SR=22.2950$, $ST=25.8269$ 统计量 $F^2=113.627612^{**}$, $F_{0.05}=4.41$
方程检验结论	因为统计量 $F^1 > F_{0.05}$, $F^2 > F_{0.05}$ 。所以认为方程线性相关性显著

[0055] 由表1、表2可知,训练好的模型能够较好的拟合测试集,并且对于侧边拍摄角度的测试集也能够较好的反应高度和距离的关系。

[0056] 进一步的,在CPU为i7-6700k、RAM为16GB、显卡为GTX1070,显存为8GB以及运行环境为Ubuntu16.04,装有CUDA8.0及CUDNN GPU加速库,PyCharm2016.3.3的集成开发环境,搭载Keras、TensorFlow框架的环境中,将拟合后的模型运用于训练集中,因数据采集摄像头拍摄的图片像素为3264*2448,而SSD检测算法检测图片固定输入为300*300,因此在步骤S5中对检测图片进行尺度变化,再输入模型得出距离。图10为最终结果图片示例。

[0057] 以上是本发明的较佳实施例,凡依本发明技术方案所作的改变,所产生的功能作用未超出本发明技术方案的范围时,均属于本发明的保护范围。

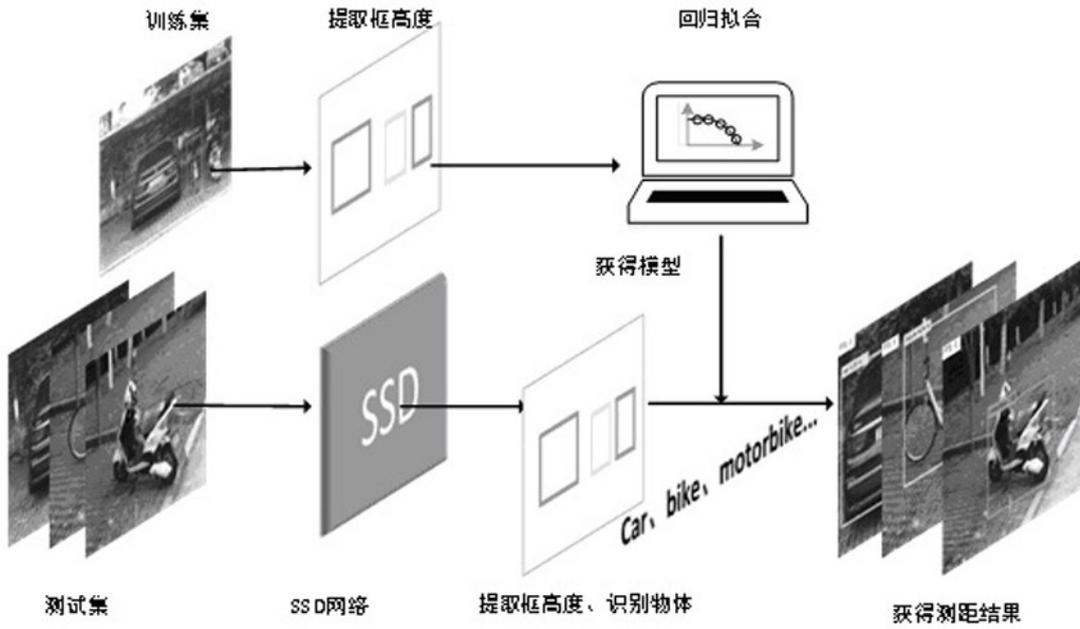


图1

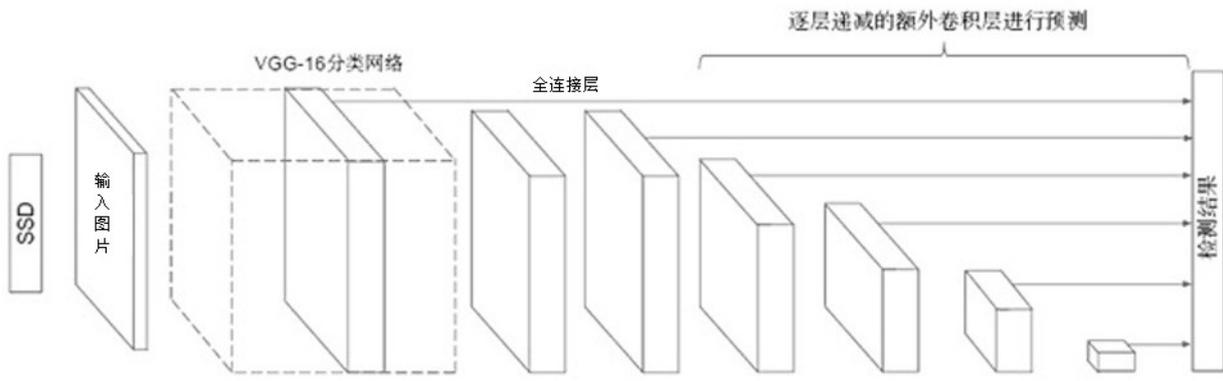


图2

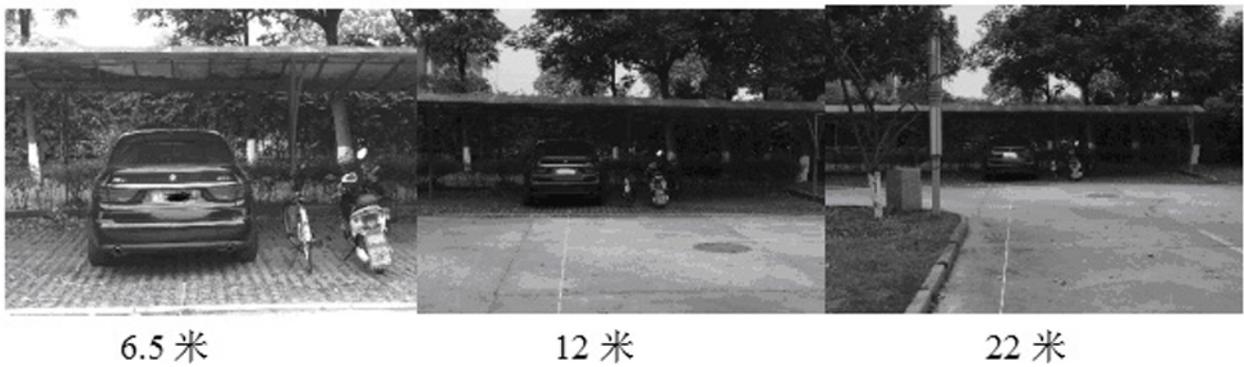


图3



图4



图5



图6



图7

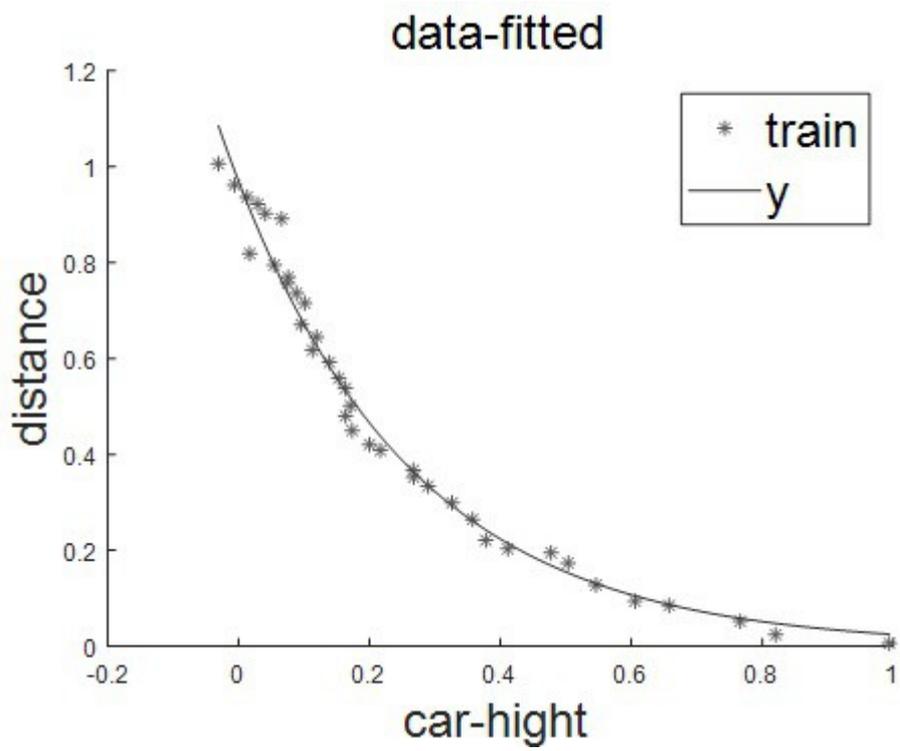


图8

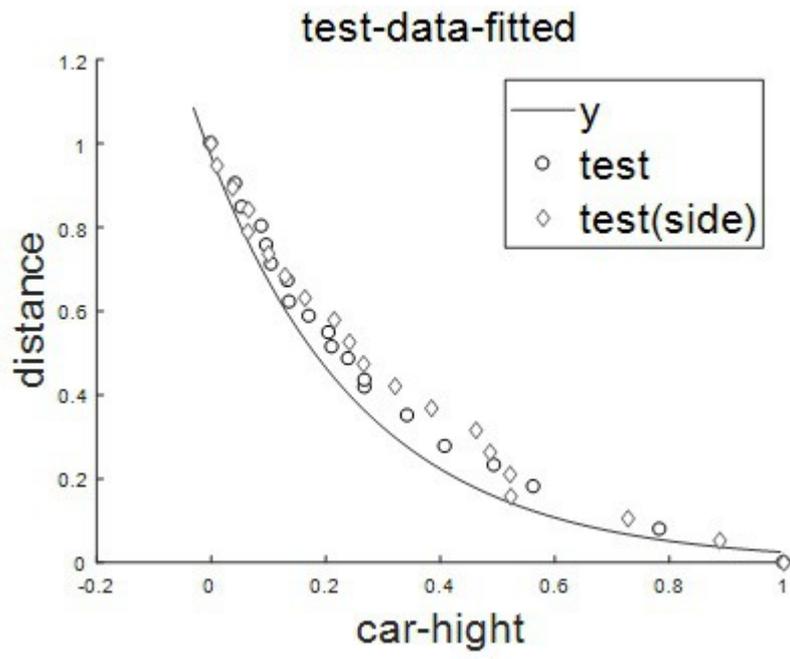


图9



第16帧

第28帧

第31帧

图10