



(12)发明专利

(10)授权公告号 CN 111311666 B

(45)授权公告日 2020.08.14

(21)申请号 202010401126.5

(22)申请日 2020.05.13

(65)同一申请的已公布的文献号

申请公布号 CN 111311666 A

(43)申请公布日 2020.06.19

(73)专利权人 南京晓庄学院

地址 211171 江苏省南京市江宁区弘景大道3601号南京晓庄学院

(72)发明人 王燕清 陈长伟 赵向军 石朝侠

肖文洁 李泳泉

(51)Int.Cl.

G06T 7/55(2017.01)

G06K 9/62(2006.01)

审查员 吕鑫

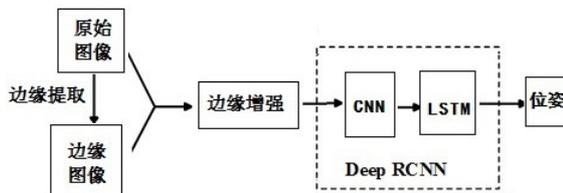
权利要求书2页 说明书8页 附图4页

(54)发明名称

一种融合边缘特征和深度学习的单目视觉里程计方法

(57)摘要

本发明公开了一种融合边缘特征和深度学习的单目视觉里程计方法,涉及视觉里程计技术领域,由于在低纹理的运动场景中,图像中可提取的特征较少,视觉里程计在进行特征匹配时会出现特征数据缺失的问题,并造成位姿估计的精度下降。本发明创新点是一种融合边缘特征和深度学习的单目视觉里程计方法,首先基于Canny边缘检测算法进行边缘增强算法的设计,边缘增强后的图像数据集作为卷积神经网络的输入并特征提取,卷积神经网络的输出再输入到循环神经网络中进行计算,最后输出整个模型对相机位姿估计,对特征提取优化。实验结果表明算法能够在模型训练时学习到更多的图像特征,提高位姿估计的准确度,并且在低纹理的场景中表现出优越的性能。



1. 一种融合边缘特征和深度学习的单目视觉里程计方法,其特征在于:基于Canny边缘检测算法进行边缘增强算法的设计,边缘增强后的图像数据集作为卷积神经网络输入并特征提取,卷积神经网络输出再输入到循环神经网络中进行计算,对特征提取优化,最后输出整个模型对相机位姿估计,具体包括以下步骤:

S1、边缘增强:提取来自边缘edge的信息结构,输出则是边缘图和原始RGB 图像的融合,将图像进行边缘增强;Canny 算子是一种基于梯度最大搜索的图像边缘检测算法,利用这个算法的部分步骤,添加操作来实现边缘锐化,整个过程如下:开始对RGB图像进行平滑处理以减少噪声,计算每个方向的一阶导数图像的渐变,使边缘模糊化;针对边缘检测图像的R,G,B的三个通道的图像矩阵采用高斯滤波器的模糊滤波方法平滑图像; H_{ij} 为卷积核大小 $(2k+1) \times (2k+1)$ 的高斯滤波器,计算公式如下所示,其中 i, j 表示核内任意点:

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right), \quad (1)$$

$$1 \leq i, j \leq (2k+1)$$

S2、内核选定:内核大小 $2k+1$ 是根据预期的模糊效果来选择的,如果内核更小,模糊不可见;在实验中,每个像素亮度值 e 计算如下:

$$e = \text{sum}(H * A) \quad (2)$$

其中 $*$ 表示卷积运算, A 是以像素的亮度值 e 为中心的窗口, sum 函数表示矩阵中所有元素的和;

S3、计算梯度:检测图像边缘的强度和方向,边缘对应于像素强度的变化,为了检测边缘使用公式2计算像素在水平 x 和垂直 y 两个方向上对应像素亮度值 e ,在每个通道 I_R, I_G, I_B 中将 A 与Sobel内核 S_x 和 S_y 卷积来实现计算图像的梯度;针对通道R图像,梯度 G_R 和方向 θ_R 的计算公式如下:

$$G_R = \sqrt{G_{R_x}^2 + G_{R_y}^2} = \sqrt{(\text{sum}(S_x * A))^2 + (\text{sum}(S_y * A))^2} \quad (3)$$

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (4)$$

其中 G_{R_x}, G_{R_y} 分别表示R图像中像素点在水平方向 x 和垂直方向 y 的梯度值; S_x 和 S_y 分别表示水平与垂直方向的Sobel算子,用于检测垂直和水平方向的边缘;得到一个RGB边缘图,并在下一步中执行非最大抑制以使边缘变薄;在非最大抑制步骤中,需要确定当前点是否在方向上是插值梯度量的局部最大值,将像素与旁边的像素进行比较,如果像素较大,则不对其进行更改;否则,将像素设置为零,从而得到图像的RGB边缘图;

S4、边缘增强:按比例将原始RGB图像的像素值与RGB边缘图相加,可以获得原始图像的边缘增强图像,将边缘增强图像作为深度神经网络的输入, R通道边缘增强方法如下:

$$I_R = (w_{raw}P_{raw} + w_{edge}P_{edge}) * \frac{255}{P_{max}} \quad (5)$$

其中 w_{raw} , w_{edge} 分别代表原始图像与边缘图像的权重, P_{raw} 和 P_{edge} 则分别表示原始图像各通道与边缘图像各通道同一位置的像素点的值; P_{max} 表示的是在前面相加计算之后图像中各通道所有像素点中的最大值, $w_{edge}=0.8$;

S5、RCNN的体系结构:将边缘增强图像作为深度神经网络的输入,深度神经网络采用了结合了CNN和RNN的基于RCNN的体系结构,具有允许VO提取特征和顺序模型的优势,网络将单目边缘增强图像序列作为输入,通过卷积层、池化层和全连接层进行训练和测试,获得网络模型权重值;

S6、针对单目边缘增强图像序列,在每个时间步减去训练集的平均RGB值,然后调整为64的倍数作为预处理RGB图像帧;将两个连续的图像堆叠在一起,作为深度RCNN的张量,以学习运动信息提取和姿势估计的方式,通过将图像张量输入到CNN中做为单目VO有效特征,然后传递到RNN中以进行顺序学习,每个时间步的图像对都会通过网络生成一个姿态估计,捕获图像后,将估算新的姿势;

S7、在网络中由所有位置p和序列方向的均方误差MSE组成的损失 L^* :

$$L^* = \arg \min_L \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \|\hat{p}_t - p_t\|_2^2 + \lambda \|\hat{\phi}_t - \phi_t\|_2^2 \quad (6)$$

其中 (p_t, ϕ_t) 表示的是在时刻t相机的真实位姿, $(\hat{p}_t, \hat{\phi}_t)$ 表示的是该时刻网络所估计的相机位姿, $\|\cdot\|$ 表示的是矩阵的L2范数计算, N 代表的是样本数据集中图像帧的数量,在相机位姿变化中,坐标平移的距离变化的尺度要远大于其角度变化的尺度,引入参数 λ 用来平衡位姿坐标在平移与方向之间的尺度差距。

一种融合边缘特征和深度学习的单目视觉里程计方法

技术领域

[0001] 本发明涉及视觉里程计技术领域,具体为一种融合边缘特征和深度学习的单目视觉里程计方法。

背景技术

[0002] 视觉里程计是一种根据输入图像估计自我运动的方法,它是即时定位与地图构建系统的核心模块。由于单目视觉里程计(Visual Odometry,VO)可以根据摄像机的馈电信号确定当前位置,它已经成为计算机视觉领域的一个热点研究领域。在自动驾驶、机器人技术等有着广泛的应用。近年来,立体摄像机的视觉里程计取得了很大的进展,由于其可靠的深度图估计能力,得到了广泛的应用。然而,一旦摄像机与基线之间的距离不同于场景的尺度,就会退化为单目的距离。

[0003] 与立体视觉里程计不同的是,单目VO不能获得真实尺度的环境地图和机器人运动,因此需要借助先验知识或摄像机高度等信息来估计绝对尺度,这使得单目VO比立体VO更容易产生较大的漂移,更具有挑战性。因此,尺度漂移是单目里程测量中亟待解决的问题,需要通过恢复绝对尺度来消除。

[0004] 深度学习(DL)以其强大的特征学习能力在计算机视觉领域得到了广泛的应用,并取得了长足的进步。动态移动的物体和缺乏特征的开放区域周围的公路,缺乏特征信息,本发明提出了一种重视边缘特征,提供结构信息的方法,该算法是在Canny边缘检测算法的基础上扩展而来的。通过这种方法,本发明可以得到一个单目视觉里程计,它在不同的场景下具有较强的鲁棒性,能够进行更准确的特征提取。

发明内容

[0005] (一)解决的技术问题

[0006] 针对环境是低纹理和忽视图像特征的重要性,导致图像特征不足,本发明提出了一种融合边缘特征提取和深度递归卷积神经网络(RCNN)的单目VO方法。该方法将传统的几何算法嵌入到基于深度学习的VO中,以增强图像中边缘特征信息的影响,该在低纹理环境下能体现很好的性能。

[0007] (二)技术方案

[0008] 根据所采用的技术和框架,算法主要分为两类:基于几何的算法和基于学习的算法。

[0009] A.基于深度学习的单目视觉里程计(Visual Odometry,VO)方法

[0010] 图像序列与特征之间的同步检测为基于深度学习的立体视觉提供了可行的方案。在本质上,它将VO视为一个分类问题,而不是姿态回归问题。使用CNN对特定场景的图像进行微调,以解决单张图像的摄像机重定位问题。使用StructurefromMotion(SfM)标记这些图像,在大规模场景中会耗费更多的时间和人力。为了克服这个问题,Costante等人首先利用CNNs对图像特征匹配得到的密集光流进行自我运动估计。Flowodomtry使用

FlowNet结合 CNN提取特征进行回归,是单目视觉里程计学习的一种端到端方法。Ls-vo使用一个自动编码器网络来表示非线性形式流形中的光流来估计自我运动。与基于模型的单目视觉系统相比,本研究采用非线性细胞神经网络进行特征提取和运动估计,积累直接回归为全局轨迹的相对姿态。阿格拉瓦尔等人提出了一种从自我运动估计学习视觉特征的算法,可以估计相机的相对位置。Ummerhofer 等人提出了一个端到端的可视里程计和网络通过将 SfM 视为一个监督式学习问题来估计深度。Wang 等提出了一种递归卷积网络结构 DeepVO,用于从视频序列中学习单目里程计。Deepvo 将 FlowNet 特征输入长短时记忆 LSTM作为单目视觉里程学习的序列到序列编解码器。为了提高性能,将额外的 IMU 读数集成到同一架构中。此外,GCN使用 RCNN体系结构来生成相应的关键点描述符。这些研究利用神经网络学习图像判读,利用神经网络的序列预测能力进行状态估计。

[0011] 一些研究试图通过将全局和相对姿态回归结合在一个共享神经网络,实现全局重定位和单目视觉里程计。基于语义的方法 VLocNet和它的高级版本VLocNet++可以根据连续的单目图像实现六自由度全局姿态回归和 VO 评估。MapNet通过学习一种数据驱动的地图表示,对相机进行全局定位,同时通过两幅连续图像的几何约束,消除估计相对位姿产生的漂移。方法只适用于已知环境,并且处理局部几何约束,相对位姿估计的效果有限,对于未知的场景,这种单目VO方法是不适用的。

[0012] 利用神经网络和基于DL的单目VO可以获得更准确的绝对尺度估计。在深度神经网络的训练中,需要比传统 VO 更多的地面真实数据。与其他学习任务相比,获取训练数据容易,本发明可以通过 GPS-RTK 或 Vicon 摄像机直接获取数据,不需要手动注释。

[0013] B.基于边缘的单目 VO

[0014] 边缘携带图像的重要信息。作为人类视觉的一部分,用于边缘识别的人工检测器可以实现复杂图像的结构信息抽取,为了提高相机估计的精度,全局一致性检验在长轨迹中发挥重要作用。本发明在单目 VO 中设计一种将边缘特征提取和深层RCNN网络相结合的方法,以便在低纹理环境中,如高速公路周围的大面积开阔地区,获得良好的性能。

[0015] c.本发明的方法

[0016] 本发明的方法结合了基于边缘检测算法 Canny 和深层RCNN 框架的边缘锐化操作。该方法的体系结构如图 1 所示。

[0017] 1.边缘增强算法

[0018] 基于Canny边缘检测算法进行边缘增强算法设计,边缘增强后的图像数据集作为卷积神经网络输入并特征提取,卷积神经网络输出再输入到循环神经网络中进行计算,对特征提取优化,最后输出整个模型对相机位姿估计,具体包括以下步骤:

[0019] S1、边缘增强:提取来自边缘edge的信息结构,主要输出是边缘图和原始RGB 图像的融合,相当于边缘锐化,不同边缘检测器的选择对最终的性能有不同的影响,鲁棒性对于跟踪连续帧中的边缘非常重要,提出的算法得益于边缘提取是一个高度健壮,高度可并行化和泛化的过程;Canny 算子是一种基于梯度最大搜索的图像边缘检测算法,利用这个算法的部分步骤,添加一些其他操作来实现边缘锐化,整个过程如下:开始对RGB图像进行平滑处理以减少噪声,计算每个方向的一阶导数图像的渐变,使边缘模糊化;利用中心差分计算梯度R,G,B的三个通道用于边缘检测的图像,高斯的模糊滤波是用于平滑图像,卷积核 $H_{i,j}$ 大小为 $(2k+1) \times (2k+1)$ 的高斯滤波器的计算公式如下所示,其中 i, j 表示核内任意点:

$$[0020] \quad H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right), \quad (1)$$

$$1 \leq i, j \leq (2k+1)$$

[0021] S2、内核大小 $2k+1$ 是根据预期的模糊效果来选择的,如果内核更小,模糊就不那么可见了,在实验中,H表示使用了一个 $5*5$ 内核大小的过滤器,每个像素 e 的亮度值可以这样计算:

$$[0022] \quad e = \text{sum}(H * A) \quad (2)$$

[0023] 其中 $*$ 表示卷积运算,A 是以像素 e 为中心的窗口,sum函数表示矩阵中所有元素的和;

[0024] S3:计算梯度:利用边缘检测算子计算图像梯度,检测图像边缘强度和方向,边缘对应于像素强度的变化,使用滤波器突出像素在水平 x 和垂直 y 两个方向上对应的亮度变化,图像平滑后计算导数 I_x 和 I_y ,通过在每个通道 I_R , I_G , I_B 中将 I_x 和 I_y 与Sobel内核 S_x 和 S_y 卷积来实现这一步骤;梯度 G_R 和方向 θ_R 的计算公式如下所示公式以R通道的计算为例:

$$[0025] \quad G_R = \sqrt{G_{R_x}^2 + G_{R_y}^2} = \sqrt{(\text{sum}(S_x * A))^2 + (\text{sum}(S_y * A))^2} \quad (3)$$

$$\theta_R = \arctan(G_{R_y} / G_{R_x}) \quad (4)$$

$$[0026] \quad S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

[0027] 其中 G_{R_x} , G_{R_y} 分别表示R中像素点在水平方向 x 和垂直方向 y 的梯度值; S_x 和 S_y 分别表示水平与垂直方向的Sobel算子,用于检测垂直和水平方向的边缘,得到一个RGB边缘图,在下一步中执行非最大抑制以使边缘变薄;在非最大抑制步骤中,需要确定该点是否在其方向上是插值梯度量的局部最大值,此步骤对边缘的性能有重大影响,将像素与其旁边的像素进行比较,如果像素较大,则不对其进行更改;否则,将像素设置为零,从而得到图像的RGB边缘图;

[0028] S4:按比例将原始RGB图像的像素值与RGB边缘图相加,可以获得原始图像的边缘增强结果,该结果是后续深度神经网络的输入,每个通道的计算方法如下,以R通道为例:

$$[0029] \quad I_R = (w_{raw}P_{raw} + w_{edge}P_{edge}) * \frac{255}{P_{max}} \quad (5)$$

[0030] 其中 w_{raw} , w_{edge} 分别代表原始图像与边缘图像的权重, P_{raw} 和 P_{edge} 则分别表示原始图像与边缘图像同一位置的像素点的值; P_{max} 表示的是在前面相加计算之后

图像中所有像素点中的最大值, $W_{edge}=0.8$ 。

[0031] 2.RCNN体系结构

[0032] 单目视觉里程计系统随时间发展并且对在运动过程中获得的图像序列进行操作, 导出运动模型连续图像帧之间的连接, 采用了深层的RCNN体系结构, 结合了CNN和RNN的RCNN的体系结构, 具有允许单目视觉里程计提取特征和顺序模型的优势, 网络将单目边缘增强图像序列作为输入。在每个时间步减去训练集的平均RGB值, 将两个连续的图像堆叠在一起, 作为深度RCNN的张量, 以学习运动信息提取和姿势估计的方式, 通过将图像张量输入到CNN中, 为单目视觉里程计生成有效特征, 然后将其传递到RNN中以进行顺序学习, 每个时间步的图像对都会通过网络生成一个姿态估计, 捕获图像估算新的姿势。本发明定义在网络中由位置 p 和序列方向的均方误差MSE组成的损失:

$$[0033] \quad L^* = \arg \min_L \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \|\hat{p}_t - p_t\|_2^2 + \lambda \|\hat{\varphi}_t - \varphi_t\|_2^2 \quad (6)$$

[0034] 其中 (p_t, φ_t) 表示的是在时刻 t 相机的真实位姿, $(\hat{p}_t, \hat{\varphi}_t)$ 表示的是该时刻网络所估计的相机位姿, $\|\cdot\|$ 表示的是矩阵的L2范数计算, N 代表的是样本数据集中图像帧的数量, 在相机位姿变化中, 坐标平移的距离变化的尺度要远大于其角度变化的尺度, 引入参数 λ 用来平衡位姿坐标在平移与方向之间的尺度差距。

[0035] 有益效果

[0036] 本发明公开了一种融合边缘特征和深度学习的单目视觉里程计方法, 涉及视觉里程计技术领域, 由于在低纹理的运动场景中, 图像中可提取的特征较少, 视觉里程计在进行特征匹配时会出现特征数据缺失的问题, 并造成位姿估计的精度下降。本发明创新点是一种融合边缘特征和深度学习的单目视觉里程计方法, 首先基于Canny边缘检测算法进行边缘增强算法的设计, 边缘增强后的图像数据集作为卷积神经网络的输入并特征提取, 卷积神经网络的输出再输入到循环神经网络中进行计算, 对特征提取优化, 最后输出整个模型对相机位姿估计。实验结果表明算法能够在模型训练时学习到更多的图像特征, 提高位姿估计的准确度, 并且在低纹理的场景中表现出优越的性能。

[0037] 为了与其它方法进行对比, 将算法的实验结果与两个基线baseline方法, 单目视觉里程计方法VISO_M以及双目视觉里程计方法VISO_S的实验结果进行对比, 通过量化位姿估计的位移误差与旋转误差曲线(如图2-5所示)更加直观地表示其差别。

[0038] 根据图中所示的实验对比结果, 可以得出结论: 本发明提出的结合边缘增强算法的视觉里程计方案在实际运行时, 尽管其性能效果略逊色于VISO_S, 但是远优于VISO_M。当车辆处于高速运行状态(运行时速高于60km/h)时, 由于采集到的图像变化较快, 在进行特征匹配时会有较大的误差, 并且车辆所处的环境较为开阔, 纹理较少, VISO_S的性能会有所下降, 本发明提出的算法此时的性能更加稳定, 其平移误差与旋转误差依然保持稳定的下降, 不会出现反向上升的情况(图5中圆圈圈出部分)。在测试集上, 不同算法的具体指标如图10所示, t_{rel} 表示子序列长度为100m到800m的平均位移RMSE漂移量, r_{rel} 表示每一百米的平均旋转RMSE漂移量($^{\circ}/100m$)。表中与本发明算法相比较的算法不仅有两个baseline方

法,还有提出的基于残差神经网络的视觉里程计算法,即ResNet。实验数据说明相较于基于残差神经网络的视觉里程计算法,本发明提出的基于边缘增强的视觉里程计算法能够生成更加精确的位姿估计结果。

[0039] 除了对实验结果进行误差数据上的比较之外,还将测试数据集的位姿估计结果进行了可视化,即对相机的运动进行轨迹重建,并与未使用边缘增强算法的DeepV0进行性能比较。运动轨迹的重建就是将模型输出的绝对位姿映射到世界坐标系中,进而还原相机在场景中的运动轨迹。图6-9分别展示了对KITTI数据集中序列04、05、07和09的运动轨迹重建结果。实验结果也表明了,所提出的该发明的实验结果与DeepV0相比能够产生更加精确的结果。也就是说通过增强输入图像中的边缘信息,网络能够提取更多更加有效的特征信息,并根据这些先验信息估计出更加准确的尺度信息。

附图说明

- [0040] 图1为结合边缘增强算法的视觉里程计结构。
- [0041] 图2为不同运动长度子序列下的位移误差比较。
- [0042] 图3为不同运动长度子序列下的旋转误差比较。
- [0043] 图4为不同运动速度下的位移误差比较。
- [0044] 图5为不同运动速度下的旋转误差比较。
- [0045] 图6对序列04的运动轨迹重建。
- [0046] 图7为对序列05的运动轨迹重建。
- [0047] 图8为对序列07的运动轨迹重建。
- [0048] 图9为在对序列09的运动轨迹重建。
- [0049] 图10在模型在测试集上运行结果的误差比较。

具体实施方式

[0050] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0051] 请参阅图1-10,本发明实施例提供一种技术方案:一种融合边缘特征和深度学习的单目视觉里程计方法:

[0052] 1.一种融合边缘特征和深度学习的单目视觉里程计,其特征在于:基于Canny边缘检测算法进行边缘增强算法设计,边缘增强后的图像数据集作为卷积神经网络输入并特征提取,卷积神经网络输出再输入到循环神经网络中进行计算,对特征提取优化,最后输出整个模型对相机位姿估计,具体包括以下步骤:

[0053] S1、边缘增强:提取来自边缘edge的信息结构,主要输出是边缘图和原始RGB 图像的融合,相当于边缘锐化,不同边缘检测器的选择对最终的性能有不同的影响,鲁棒性对于跟踪连续帧中的边缘非常重要,提出的算法得益于边缘提取是一个高度健壮,高度可并行化和泛化的过程;Canny 算子是一种基于梯度最大搜索的图像边缘检测算法,利用这个算法的部分步骤,添加一些其他操作来实现边缘锐化,整个过程如下:开始对RGB图像进行平

滑处理以减少噪声,计算每个方向的一阶导数图像的渐变,使边缘模糊化;利用中心差分计算梯度R,G,B的三个通道用于边缘检测的图像,高斯的模糊滤波是用于平滑图像,卷积核 H_{ij} 大小为 $(2k+1) \times (2k+1)$ 的高斯滤波器的计算公式如下所示,其中 i, j 表示核内任意点:

$$[0054] \quad H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right), \quad (1)$$

$$1 \leq i, j \leq (2k+1)$$

[0055] S2、内核大小 $2k+1$ 是根据预期的模糊效果来选择的,如果内核更小,模糊就不那么可见了,在实验中,H表示使用了一个 $5*5$ 内核大小的过滤器,每个像素 e 的亮度值可以这样计算:

$$[0056] \quad e = \text{sum}(H * A) \quad (2)$$

[0057] 其中 $*$ 表示卷积运算,A 是以像素 e 为中心的窗口,sum函数表示矩阵中所有元素的和;

[0058] S3:计算梯度:利用边缘检测算子计算图像梯度,检测图像边缘强度和方向,边缘对应于像素强度的变化,使用滤波器突出像素在水平 x 和垂直 y 两个方向上对应的亮度变化,图像平滑后计算导数 I_x 和 I_y ,通过在每个通道 I_R , I_G , I_B 中将 I_x 和 I_y 与Sobel内核 S_x 和 S_y 卷积来实现这一步骤;梯度 G_R 和方向 θ_R 的计算公式如下所示公式以R通道的计算为例:

$$[0059] \quad G_R = \sqrt{G_{R_x}^2 + G_{R_y}^2} = \sqrt{(\text{sum}(S_x * A))^2 + (\text{sum}(S_y * A))^2} \quad (3)$$

$$\theta_R = \arctan(G_{R_y} / G_{R_x}) \quad (4)$$

$$[0060] \quad S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

[0061] 其中 G_{R_x} , G_{R_y} 分别表示R中像素点在水平方向 x 和垂直方向 y 的梯度值; S_x 和 S_y 分别表示水平与垂直方向的Sobel算子,用于检测垂直和水平方向的边缘,得到一个RGB边缘图,在下一步中执行非最大抑制以使边缘变薄;在非最大抑制步骤中,需要确定该点是否在其方向上是插值梯度的局部最大值,此步骤对边缘的性能有重大影响,将像素与其旁边的像素进行比较,如果像素较大,则不对其进行更改;否则,将像素设置为零,从而得到图像的RGB边缘图;

[0062] S4:按比例将原始RGB图像的像素值与RGB边缘图相加,可以获得原始图像的边缘增强结果,该结果是后续深度神经网络的输入,每个通道的计算方法如下,以R通道为例:

$$[0063] \quad I_R = (w_{raw}P_{raw} + w_{edge}P_{edge}) * \frac{255}{P_{max}} \quad (5)$$

[0064] 其中 W_{raw} , W_{edge} 分别代表原始图像与边缘图像的权重, P_{raw} 和 P_{edge} 则分别表示原始图像与边缘图像同一位置的像素点的值; P_{max} 表示的是在前面相加计算之后图像中所有像素点中的最大值, $W_{edge}=0.8$ 。

[0065] 单目视觉里程计系统随时间发展并且对在运动过程中获得的图像序列进行操作, 导出运动模型连续图像帧之间的连接, 采用了深层的RCNN体系结构, 结合了CNN和RNN的RCNN的体系结构, 具有允许单目视觉里程计提取特征和顺序模型的优势, 网络将单目边缘增强图像序列作为输入。

[0066] 在每个时间步减去训练集的平均RGB值, 将两个连续的图像堆叠在一起, 作为深度RCNN的张量, 以学习运动信息提取和姿势估计的方式, 通过将图像张量输入到CNN中, 为单目视觉里程计生成有效特征, 然后将其传递到RNN中以进行顺序学习, 每个时间步的图像对都会通过网络生成一个姿态估计, 捕获图像估算新的姿势。

[0067] 本发明定义在网络中由位置 p 和序列方向的均方误差MSE组成的损失:

$$[0068] \quad L^* = \arg \min_L \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \|\hat{p}_t - p_t\|_2^2 + \lambda \|\hat{\phi}_t - \phi_t\|_2^2 \quad (6)$$

[0069] 其中 (p_t, ϕ_t) 表示的是在时刻 t 相机的真实位姿, $(\hat{p}_t, \hat{\phi}_t)$ 表示的是该时刻网络所估计的相机位姿, $\|\cdot\|$ 表示的是矩阵的L2范数计算, N 代表的是样本数据集中图像帧的数量, 在相机位姿变化中, 坐标平移的距离变化的尺度要远大于其角度变化的尺度, 引入参数 λ 用来平衡位姿坐标在平移与方向之间的尺度差距。

[0070] 应用实验

[0071] KITTI数据集进行算法模型训练与测试, 并对测试结果进行了分析与比较。

[0072] 1 数据集介绍

[0073] 本发明实验使用的是德国Karlsruhe Institute of Technology提供的KITTI基准数据集。该数据集中有众多的视觉算法的数据与评测项目, 包括视觉里程计、目标检测、目标跟踪以及图像语义分割等。其中视觉里程计的数据集包含左右两个照相机采集的图像数据, 既可用于双目, 也可选择其中单个相机的图像序列进行单目视觉里程计研究。

[0074] 在该数据集中共有22个图像帧序列(00-21), 其中前11个序列(00-10, 序列03不可用)提供了位姿的真值数据(ground truth)。在真值数据中, 从每个序列的第一帧开始逐帧提供位姿矩阵数据, 以此给出连续帧对应的轨迹数据。在进行模型训练时, 采用前11个序列的图像数据和真实位姿数据, 为了适应卷积神经网络对输入的图像尺寸要求, 在实验时会将所有图片的尺寸调整为统一大小。

[0075] 实验使用了KITTI V0数据集中提供真值的前11个图像序列(即序列00-10,) , 其中序列00, 01, 02, 06, 08和10场景的运行轨迹较长, 数据集的规模较大, 能够提供较为丰富的图像数据, 作为算法模型训练的数据集。剩余的04, 05, 07和09场景序列则作为测试数据集对训练好的模型进行性能评估。

[0076] 2性能指标

[0077] 在对视觉里程计算法模型进行性能评估时,通常是比较算法模型估计位姿向量的精度,也就是计算位移的误差和角度的误差。其中相机的位移可以用三维向量表示,分别代表相机在前后、左右和上下移动的信息,相机的旋转则可以用旋转矩阵、欧拉角和四元数表示。在实验中相机的旋转信息采用的是欧拉角的表示方法。所谓欧拉角方法,就是将相机的旋转过程分解成三个分离的转角,三个转角分别表示相机绕着三个坐标轴的旋转。三个欧拉角分别是绕X轴旋转的俯仰角,绕Y轴旋转的航向角以及绕Z轴旋转的滚动角。通过计算三个旋转矩阵的乘积得到相机在三维空间的旋转矩阵,再将旋转矩阵转化为三个欧拉角,组成一个三维向量表示相机的角度变化。

[0078] 本发明的实验中对位移误差和角度误差进行比较时,使用了KITTI VO/SLAM的评价指标,即取100米到800米一共8种不同运动长度子序列和不同速度范围内位移误差与旋转误差的均方根误差(Root Mean Square Error, RMSE)。所谓均方根误差,就是估计值与真值之间偏差的平方和与估计次数n比值的平方根,可以用来衡量估计值与真值之间的偏差,其计算公式如下所示:

$$[0079] \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

[0080] 在计算不同运动长度子序列的位移误差和角度误差时,先采集轨迹中相同运动长度序列的样本,并计算其偏差值,再对所有采集的样本序列计算平均值得到该长度子序列的平均误差,最后将各长度子序列的平均误差进行汇总。误差值越小,表示估计值与真值之间的偏差越小,估计越准确,其最终输出的轨迹曲线与真实曲线越吻合。

[0081] 需要说明的是,在本发明中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。

[0082] 尽管已经示出和描述了本发明的实施例,对于本领域的普通技术人员而言,可以理解在不脱离本发明的原理和精神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,本发明的范围由所附权利要求及其等同物限定。

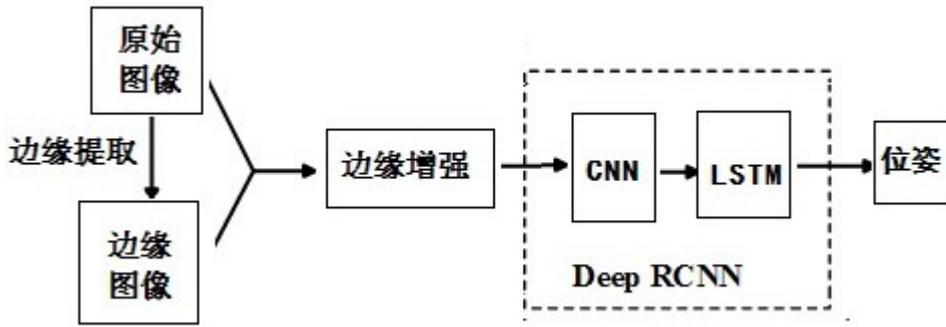


图1

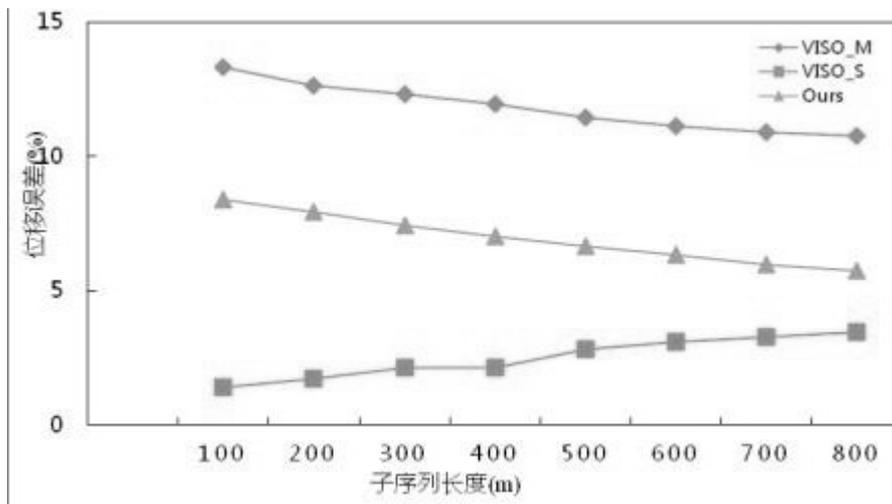


图2

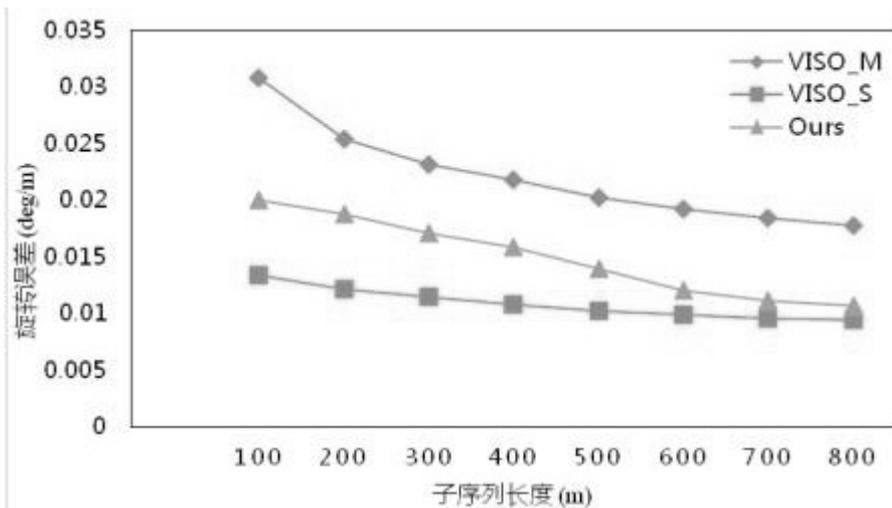


图3

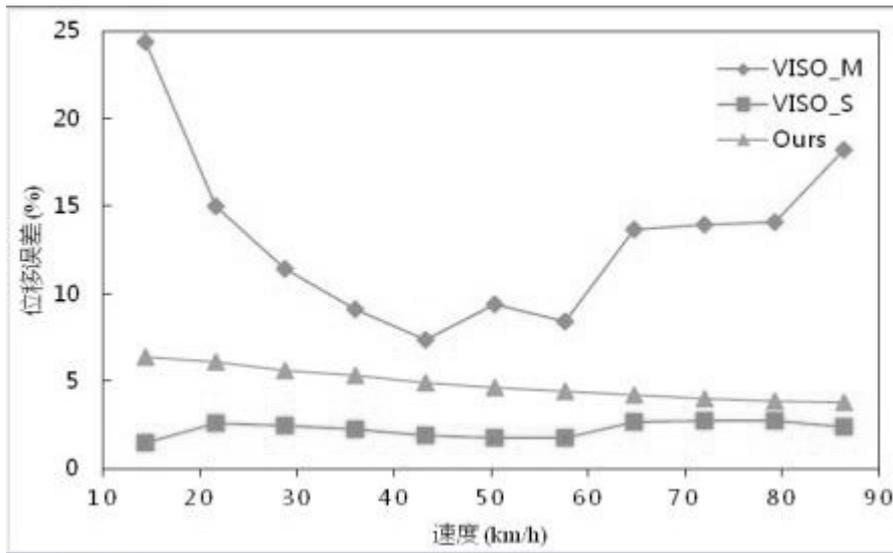


图4

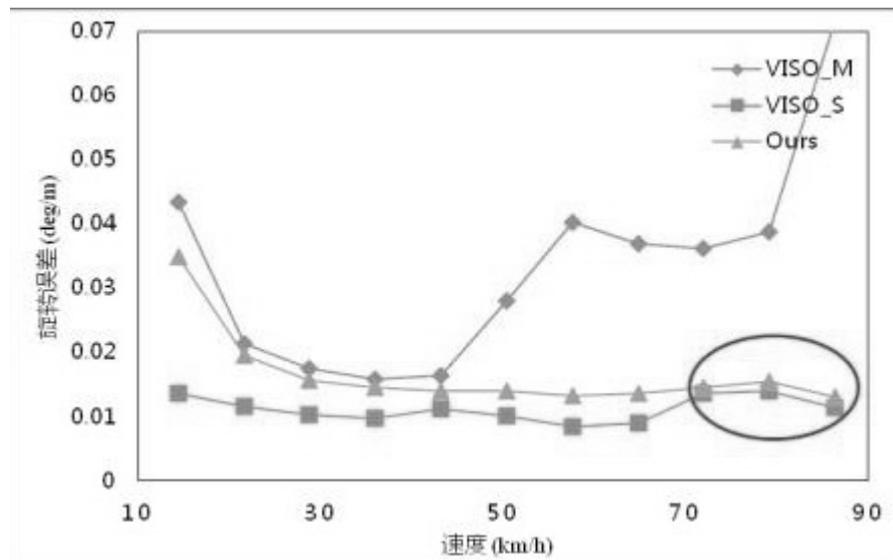


图5

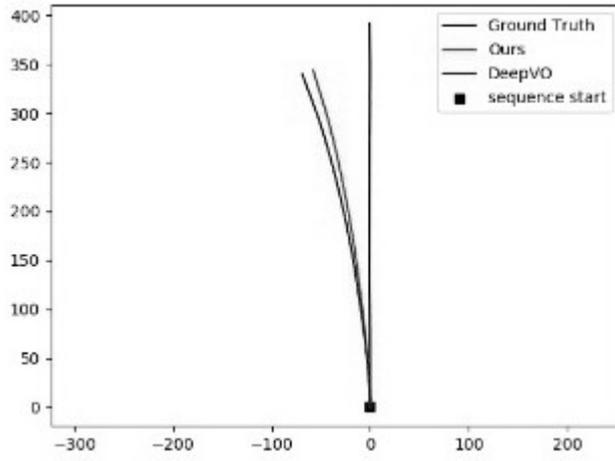


图6

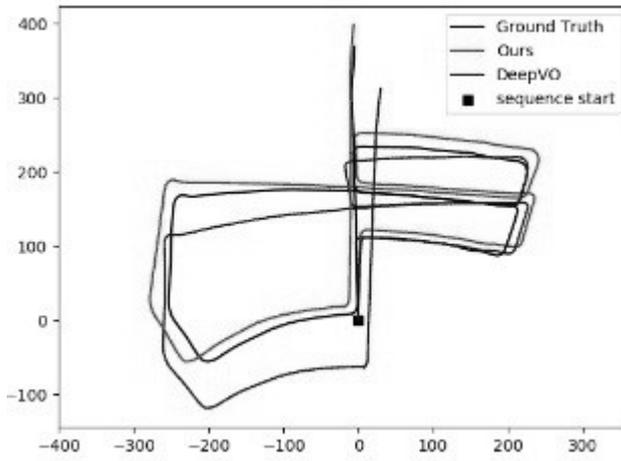


图7

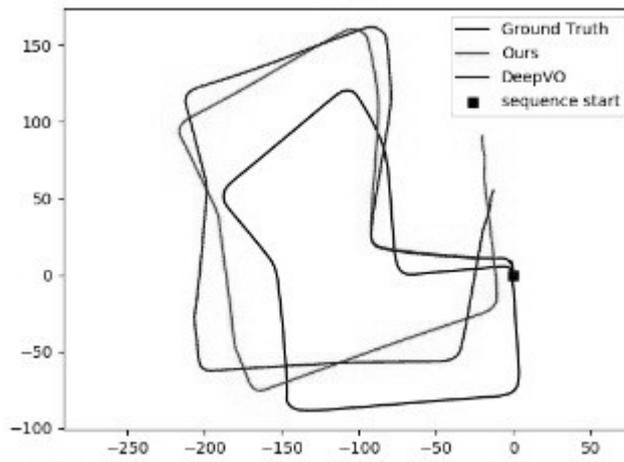


图8

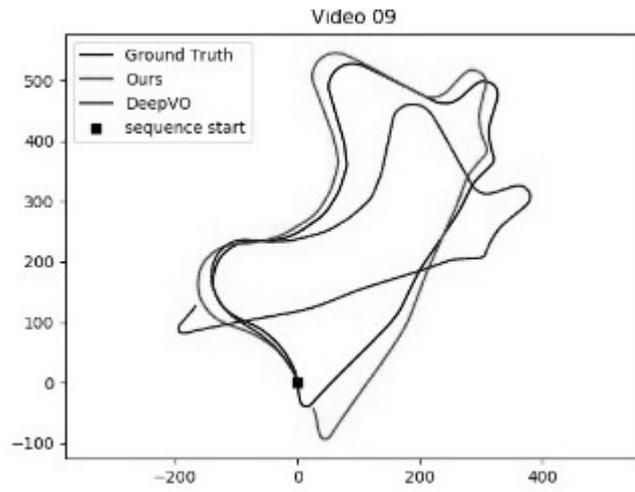


图9

序列号	Ours		VISO2_M		VISO2_S		ResNet	
	$t_{\text{rel}}(\%)$	$r_{\text{rel}}(^{\circ})$	$t_{\text{rel}}(\%)$	$r_{\text{rel}}(^{\circ})$	$t_{\text{rel}}(\%)$	$r_{\text{rel}}(^{\circ})$	$t_{\text{rel}}(\%)$	$r_{\text{rel}}(^{\circ})$
04	3.51	1.16	4.69	4.49	2.12	2.12	--	--
05	3.05	2.61	19.22	17.58	1.53	1.60	7.43	4.21
06	4.98	4.97	7.30	6.14	1.48	1.58	--	--
07	3.93	5.90	23.61	29.11	1.85	1.91	9.85	7.36
10	7.99	2.03	41.56	32.99	1.17	1.30	16.90	8.72
平均值	4.69	3.33	17.48	16.52	1.89	1.96	11.39	6.76

图10