

融合深度学习和几何定姿的光流视觉里程计研究

李鹏¹, 刘琪¹, 何宸宇¹, 满春涛², 刘宗明³

(1. 湘潭大学 自动化与电子信息学院 湖南 湘潭 411105; 2. 哈尔滨体育学院 哈尔滨 150080;

3. 上海航天控制技术研究所 上海 201109)

摘要: 针对多视图几何视觉里程计鲁棒性差和尺度漂移问题, 提出一种融合深度学习和几何定姿的光流视觉里程计。首先通过深度学习对两帧图像间光流和单目深度进行估计, 从光流估计中得到 2D-2D 对应关系, 并求解本质矩阵, 得到相机相对位姿; 然后从像素对应关系中求解三角化深度, 将预测深度与三角化深度进行尺度自适应对齐, 并将转换后的预测深度与三角化深度之间的深度误差加入损失函数中, 优化深度网络模型, 解决尺度不一致问题, 得到鲁棒视觉里程计 DOF-VO。在 KITTI 数据集上进行大量仿真实验, 结果表明该算法在通用评价指标上与 ORB_SLAM2 和 SC-SfMLearner 相比有明显提升, 在轨迹跟踪上具有良好的表现, 验证了该算法的有效性。

关键词: 深度学习; 视觉里程计; 位姿估计; 光流估计; 三角测量; 尺度

DOI: 10.15938/j.emc.2020.12.017

中图分类号: TP 391.41

文献标志码: A

文章编号: 1007-449X(2020)12-0142-10

Research on optical flow visual odometer combined deep learning and geometric attitude determination

LI Peng¹, LIU Qi¹, HE Chen-yu¹, MAN Chun-tao², LIU Zong-ming³

(1. School of Automation and Electronic Information, Xiangtan University, Xiangtan 411105, China; 2. Harbin Sport University, Harbin 150080, China; 3. Shanghai Institute of Spaceflight Control Technology, Shanghai 201109, China)

Abstract: Aiming at the poor robustness and scale drift of multi-view geometry vision odometer, an optical flow vision odometer with deep learning and geometric attitude determination is proposed. Firstly, the optical flow and monocular depth between two images were estimated by deep learning. The 2D-2D correspondence was obtained from the optical flow estimation, and the camera relative pose was received by solving the essential matrix. Then, the triangulated depth was got from the pixel correspondence, and the predicted depth and the triangulated depth were scale-aligned adaptively. The depth error between the converted prediction depth and the triangulation depth was added to the loss function to optimize the depth network model, and the robust visual odometer DOF-VO was deduced to overcome the scale inconsistency. Extensive simulation experiments on the KITTI dataset show that the algorithm has a significant improvement compared to ORB_SLAM2 and SC-SfMLearner in the general evaluation index, and has good performance in trajectory tracking, which verifies the validity of the algorithm.

Keywords: deep learning; visual odometer; pose estimation; optical flow estimation; triangulation; scale

收稿日期: 2020-07-03

基金项目: 国家自然科学基金(61773330); 湖南省科技厅重点研发计划项目(2018GK2014)

作者简介: 李 鹏(1978—), 男, 博士, 副教授, 研究方向为机器视觉和先进控制理论;

刘 琪(1996—), 男, 硕士研究生, 研究方向为视觉 SLAM 和机器人控制;

何宸宇(1996—), 男, 硕士研究生, 研究方向为机器人软件设计和路径规划算法;

满春涛(1965—), 男, 博士, 教授, 研究方向为工业过程稳态优化;

刘宗明(1986—), 男, 博士, 高级工程师, 研究方向为机器视觉。

通信作者: 刘宗明

0 引言

随着科技的发展,人们对移动机器人的自主性要求越来越高,在陌生环境中实现移动机器人的自主导航成为机器人领域的研究热点^[1]。基于视觉的同时定位与地图构建(simultaneous localization and mapping, SLAM)是实现这一技术的核心,视觉里程计(visual odometry, VO)是根据图像帧的运动估计相机位姿,是视觉 SLAM 中关键的一环,在无法预知地图的情况下,视觉里程计对于定位来说至关重要^[2]。

传统视觉里程计都是基于多视图几何的方法,在理想的场景下,它的精度是可靠且准确的。ORB-SLAM2^[3]是目前最精确的 SLAM 系统之一,但是它的视觉里程计基于稀疏 ORB 特征的方法,对特征点的提取和匹配有着较高要求,在动态场景下,定位比较困难。基于直接法的 VO 中比较经典的是 DSO-SLAM^[4],该方法直接基于关键帧的稀疏视觉里程计,对光照变化敏感,闭环难度大。传统视觉里程计方法,在实际运用场景下存在尺度漂移问题。

随着深度学习技术的不断发展,基于深度学习的视觉里程计逐渐成为热门。Agrawal^[5]等建议从运动估计任务中学习良好的视觉特征,该方法能够对相机相对姿态进行有效估计。Zhou^[6]等提出将运动结构作为一个监督学习问题,以端到端的方式学习单目深度和位姿。Tinghui Zhou^[7]等提出了 SfM-Learner 方法,第一个提出用自监督方法将相机位姿和深度进行联合学习。Jiawang Bian^[8]等在 SfM-Learner 的基础上提出了 SC-SfMLearner 方法,通过加强深度一致性来解决尺度不一致问题。上述基于学习的方法都是在假设所有图像的深度图和位姿尺度一致的基础上进行的,且没有明确考虑相机运动而存在的多视图几何约束,导致泛化能力降低。在经典光流理论的基础上,很多学者发现深度神经网络在特征提取和相邻帧之间的密集对应估计方面具有强大的性能。Dosovitskiy^[9]等首次提出将 CNN 引入到光流估计领域,提出了 FlowNet 网络结构,输入为两张连续的图像,直接利用 CNN 进行端到端的训练,输出为原图大小的光流图。Ilg^[10]等在 FlowNet 的基础上提出了 FlowNet2.0,通过新数据集合成及训练方法改进、多个网络堆叠、小位移网络设计等三个策略,显著提升了光流估计的性能。

本文针对上述问题,提出了融合深度学习和几何定姿的光流视觉里程计,主要贡献有:(1)设计了光流估计卷积神经网络模型,该模型由特征金字塔、光流估计器和上下文网络三部分组成。采用特征金字塔网络对不同尺度的图像进行特征提取,采用空洞卷积来整合光流的上下文信息,结合逐层学习的方法,缩小光流模型,改进光流估计性能;(2)将预测深度与三角化深度之间的深度误差引入损失函数,优化深度网络模型,改善位姿估计尺度不一致问题,得到鲁棒视觉里程计 DOF-VO;(3)在 KITTI 数据集上进行了大量仿真实验,验证了算法的有效性,并与其他典型方法进行了对比,结果表明,本文视觉里程计 DOF-VO 在轨迹跟踪上具有良好的表现,尺度漂移问题得到有效改善。

1 融合深度学习的相对位姿估计

1.1 传统的光流估计

经典的 LK(Lucas-Kanade)光流算法是一种两帧差分的光流估计算法,它利用两帧图像之差来追踪特征点瞬时速度,通过迭代计算得到稀疏的光流场。为了将光流估计进行建模,LK 算法做了两个重要的假设,分别是亮度不变假设和空间一致性假设。

假设 t 时刻,位于 (x, y) 像素位置的物体,在 $t + \Delta t$ 时刻位于 $(x + u, y + v)$ 位置,基于亮度不变假设,有

$$I(x, y, t) = I(x + u, y + v, t + \Delta t) \quad (1)$$

将等式右边进行一阶泰勒展开得

$$I(x + u, y + v, t + \Delta t) \approx I(x, y, t) + I_x u + I_y v + I_t \Delta t \quad (2)$$

即有

$$I_x u + I_y v + I_t \Delta t = 0 \quad (3)$$

写成矩阵形式有

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \Delta t \quad (4)$$

式中, I_x, I_y 分别为 (x, y) 像素点处图像亮度在 x 方向和 y 方向的偏导数,即图像 x 方向和 y 方向的梯度。 I_t 为 t 时刻 (x, y) 处像素亮度对时间的导数, $I_t \Delta t$ 即为两图之间的坐标位置的亮度差。

给定两张图像, $I_x, I_y, I_t \Delta t$ 均为已知量, u, v 即为待求的光流。式(4)为二元一次方程,仅由其无法求解算出运动参数 u, v ,LK 光流对此进行了空间一致性假设,即假设 $w \times w$ 窗口内的像素具有相同的运动,有:

$$\begin{bmatrix} I_x^{(1)} & I_y^{(1)} \\ I_x^{(2)} & I_y^{(2)} \\ \vdots & \vdots \\ I_x^{(n)} & I_y^{(n)} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_t^{(1)} \Delta t \\ -I_t^{(2)} \Delta t \\ \vdots \\ -I_t^{(n)} \Delta t \end{bmatrix} \quad (5)$$

上式即为 $Ax = b$ 的形式,可求得光流 $x = (u, v)$ 的最小二乘解:

$$x = (A^T A)^{-1} A^T b \quad (6)$$

考虑物体的运动速度较大时,算法会出现较大的误差的情况。Lucas-Kanade 改进算法^[11]借助了图像金字塔(Pyramid)的方法,将源图像逐层缩放,当图像为 400×400 时,物体速度为 $[16, 16]$,那么

图像缩小为 100×100 时,速度减少到 $[4, 4]$ 。在高层低分辨率图像上,大的偏移将变为小的偏移,通过生成原图像的金字塔图像,逐层求解光流。

1.2 基于深度学习的光流估计

深层神经网络将光流估计看作一个学习问题,采用端到端的网络模型对光流进行估计,在特征提取和相邻图像帧之间的密集对应估计方面具有很好的效果,在速度和准确率上都超过了传统光流算法。光流估计的网络架构^[12-14]包括特征金字塔提取网络、光流估计网络和上下文网络,整体网络架构如图1所示。

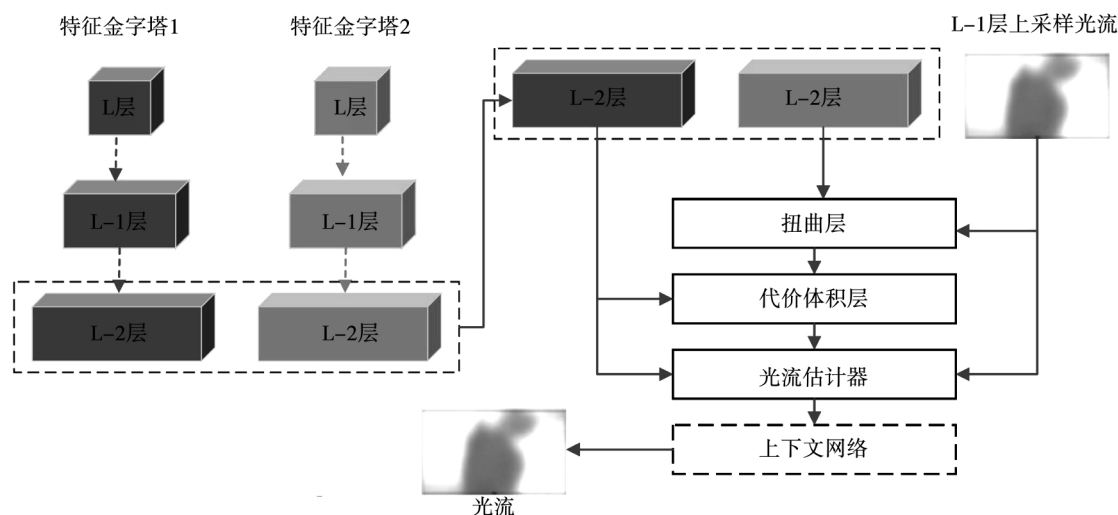


图1 光流网络架构

Fig.1 Architecture of optical flow network

特征金字塔提取网络中,因为原始图像的光照一直变化,将传统的图像金字塔方法变换为可学习的特征金字塔。给定两帧输入图像 I_1, I_2 , 两帧图像使用相同的孪生神经网络进行编码,金字塔层数为6,每一层包含一个卷积层和一个卷积下采样层,每个卷积层后面都有一个 Leaky ReLU 单元,结构如图2(a)所示。

在光流估计网络中,本文将传统方法中的扭曲运算作为网络中的一层来估计大运动,由于代价体积(cost volume)表示两幅图像各自像素的匹配程度,比原始图像更能区分光流,网络中设计了一个代价体积层,用于存储下一帧像素与其对应帧像素相关联的匹配代价,光流估计器用来估计光流。扭曲层和代价层不包含可训练参数。光流估计器一个多层的卷积神经网络,它的输入是代价体积、第一帧图像当前层的特征和下一层的上采样光流,输出当前

层的光流。不同金字塔层的光流估计不共享相同的参数,而是有自己的参数,重复这个估计过程,直到第0层。除最后一个卷积层外,其他卷积层后面都有一个 Leaky ReLU 单元, DenseNet 保证网络中层与层之间最大程度的信息传输的前提下,直接将所有层连接起来,故使用 DenseNet 连接来增强光流估计器架构性能,金字塔第2层光流估计器网络结构如图2(b)所示。

上下文网络从光流估计器获取估计光流和特征,使用上下文信息来优化光流,上下文网络结构设计基于空洞卷积,它由7个卷积层组成,这些层有不同的膨胀常数,能够有效地将每个输出单元的感知野大小扩大到所需的金字塔级,而不会引起很大的计算负担。结构如图2(c)所示,每个卷积层的最后一个数字表示膨胀系数。

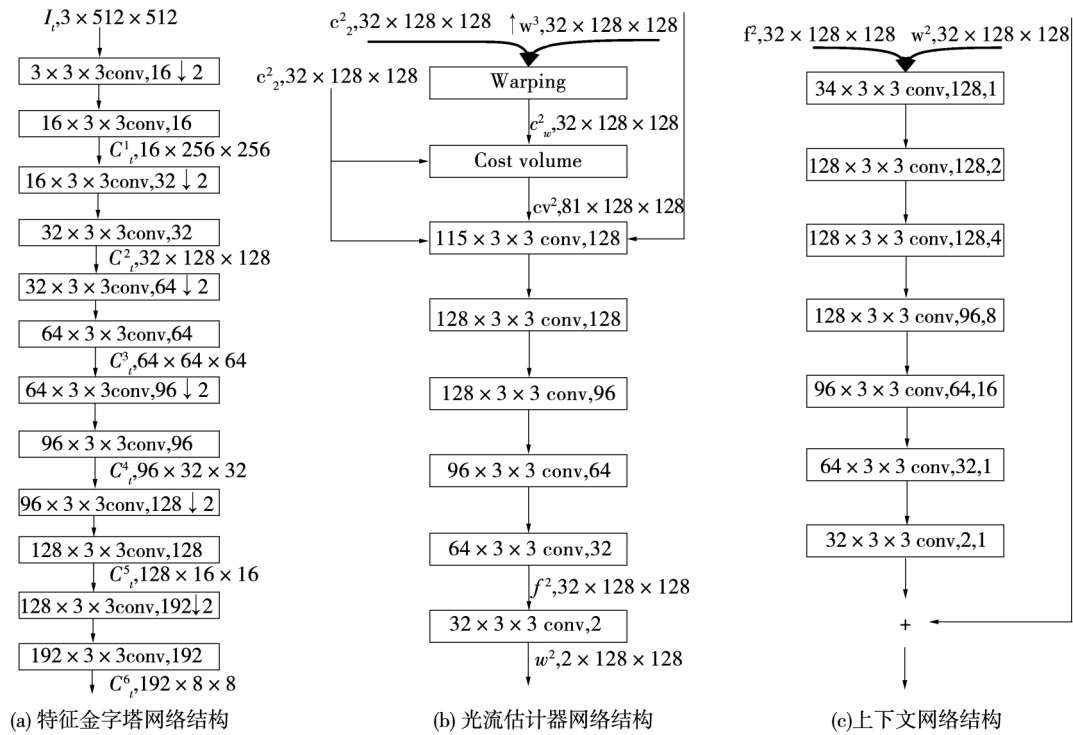


图 2 光流网络结构图

Fig. 2 Optical flow network structure diagram

光流估计算法: (1) 以 I_t 和 I_{t+1} 两帧图像为输入, 对两帧图像分别进行卷积下采样, 将其编码空间分辨率逐层降低的特征金字塔 (0, 1, 2, ..., L, 0 层为原始图像); (2) 光流编码器以从低到高的方式从相应分辨率的金字塔层中提取特征, 从金字塔顶层 L

层估计光流的上采样光流来扭曲 I_{t+1} 帧 $L-1$ 层特征, 计算两帧图像间的代价体积 (cost volume), 作为光流估计器的输入, 输出为当前分辨率的光流; (3) 重复该过程得到原始分辨率光流。光流估计算法框图如图 3 所示。

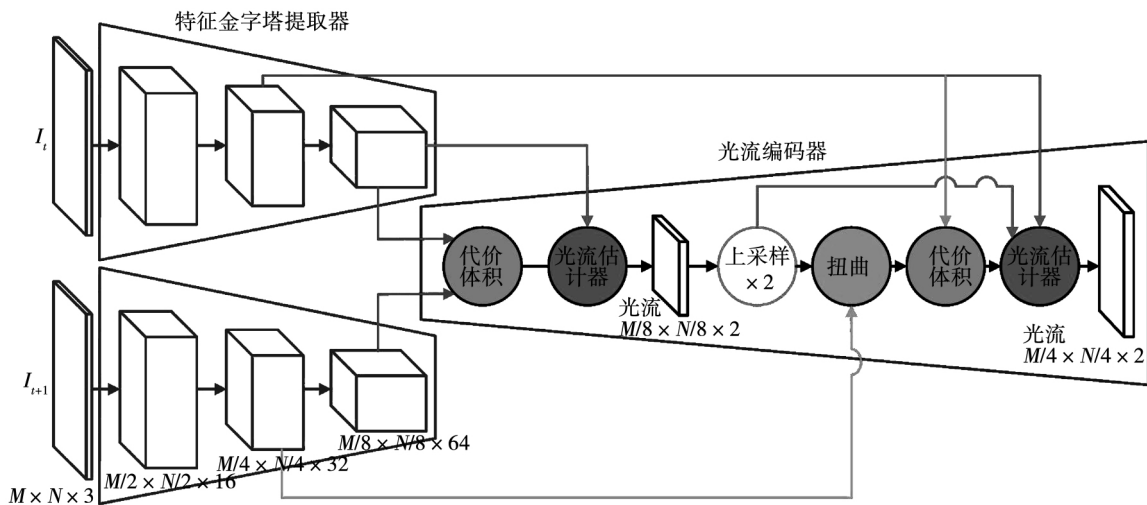


图 3 光流估计算法框图

Fig. 3 Diagram of optical flow estimation algorithm

2D-2D 对应关系的精度对于准确估计相对位姿具有重要意义, 为了提高光流估计精确度, 采用前后

向光流一致性^[15]作为衡量对应关系的标准来滤除光流异常值, 选择一致性评估中表现最佳的光流来

形成 2D-2D 匹配,用于估计相机相对姿态。一致性评估标准

$$M_s = 1/(0.1 + d_f) \quad (7)$$

式中, d_f 为前后向光流对应像素距离。光流估计效果图如图 4 所示,其中图 4(a) 是两帧图像间的 2D-2D 匹配,图 4(b)、(c) 是前后向光流。

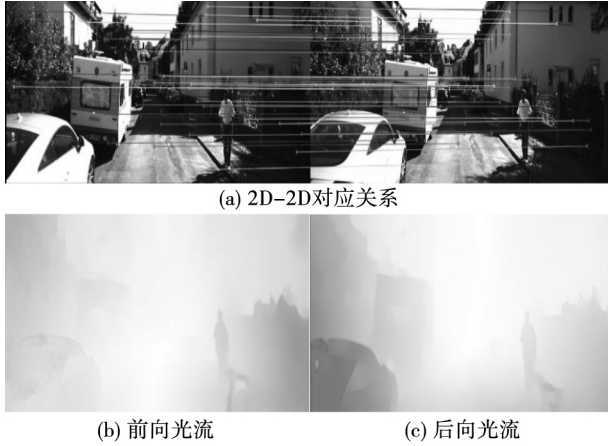


图 4 光流估计图

Fig. 4 Optical flow estimation diagram

1.3 光流损失函数文

光流估计网络中损失函数由光度损失和光流平滑度损失组成。光度损失是由参考图像和扭曲图像间的像素差异造成的,采用最小化源图像的光度损失函数 L_f 来克服视野外像素和遮挡有关问题^[16]。

$$L_f(I_i, I_j^i) = \frac{\alpha}{2} (1 - SSIM(I_i, I_j^i)) + (1 - \alpha) \|I_i - I_j^i\| \quad (8)$$

式中 $\alpha = 0.85$, I_i 和 I_j^i 是参考图像和合成图像, $SSIM$ 是结构相似性,是衡量两幅图像相似度的指标。

光度平滑度损失是由光度损失在低纹理场景或均匀区域无法提供充分信息而造成,为解决该问题,对光流估计进行正则化前先引入平滑度损失函数 L_s :

$$L_s(D_i, I_i) = |\partial_x D_i| e^{-|\partial_x I_i|} + |\partial_y D_i| e^{-|\partial_y I_i|} \quad (9)$$

式中: I_i 和 D_i 是图像及其深度; ∂ 是沿空间方向的一阶导数,它可以确保平滑度受图像边缘的影响。

1.4 对极几何

给定两帧图像 I_1, I_2 , 估计相机相对姿态的基本方法是求解本质矩阵, 即当两帧图像之间形成 2D-2D 像素(p_1, p_2) 对应关系时, 采用对极约束求解本质矩阵, 从而得到相机的相对姿态 $[R \ t]$ 。

如图 5 所示, 两帧图像 I_1, I_2 分别为相机在 O_1 ,

O_2 两位置的图像, 共同观测到 P 点, 旋转平移矩阵分别为 R, t , (p_1, p_2) 为空间中一对匹配点, 左图投影点像素坐标 $p_1 = [u_1 \ v_1 \ 1]^T$, 右图投影点 $p_2 = [u_2 \ v_2 \ 1]^T$ 。以左相机 O_1 相机为参考坐标系有:

$$\begin{cases} s_1 p_1 = K P, \\ s_2 p_2 = K(RP + t) \end{cases} \quad (10)$$

K 为相机内参矩阵, 可逆, 可得:

$$\begin{cases} s_1 K^{-1} p_1 = P, \\ s_2 K^{-1} p_2 = RP + t \end{cases} \quad (11)$$

式中: P 为观测点 P 的空间位置; s_1, s_2 分别为 P 点在两相机坐标系 O_1, O_2 下的深度, 取 $x_1 = K^{-1} p_1$, $x_2 = K^{-1} p_2$, 如果使用齐次坐标, 则等价于:

$$x_2 = R x_1 + t \quad (12)$$

两边同时左乘 t^A 得到:

$$t^A x_2 = t^A R x_1 \quad (13)$$

两边同时左乘 x_2^T 得到:

$$x_2^T t^A x_2 = x_2^T t^A R x_1 = 0 \quad (14)$$

将 $x_1 = K^{-1} p_1, x_2 = K^{-1} p_2$ 代入可得:

$$p_2^T K^{-T} t^A R K^{-1} p_1 = 0 \quad (15)$$

式(15)即为对极约束公式, 简洁地给出了两个匹配点的空间位置关系。

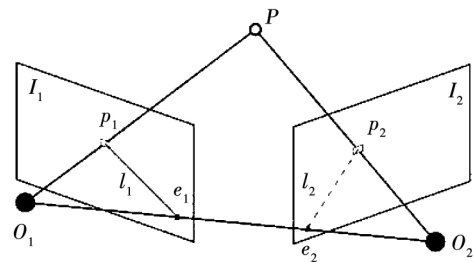


图 5 对极几何

Fig. 5 Epipolar geometry

光流为每个像素提供对应关系, 而其中一些像素是有噪声的, 不适合求解本质矩阵。因此本文使用前后向光流一致性评估来选择可靠的对应关系, 对非遮挡区域和前后向光流一致性评估中表现最佳的前 15% 的对应关系进行采样, 然后利用 RANSAC 算法随机抽取 6 K 个样本, 通过归一化“八点法”求解本质矩阵 E , 然后通过 SVD 分解求得 $[R \ t]$ 。

2 尺度对齐

2.1 深度估计

得到相机的相对位姿后, 需要估计地图点的空间位置, 由于在深度学习训练过程中不能保证图像

片段之间使用一致的比例,所以仍然存在尺度不一致的问题。从基于光流估计的匹配关系中得到三角化深度,将预测深度与三角化点云深度进行尺度对齐,并将转换后的预测深度与三角化深度之间的深度误差加入损失函数中,作为深度预测模型训练过程的监督,解决尺度不一致问题,使学习问题更简单,提高泛化能力。

尽管大多数三角化深度的效果都不错,但被遮挡区域和极点区域周围的深度值是有噪声的。对精确的像素匹配重新进行采样,通过三角测量得到稀疏的三角化深度,并采用中点三角测量方法^[17]来求解三角化深度。如图 6 所示,理想情况下 O_1p_1 和 O_2p_2 会相交于空间中的一点,但是由于图像分辨率以及噪声的存在,实际的情况更可能是图中所描述的那样: O_1p_1 和 O_2p_2 在空间中并没有交点,中点三角测量方法通过最小化从该点到所有光线的距离的平方和来确定最佳点,如式(16)所示:

$$e = \sum_{i=1}^N d(L_i, x)^2. \quad (16)$$

式中 d 是空间点到相机相应光线的距离。在两视图的特定情况下,选择 O_1p_1 与 O_2p_2 之间的公垂线,取其上的中点作为估计的最佳空间位置 P 。

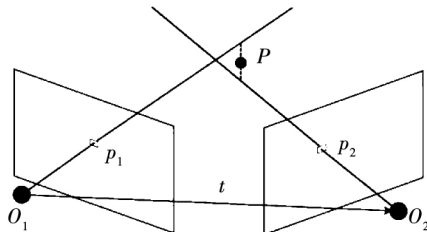


图 6 三角测量

Fig. 6 Triangulation

预测深度模型本文采用带有跳跃连接的标准全卷积编码器/解码器网络架构^[18-20], ResNet18 作为编码器, DispNet 作为解码器, 深度预测效果图如图 7 所示。



图 7 深度预测

Fig. 7 Depth prediction

2.2 深度损失函数

深度损失函数由边缘感知深度平滑度损失^[21]和三角测量深度误差损失组成, 边缘感知深度平滑度损失与式(9)一致。将预测深度与三角化点云深度进行尺度自适应对齐, 并将转换后的预测深度与三角化深度之间的深度误差最小化, 以最小误差作为深度损失进行反向传播, 深度误差损失 L_d 如下:

$$L_d = \left(\frac{D_{tri} - sD}{D_{tri}} \right)^2. \quad (17)$$

式中: D 是深度模型预测深度; D_{tri} 是三角测量得到的伪地面真实深度; s 是自适应尺度。

3 算法流程

本文视觉里程计算法框架如图 8 所示, 主要流程如下:

1) 采用深度学习方法训练光流网络模型和深度网络模型, 得到图像的前后向光流图以及深度图;

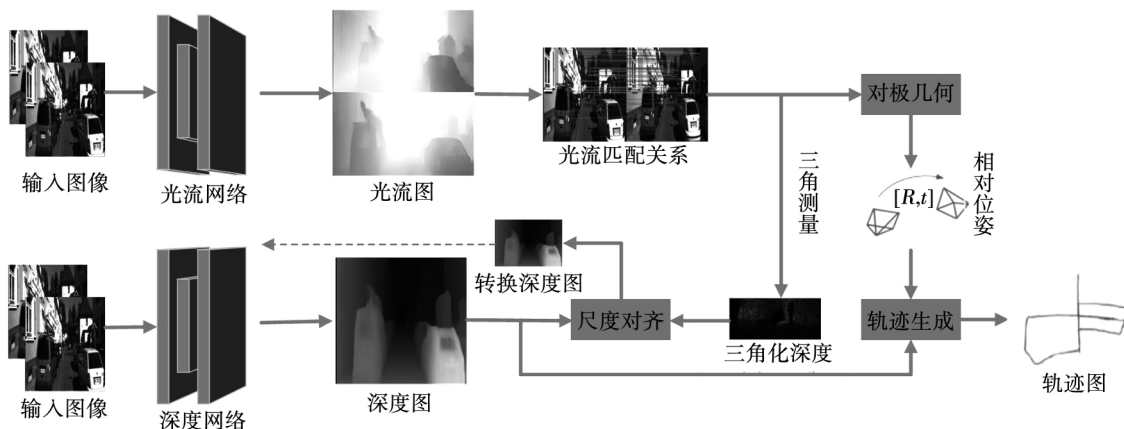


图 8 算法流程

Fig. 8 Flowchart of algorithm

2) 利用训练好的网络模型对两帧图像进行光流估计,得到两帧图像间的像素对应关系。选择前后向光流一致性评估中表现最佳的光流来形成 2D-2D 匹配关系;

3) 根据得到的 2D-2D 匹配关系,采用对极几何求解本质矩阵,从而得到相机的相对位姿 $[R, t]$;

4) 利用三角测量的方法从精确 2D-2D 匹配关系中得到三角化深度,将深度网络模型预测深度与三角化深度进行尺度对齐,并将转换后的预测深度与三角化深度之间的深度误差加入损失函数中,作为深度预测模型训练过程的监督,优化深度网络模型;

5) 采用三角测量深度对预测深度进行尺度矫正,预测深度的尺度与相对位姿进行匹配,从而优化尺度模糊问题,得到尺度一致的轨迹。

4 实验结果与分析

实验所使用的设备内存为 16 GB, CPU 为 Intel CORE i5-7500H, GPU 为 NVIDIA GTX1080Ti, 显存为 11GB, 运行系统为 Ubuntu16.04, 使用 Pytorch 深度学习框架对光流网络和深度网络进行训练。为了保证实验结果的有效性和实验方法对比的一致性,实验在同一硬件及软件环境下完成。

4.1 数据集与训练参数设置

实验采用 KITTI 视觉里程计数据集,为了使实

验结果可对比和评估,选取提供真实轨迹的 00 ~ 10 共 11 个序列,使用序列 00 ~ 07 进行训练,序列 08 ~ 10 进行测试并用来评估视觉里程计的性能。在训练过程中,使用 Adam 优化器对 20 个周期进行训练,其中 1 阶矩指数衰减率 $\beta_1 = 0.9$, 2 阶矩指数衰减率 $\beta_2 = 0.999$, 每一小批样本量为 8, 初始学习率设置为 10^{-4} , 经过 15 次迭代训练后调整为 10^{-5} , KITTI 数据集图片大小调整为 640×192 。

4.2 仿真实验结果与分析

为了验证本文视觉里程计的性能,对本文视觉里程计 DOF-VO 与采用 ORB_SLAM2 方法和 SC-SfMLearner 方法的视觉里程计进行仿真实验, 08 ~ 10 序列测试实验结果如图 9、图 10、图 11 所示,其中,灰色虚线代表真实轨迹 GT, 蓝色实线表示 DOF-VO 跟踪轨迹, 红色实线表示 ORB_SLAM2 跟踪轨迹, 绿色实线表示 SC-SfMLearner 跟踪轨迹。从图中可以看出 DOF-VO 对真实轨迹的跟踪效果较好,在精度方面优于 ORB_SLAM2 和 SC-SfMLearner 预测轨迹。

为了进行更详细的跟踪精度分析,在 00 ~ 10 这 11 个不同长度序列(100 m, 200 m, ..., 800 m)上计算平移矢量、旋转量、绝对轨迹误差、相对位姿误差的均方误差,并计算其均值作为视觉里程计的评价指标, DOF-VO 和 ORB_SLAM2、SC-SfMLearner 在平移矢量、旋转量、绝对轨迹误差、相对位姿误差指标上的对比如表 1、表 2 所示。

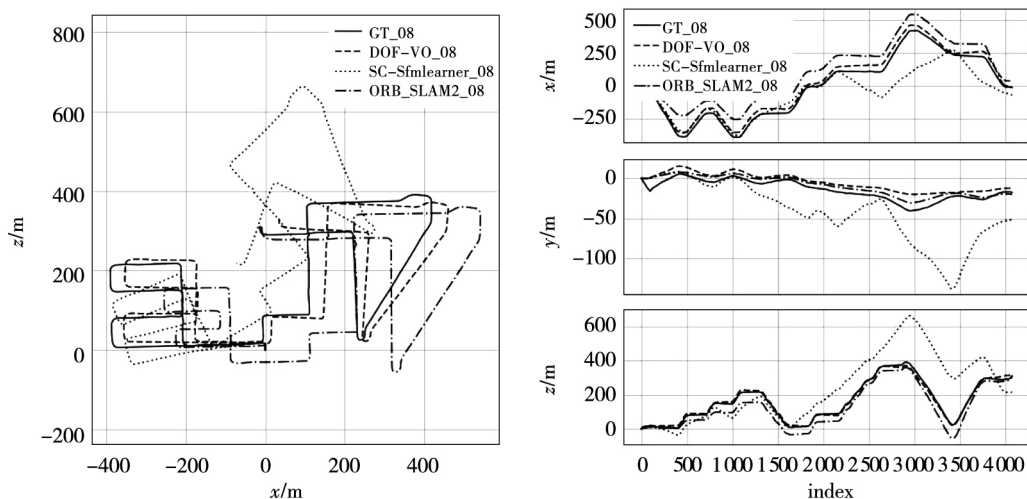


图9 序列08轨迹

Fig. 9 Track of sequence 08

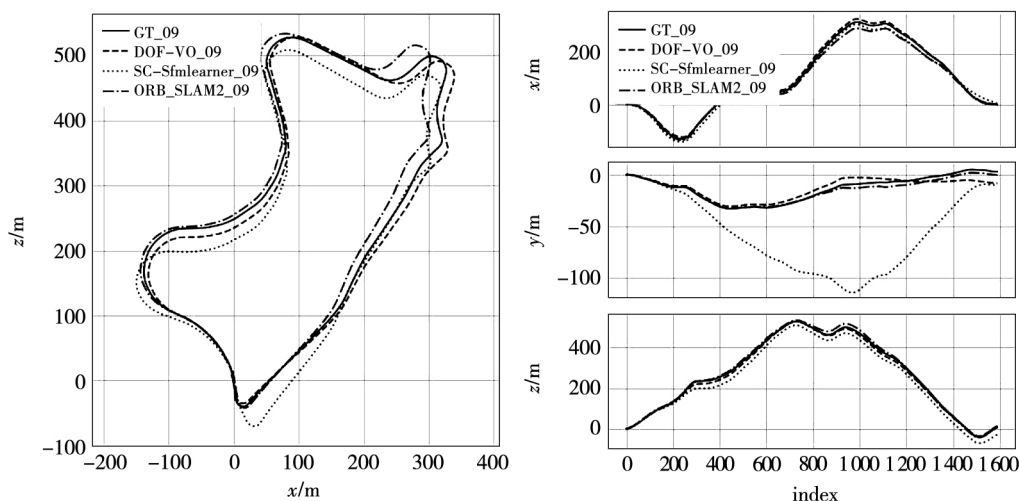


图 10 序列 09 轨迹

Fig. 10 Track of sequence 09

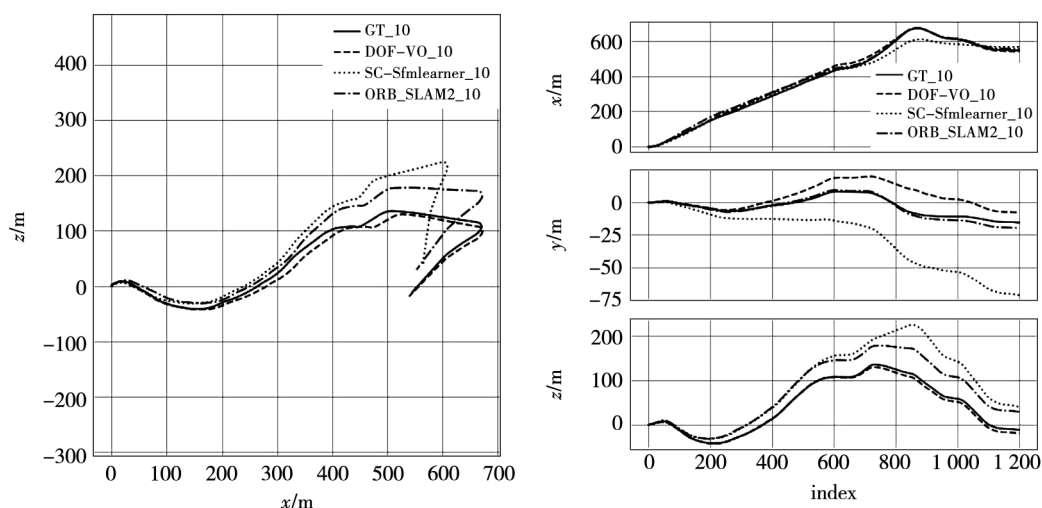


图 11 序列 10 轨迹

Fig. 11 Track of sequence 10

表 1 00-05 序列对比

Table 1 Comparison of 00-05 sequence

方法	评价指标	00	01	02	03	04	05
ORB_SLAM2	平移误差(%)	11.43	107.5	10.34	0.97	1.30	9.04
	旋转误差/(°)	0.58	0.89	0.26	0.19	0.27	0.26
	绝对轨迹误差	40.65	502.2	47.82	0.94	1.30	29.95
	相对位姿误差/m	0.169	2.970	0.172	0.031	0.078	0.140
	相对位姿误差/(°)	0.079	0.098	0.072	0.055	0.079	0.058
SC-SfmLearner	平移误差(%)	11.01	27.09	6.74	9.22	4.22	6.70
	旋转误差/(°)	3.39	1.31	1.96	4.93	2.01	2.38
	绝对轨迹误差	93.04	85.90	70.37	10.27	2.97	40.56
	相对位姿误差/m	0.139	0.888	0.092	0.059	0.073	0.070
	相对位姿误差/(°)	0.129	0.075	0.087	0.068	0.055	0.069
DOF-VO	平移误差(%)	1.83	4.70	3.07	1.19	0.769	1.846
	旋转误差/(°)	0.33	0.16	0.314	0.26	0.181	0.228
	绝对轨迹误差	12.93	19.7	22.67	1.25	0.608	6.758
	相对位姿误差/m	0.035	0.14	0.056	0.049	0.028	0.039
	相对位姿误差/(°)	0.05	0.03	0.04	0.077	0.034	0.039

表 2 06-10 序列对比
Table 2 Comparison of 06-10 sequence

方法		06	07	08	09	10	平均误差
ORB_SLAM2	平移误差(%)	14.56	9.77	11.46	9.30	2.57	8.074
	旋转误差/(°)	0.26	0.36	0.28	0.26	0.32	0.304
	绝对轨迹误差	40.82	16.04	43.09	38.77	5.42	26.48
	相对位姿误差/m	0.237	0.105	0.192	0.128	0.045	0.130
	相对位姿误差/(°)	0.055	0.047	0.061	0.061	0.065	0.063
SC-SfmLearner	平移误差(%)	5.36	8.29	8.11	7.64	10.74	7.803
	旋转误差/(°)	1.65	4.53	2.61	2.19	4.58	3.023
	绝对轨迹误差	12.56	21.01	56.15	15.02	20.19	34.21
	相对位姿误差/m	0.069	0.075	0.085	0.095	0.105	0.086
	相对位姿误差/(°)	0.066	0.074	0.074	0.102	0.107	0.083
DOF-VO	平移误差(%)	0.929	0.540	2.464	2.314	5.146	2.254
	旋转误差/(°)	0.136	0.230	0.319	0.303	0.499	0.269
	绝对轨迹误差	2.459	0.709	10.64	6.472	10.64	8.621
	相对位姿误差/m	0.043	0.021	0.043	0.051	0.065	0.052
	相对位姿误差/(°)	0.045	0.037	0.034	0.036	0.054	0.043

从表 1、表 2 的平均误差项可以看出 ,由于存在尺度漂移问题 ,几何方法 ORB_SLAM2 表现出较低的旋转误差和较高的平移误差 ,深度学习方法 SC-SfmLearner 在相对位姿估计上具备一定的优势 ,表现出较低的相对位姿误差 ,但平移误差和旋转误差较高。本文视觉里程计 DOF-VO 使用尺度一致的预测深度来进行尺度恢复 ,有效缓解了尺度漂移问题 ,与 ORB_SLAM 和 SC-SfmLearner 相比 ,在平移误差和旋转误差方面得到了显著改善 ,相对位姿误差明显降低 ,绝对轨迹误差有所下降 ,具有更好的定位精度。

5 结 论

本文提出了融合深度学习和几何定姿的光流视觉里程计 DOF-VO ,采用深度学习方法进行光流估计 ,并从光流估计中得到两帧图像间的 2D-2D 对应关系 ,从而得到相机相对位姿。实验结果表明 ,与几何方法相比 ,基于深度学习方法在图像帧对帧跟踪中具备更好的鲁棒性 ,能够得到更精确的相对位姿估计。将预测深度与三角化深度进行尺度自适应对齐 ,平移误差和旋转误差方面得到了显著改善 ,优化了尺度漂移问题。在 KITTI 数据集上进行大量实验 ,在 KITTI 数据集的 11 个序列中 ,本文的视觉里程计算法在轨迹跟踪上具有良好的表现 ,在平移矢

量、旋转量、绝对轨迹误差、相对位姿误差等评价指标上与几何方法 ORB_SLAM2 和深度学习方法 SC-SfmLearner 相比有明显提升 ,实验结果表明融合深度学习和几何定姿的光流视觉里程计方法具备一定的可行性和优越性。但也存在着一些不足:光流估计方法通过对估计的前向和后向光流进行一致性检查 ,然后将光流外推到遮挡区域 ,此时初始光流已经受到遮挡的不利影响 ,由于代价体积中也包含了遮挡相关的信息 ,可以合理利用这些信息。未来的工作主要对光流估计模型进行进一步改进 ,在代价体积计算之后直接在估计光流之前连接一个遮挡估计器 ,进一步减少误差 ,提高光流估计精度。

参 考 文 献:

[1] CADENAC ,CARLONE L ,CARRILLO H ,et al. Past , present , and future of simultaneous localization and mapping: toward the robust-perception age [J]. IEEE Transactions on Robotics , 2016 , 32 (6) : 1309.

[2] Fuentes-Pacheco , Jorge , et al. Visual simultaneous localization and mapping: a survey[J]. Artificial Intelligence Review , 2015 , 43(1) : 55.

[3] MURARTAL R , TARDOS J D. ORB-SLAM2: An open-source slam system for monocular , stereo , and RGB-D cameras [J]. IEEE Transactions on Robotics , 2017 , 33(5) : 1255.

[4] ENGEL J , KOLTUN V , CREMERS D , et al. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine In-

- telligence, 2018, 40(3): 611.
- [5] AGRAWAL P, CARREIRA J, MALIK J, et al. Learning to see by moving[C]. //2015 IEEE International Conference on Computer Vision (ICCV), December 7 – 13, 2015, Santiago, USA. 2015: 37 – 45.
- [6] ZHOU H, UMMENHOFER B, BROX T, et al. DeepTAM: Deep Tracking and Mapping[C]. //European conference on computer vision, September 8 – 14, 2018, Munich, Germany. 2018: 851 – 868.
- [7] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21 – 26, 2017, Honolulu, USA. 2017: 6612 – 6619.
- [8] BIAN J, LI Z, WANG N, et al. Unsupervised scale-consistent depth and ego-motion learning from monocular video[C]. //Neural information processing systems, October, 10 – 12, 2019, Vancouver, Canada. 2019: 35 – 45.
- [9] DOSOVITSKIY A, FISCHERY P, ILG E, et al. FlowNet: learning optical flow with convolutional networks[C]. //2015 IEEE International Conference on Computer Vision (ICCV), December 7 – 13, 2015, Santiago, Chile, USA. 2015: 2758 – 2766.
- [10] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21 – 26, 2017, Honolulu, USA. 2017: 1647 – 1655.
- [11] YIC, LIYUN C, CHUNGUANG L. Moving target tracking algorithm based on improved optical flow technology[J]. Open Automation & Control Systems Journal, 2015, 7(1): 1387.
- [12] HUI T, TANG X, LOY CC, et al. LiteFlowNet: a lightweight convolutional neural network for optical flow estimation[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18 – 23, 2018, Salt Lake City, USA. 2018: 8981 – 8989.
- [13] KENDALL A, GRIMES M, CIPOLLA R. PoseNet: A convolutional network for real-time 6-dof camera relocalization[C]. //2015 IEEE International Conference on Computer Vision (ICCV), December 7 – 13, 2015, Santiago, Chile, USA. 2015: 2938 – 2946.
- [14] WANG Y, YANG Y, YANG Z, et al. Occlusion Aware Unsupervised Learning of Optical Flow[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18 – 23, 2018, Salt Lake, USA. 2018: 4884 – 4893.
- [15] YIN Z, SHI J. GeoNet: unsupervised learning of dense depth, optical flow and camera pose[C]. //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18 – 23, 2018, Salt Lake, USA. 2018: 1983 – 1992.
- [16] GODARD C, AODHA O M, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]. //International conference on computer vision, October 27–November 2, 2019, Seoul, Korea. 2019: 3828 – 3838.
- [17] YANG K, FANG W, ZHAO Y, et al. Iteratively reweighted midpoint method for fast multiple view triangulation[J]. IEEE Robotics and Automation Letters, 2019, 4(2): 708.
- [18] RONNEBERGER O, FISCHER P, BROXT, et al. U-Net: convolutional networks for biomedical image segmentation[C]. //Medical image computing and computer assisted intervention, October 5 – 9, 2015, Munich, Germany. 2015: 234 – 241.
- [19] GODARD C, AODHA O M, BROSTOW G J, et al. Unsupervised monocular depth estimation with left-right consistency[C]. //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21 – 26, 2017, Honolulu, USA. 2017: 6602 – 6611.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 26 – July 1, 2016, Las Vegas, USA. 2016: 770 – 778.
- [21] WANG C, BUENAPASADA J M, ZHU R, et al. Learning depth from monocular videos using direct methods[C]. //2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18 – 23, 2018, Salt Lake, USA. 2018: 2022 – 2030.

(编辑: 刘素菊)