DOI:10.19651/j.cnki.emt.2108299

宽视角相机相对姿态测量方法研究*

李威静[1,2] 王 鉴[1,2] 郭锦铭[1,2] 张 璇[1,2] 韩 焱[1,2]
(1.中北大学山西省信息探测与处理重点实验室 太原 030051;2.中北大学信息与通信工程学院 太原 030051)

摘 要: 针对目前相机姿态估计方法都存在视觉局限性的问题,使用鱼眼镜头作为视觉传感器进行姿态估计,进而实现宽视角相机相对姿态估计。但鱼眼成像在具有宽视场优点的同时,伴随着严重的非线性畸变导致其在不同的方位和距离下具有不同畸变扩散的问题,为此提出了一种直接利用鱼眼图像的非线性进行相机相对姿态测量的方法。首先,构建鱼眼数据集 kitti_FE;其次,使用卷积神经网络进行特征提取后结合长短时记忆网络进行双向循环训练,实现相机相对姿态的端对端输出;最后利用迁移学习的方法对实际场景进行相机姿态估计。为了验证所提方法的鲁棒性和精确度,在相同实际场景下,利用所提方法分别与现有框架 CNN、DeepVO 和 CNN-LSTM-VO-cons 进行对比。实验表明,该方法分别比现有框架的相机姿态估计精度提高了 32%、29%和 25%,而且在高速运动下该方法更具有稳定性。

关键词: 宽视角运动姿态估计;非线性光学;鱼眼成像模型;深度学习
中图分类号: TP183;TP277 **文献标识码:** A **国家标准学科分类代码:** 510.4050;520.2060

Research on relative attitude measurement method of wide-FOV camera

Li Xianjing[1,2] Wang Jian[1,2] Guo Jinming[1,2] Zhang Xuan[1,2] Han Yan[1,2]
(1. Shanxi Key Laboratory of Signal Capturing and Processing, North University of China, Taiyuan 030051, China;
2. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China)

Abstract: In view of the visual limitations of the current camera attitude estimation methods. In order to realize the relative attitude estimation of wide-FOV(fields of view camera), this paper used fisheye lens as visual sensor for attitude estimation. While fisheye imaging has the advantages of a wide-FOV, it is accompanied by serious nonlinear distortions, which leads to the problem of different distortion diffusion at different azimuths and distances. Therefore, this paper proposed a method to directly use the non-linear characteristics of the fisheye image to measure the relative pose of the camera. First, established the fisheye dataset kitti_FE. Secondly, used convolutional neural network for feature extraction and then combined with long short-term memory network for bidirectional loop training to achieve the end-to-end output of the relative posture of the camera. Finally, the method of transfer learning was used to estimate the pose of the fisheye camera in the actual scene. Experiments show that the proposed method is 32%、29% and 25% higher than the camera pose estimation accuracy under the existing frameworks of CNN, DeepVO and CNN-LSTM-VO-cons, respectively, and the proposed method is more stable under high-speed motion.

Keywords: wide-FOV motion pose estimation;nonlinear optics;fisheye imaging model;deep learning

0 引 言

单目视觉实时定位与地图构建 SLAM(simultaneous localization and mapping)设备体积小采集成本较低,便于在未知环境中为无人平台提供了自身的位置信息和环境的结构信息,因此,在自动驾驶[1-3]、无人监控[4]和机器人[5-6]的研究中发挥着重要的作用。单目相机姿态估计是 SLAM 中的一个重要环节,对地图的构建有着重要的影响[7-8]。

对于现有相机姿态估计方法而言,虽然它们的算法流程不同,但是都有一个共同之处,即必须依赖于一个信息充分的观察环境,就是相邻两帧间必须有足够的重叠区域,只有这样才满足两帧图像相机相对姿态估计的条件[9]。然而由于常用传统相机视场有限,在许多实际情况下,例如,自动驾驶中汽车快速运动或者快速旋转时会导致采集的图像

收稿日期:2021-11-08
* 基金项目:国家自然科学基金(61801437,61871351,61971381)、山西省研究生创新项目(2021Y609)资助

相邻帧间重叠减少，这时会导致有效特征点无法匹配，最终无法进行位姿测量。由此可见，增加相机的视场范围可使相机位姿估计有更好的效果。近年来，众多学者提出了一些技术来扩大视觉传感器的视角覆盖范围，主要分为单镜头旋转/步进扫描技术、多镜头拼接技术和宽视角镜头凝视技术。由于单镜头旋转/步进扫描技术需要足够的时间，故上述方法的突出缺点之一就是无法实时的获得大视野图像，另外在实际应用中还受系统转动惯量、扫描系统的架构影响。多镜头拼接式的传感器成本高而且受到诸多因素的限制，在实际应用中可维修性差。毫无疑问，从各种使用性能考虑，单目宽视角凝视技术是最有应用前景的实用技术。其中，鱼眼镜头的视角可达到甚至超过 180°，比超广角镜头更具有宽视角环境感知优势且安装简单紧凑[10-12]，因此本文将单目鱼眼镜头作为视觉传感器进行宽视角相机相对姿态估计。

然而，利用鱼眼宽视野成像的同时伴随严重的非线性畸变，导致在不同的方位和距离下具有不同畸变扩散，从光轴不同角度看过去的物体大不相同，制约了它的应用。为了既实现小畸变、宽视场，需对非线性畸变图像进行矫正。矫正非线性畸变的方法主要有光学矫正法和算法矫正法。前者通过多组光学透镜将畸变图像矫正为线性图像，后者采用算法进行矫正，例如对鱼眼图像进行特征点匹配或者使用多项式畸变模型[13]进行径向畸变矫正。然而，光学矫正方法镜头结构复杂，在振动冲击环境下光学性能会发生变化；而算法矫正后的图像会存在不同程度上的伪影或者模糊，尤其是在图像边缘[14]。致使鱼眼图像相对姿态测量精度低或者有效特征点无法匹配，最终无法测量。

针对以上问题，本文提出通过鱼眼成像三维空间点到二维平面的非线性畸变特性，结合深度学习，直接对宽视角图像进行相对姿态估计。由于深度学习的方法缺乏大量的鱼眼相机姿态估计数据集，因此本文将自动驾驶通用数据集 kitti[15]中加入特定畸变变成具有鱼眼图像特征的虚拟鱼眼数据集进行预训练，然后利用迁移学习的方法在少量的实际场景样本中实现宽视角相机相对姿态估计，证明了本文所提方法的鲁棒性和准确性。为了便于实现迁移学习，所以本文使用相机的相对姿态作为输出标签。

1　鱼眼相机相对姿态测量研究方法

基于深度学习的相机姿态估计方法中，2015 年，Kendall 等[16]提出的 PoseNet 首次将深度学习用于相机姿态估计的参数回归中，室外定位误差约为 2 m，3°。2016 年，Walch 等[17]在 PoseNet 的基础上添加了 LSTM(long short-term memory)部分，从而提高了相机姿态估计精度。2017 年，Clark 等[18]提出的 VidLoc，首次采用 CNN (convolutional neural networks)＋LSTM 直接对图片进行位姿估计，在 PoseNet 的效果上有明显提升。同年，Wang 等[19]提出 DeepVO，同样将序列图像作为输入，然后采用 CNN 进行特征提取，其次输入到 RNN 网络学习帧间几何关系，与之前研究不同的是，该文献以多帧图片的相对姿态作为输出而且相对于前面所有的研究工作，效果非常好，为在相机姿态估计方面实现迁移学习提供了可能。2018 年，张林箭[20]提出的 CNN-LSTM-VO-cons，在 DeepVO 的基础上加入了逆序训练，效果相对于 DeepVO 有了一定的提高。本文在张林箭提出的 CNN-LSTM-VO-cons 训练框架上加入了迁移学习过程，研究了利用宽视场非线性畸变图像进行相机相对姿态测量的方法。

1.1　CNN-LSTM-VO-cons 整体框架介绍

CNN-LSTM-VO-cons 整体网络结构如图 1 所示，属于端到端的模型，每两张相邻的图片对应一个表示相机相对位姿的 1×6 维向量 $\boldsymbol{Lab}$，其中包括 1×3 维的位移变化量 $\Delta\boldsymbol{P}$ 和方向偏移量 $\Delta\boldsymbol{\Phi}$。因为两张图片之间的相对姿态可以分别以各自为基准，与图片的先后顺序相关。在常用框架卷积 CNN 进行特征提取后结合了 LSTM 进行双向循环训练，进一步提取两张图片正反序之间的几何特征，对相机相对姿态进行回归。整个网络分为左右两边，左半边和右半边的网络完全对称，唯一不同的是分别采用正逆序序列图像和其相对应标签作为输入输出。整体的损失函数由所有位移 $\boldsymbol{P}$ 和方向角 $\boldsymbol{\Phi}$ 的均方误差(MSE)组成，如式(1)所示。

$$Loss=\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\|\boldsymbol{P}_{1ij}-\hat{\boldsymbol{P}}_{1ij}\|_2^2+\beta_1\|\boldsymbol{\Phi}_{1ij}-\hat{\boldsymbol{\Phi}}_{1ij}\|_2^2+\|\boldsymbol{P}_{2ij}-\hat{\boldsymbol{P}}_{2ij}\|_2^2+\beta_2\|\boldsymbol{\Phi}_{2ij}-\hat{\boldsymbol{\Phi}}_{2ij}\|_2^2 \tag{1}$$

其中，N 为每个序列图像中的样本对数，M 为序列数，$(\boldsymbol{P}_{1ij},\boldsymbol{\Phi}_{1ij})$ 和 $(\boldsymbol{P}_{2ij},\boldsymbol{\Phi}_{2ij})$ 分别表示真实的第 i 个序列中第 j 个时刻的正序和反序输入的位移和旋转角，$(\hat{\boldsymbol{P}}_{1ij},\hat{\boldsymbol{\Phi}}_{1ij})$ 和 $(\hat{\boldsymbol{P}}_{2ij},\hat{\boldsymbol{\Phi}}_{2ij})$ 为预测结果，β_2 和 β_2 为尺度因子。

1.2　鱼眼非线性特性及数据集生成

基于深度学习的宽视场图像的目标测量，首先需要有大量的学习场景样本，然而，由于非线性畸变的多样性，目前尚无可借鉴的数据库，但国内外建立了许多的无畸变镜头的图像数据库。因此，可通过构建相机的非线性畸变函数的映射关系，由针孔图像库可产生畸变图像数据库。

1)鱼眼镜头成像模型及建立 RGB 图像数据集

鱼眼镜头在获取大视场的同时会发生非线性形变，传统的成像模型不再成立。鱼眼镜头的几种数学模型中，最普遍的是等距投影模型，如式(2)所示。针孔照相机的投影模型如式(3)所示。

$$r_f=f\theta \tag{2}$$

$$r=f\tan\theta \tag{3}$$

其中，θ 是光轴和入射光线的夹角，r 是图像点到原点的距离，f 是焦距。

由于目前的鱼眼数据集不够完善，不能满足本研究的要求。因此，本文使用等距鱼眼相机模型给自动驾驶通用

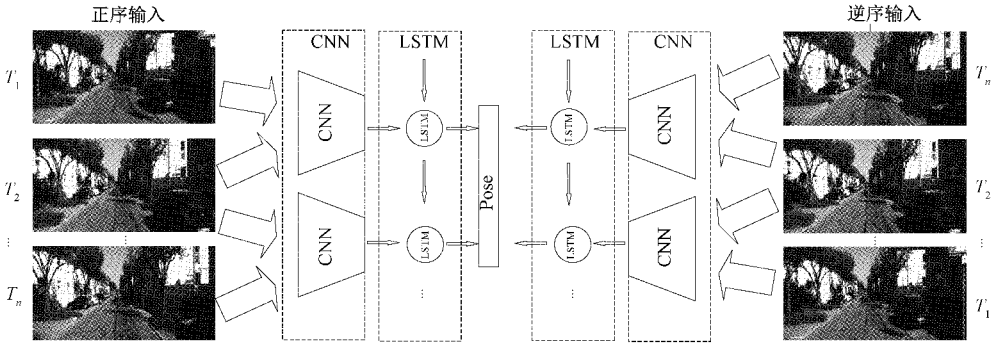


图 1 CNN-LSTM-VO-cons 网络总体结构

数据集 kitti 中加入畸变,使其变成具有鱼眼图像非线性特性的鱼眼数据集 kitti_FE(KITTI of fish eye),从而为实现迁移学习提供基础。根据式(2)和(3),可以对针孔图像($p_c=(u_c,v_c)$)和合成鱼眼图像($p_f=(u_f,v_f)$)上类似的像素点之间进行重新映射,合成具有不同畸变程度的虚拟鱼眼图像。如式(4)所示,虚拟鱼眼图像径向畸变的程度仅取决于 f 参数,效果如图 2 所示,本文将 f 设为 150。

$$d_c = f\tan(d_f/f) \tag{4}$$

其中,$d_c=\sqrt{(u_c-u_{cu})^2+(v_c-v_{cv})^2}$ 和 $d_f=\sqrt{(u_f-u_{fu})^2+(v_f-v_{fv})^2}$ 分别是针孔图像和虚拟鱼眼图像中单个像素点与主点之间的距离。

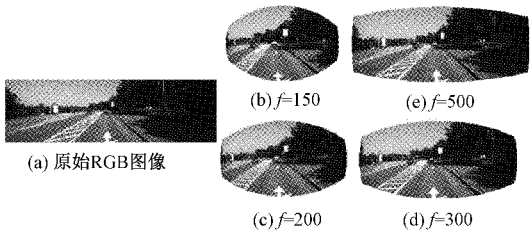


图 2 不同畸变程度的虚拟鱼眼图像

2)建立相对姿态标签

绝对姿态估计在相同场景中可以表现出比较高的定位精度,但是无法实现场景迁移,这意味着对于每个场景都必须单独训练一个模型,所以成本较高。所以本文将 kitti 数据集中实际测量的绝对姿态转换为相对姿态作为 kitti_FE 数据集的标签。

两张相邻图片的相对位姿旋转矩阵如式(5)所示。

$$\boldsymbol{M}_i=\boldsymbol{R}_{i|1}^{1}\boldsymbol{R}_i=\begin{bmatrix} & \Delta r & & \Delta x\\ & & & \Delta y\\ & & & \Delta z\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{5}$$

其中,$\boldsymbol{R}_i$ 表示第 i 张图像的齐次变换矩阵;$\boldsymbol{R}_{i-1}$ 为正序相邻帧图像的齐次变换矩阵;当逆序产生图像对时,相对姿态标签求解与正序类似,表示为 $\boldsymbol{M}_i=\boldsymbol{R}_{i-1}^{-1}\boldsymbol{R}_i$。最后用 6 维向量作为相邻两张图片的相对位姿标签如式(6)所示。

$$\boldsymbol{Lab}_i=[\Delta x \quad \Delta y \quad \Delta z \quad \Delta\varphi \quad \Delta\chi \quad \Delta\phi] \tag{6}$$

其中,$[\Delta x \quad \Delta y \quad \Delta z]$ 和 $[\Delta\varphi \quad \Delta\chi \quad \Delta\phi]$ 分别表示 $\boldsymbol{M}_i$ 中的位移向量和旋转部分经过变换得到的欧拉角。

1.3 鱼眼图像的相对姿态测量方法

本文的鱼眼相机相对姿态估计网络框架 CNN-LSTM-VO-cons-FE,在 CNN-LSTM-VO-cons 的基础上加入了迁移学习,并将输入改为 kitti_FE 数据集中时间序列上的两张相邻图片,对应标签为 1×6 维的相对姿态矩阵。

1)CNN 特征提取网络框架

CNN 特征提取部分的具体网络配置如表 1 所示。该网络包含 10 个卷积层,每一层后面都有一个整流线性单元(ReLU)激活。网络中卷积核的大小从 7×7 逐渐减小到 5×5,然后是 3×3,用于提取小的局部特征。其输入为尺寸一致的 RGB 图像,用以学习不同姿态下,图片中共同物体的几何畸变特性进行精确姿态测量。图片经过 10 层卷积层提取特征后得到 4×4×1 024 的特征图谱,最后将其拉为一维向量传递给 LSTM 进行顺序建模。

表 1 CNN 网络参数配置表

层次	卷积核	步长	权重	张量尺寸
Conv1	7×7	2	6×64	256×256×64
Conv2	5×5	2	64×128	128×128×128
Conv3	5×5	2	128×256	64×64×256
Conv3_1	3×3	1	256×256	64×64×256
Conv4	3×3	2	256×512	32×32×512
Conv4_1	3×3	1	512×512	32×32×512
Conv5	3×3	2	512×512	16×16×512
Conv5_1	3×3	1	512×512	16×16×512
Conv6	3×3	2	512×1 024	8×8×1 024
Conv6_1	3×3	1	1 024×1 024	8×8×1 024
Max-pooling	2×2	2	—	4×4×1 024

2)CNN+LSTM 单向框架

本文所用单向的基于 CNN+LSTM 单目宽视场相机姿态估计系统如图 3 所示。整个网络采用多帧图像作为输入,所有图片的大小都调为 512×512,假设序列长度为

$n+1$，当以滑动窗口形式顺序组合相邻两帧图片时，可以得到 n 组图片对。这些图片对都分别经过 CNN 网络进行特征提取，得到两两之间的代表其几何关系的特征张量（共 n 个），张量大小为 4×4×1 024。把这些张量看成时间序列上的特征，依次输入到两层 LSTM 网络中，经过时序上的处理后，LSTM 将在每个时刻产生一个输出。然后经过全连接层进行降维，得到 6 维的姿态向量，分别表示相邻两张图片之间的相对姿态 $[\Delta x \quad \Delta y \quad \Delta z \quad \Delta\varphi \quad \Delta\chi \quad \Delta\phi]$。

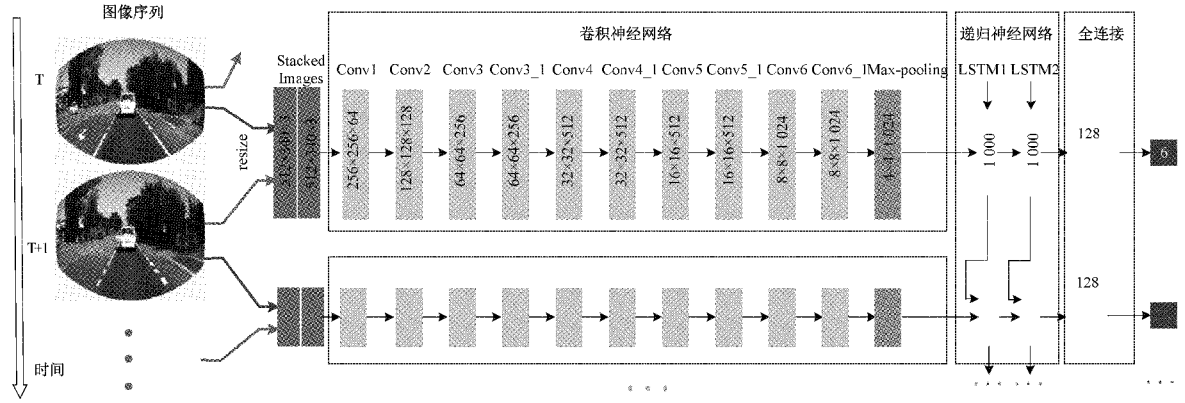


图 3　基于 CNN+LSTM 单目宽视场相机姿态估计系统数据流图

可以把基于 CNN+LSTM 单目宽视场相机姿态估计看成一个条件概率的问题：给定一个序列的单目 RGB 图像 $X=(X_1,X_2,\cdots,X_{n+1})$，计算得到序列图像中相邻图像对的相对姿态 $Y=(Y_1,Y_2,\cdots,Y_n)$ 的概率，如式(7)所示。

$$p(Y\mid X)=p(Y_1,Y_2,\cdots,Y_n\mid X_1,X_2,\cdots,X_{n+1}) \tag{7}$$

为了求得网络的最优参数 ω^*，需使得上述概率最大化，如式(8)所示。

$$\omega^*=\underset{\omega}{\operatorname{argmax}}p(X\mid Y;\omega) \tag{8}$$

对于 M 个序列，为了使参数 ω 最优化，需使得第 i 个序列中第 j 个时刻相对于下一时刻图像的真实姿态 $(\boldsymbol{P}_{ij},\boldsymbol{\Phi}_{ij})$ 和预测结果 $(\hat{\boldsymbol{P}}_{ij},\hat{\boldsymbol{\Phi}}_{ij})$ 之间的欧几里得距离最小化，如式(9)所示。

$$\omega^*=\operatorname{argmin}\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\|\boldsymbol{P}_{ij}-\hat{\boldsymbol{P}}_{ij}\|_2^2+\beta\|\boldsymbol{\Phi}_{ij}-\hat{\boldsymbol{\Phi}}_{ij}\|_2^2 \tag{9}$$

其中，N 为每个序列中的图像对数，β 为尺度因子，$\|\cdot\|_2$ 表示二范数。因此，本文整体损失函数如式(1)所示。由所有位移 $\boldsymbol{P}$ 和方向角 $\boldsymbol{\Phi}$ 的 MSE 组成，所用的双向循环网络的左边和 CNN+LSTM 单向框架是一致的，网络的右边和左边完全对称，只是输入为逆序图像。

3)CNN-LSTM-VO-cons-FE 框架

深度学习需要大量的高质量标注数据，是一项昂贵且耗时的任务而且实景图和合成鱼眼图像存在差异。因此，本文为了在少量的现实场景中达到好的相对姿态估计效果，利用迁移学习的方法，将上述在 kitti_FE 数据集的训练权重作为实际场景训练的初始化权重，然后将前面的卷积层冻结，只使用第 5、6 层卷积对参数进行微调后，传入 LSTM 中经过全连接层降至 1×6 维，实现了真实场景中的相机相对姿态估计，具体实现方法如图 4 所示。

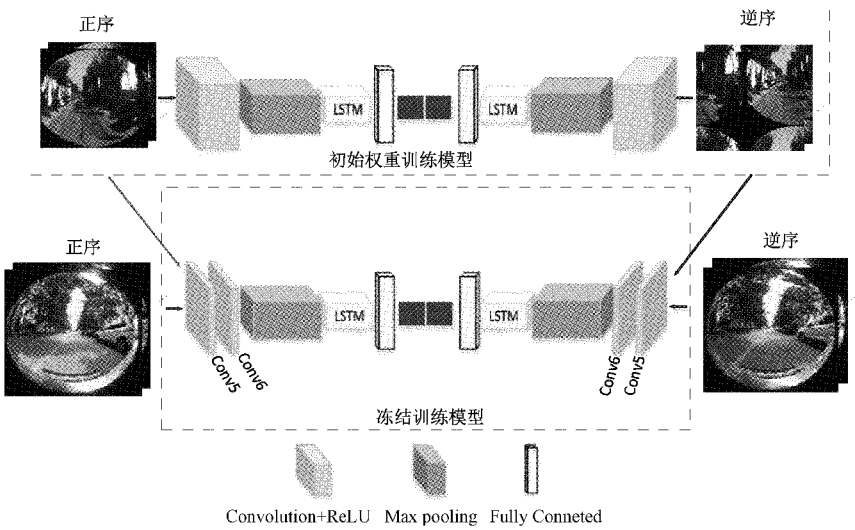


图 4　本文相机姿态估计 CNN-LSTM-VO-cons-FE 框架

2 实验与分析

2.1 实验环境

本文实验运行的服务器参数配置如表 2 所示。

表 2 服务器参数配置

硬件	参数
处理器	Intel(R) Xeon(R) Gold 5118 CPU@2.30 GHz
内存	DIMM,2 666 MHz,512 G
硬盘	东芝,2.4 TB,10 500 r/min
显存	NVIDIA Tesla V100-FHHL-16 GB

2.2 合成数据集结果分析

为了对本文所提方法进行验证,首先采用 CNN-LSTM-VO-cons 的训练模型作为初始权重、kitti_FE 数据集作为实验数据基础,将序列 00,02,05,08,09 作为训练集,因为这些场景的运行轨迹相对较长,包含了更丰富的数据用于模型的训练;其余的 6 个场景(序列 01,03,04,06,07,10)作为测试集;而验证集随机挑选自训练集,而且属于不放回抽样。具体数据如表 3 所示。其中,将场景 05,09 和 01,04 分别作为训练集和测试集中例子来直观分析现有方法在线性数据集 kitti 上的相对姿态估计轨迹和本文所用方法之间的相对姿态估计轨迹,如图 5 所示。

表 3 训练集、验证集及测试集组成

数据集	序列	总图像对
训练	00 02 05 08 09	17 625
验证		640
测试	01 03 04 06 07 10	5 576

实验设置初始学习率为 10^{-4},$\beta_1=\beta_2=10$,batchsize 为 40,为了使数据集可以随意截取长度进行单独训练,增加数据集的多样性,对于每个 batchsize,都将 LSTM 的初始状态设为 0,迭代了 100 个 epoch 的网络模型进行测试。实验结果显示在 epoch=50 时的测试效果最好,如图 5 所示,分别显示了测试集上序列 05、09 和测试集上序列 01、04 的轨迹图(实线为本文方法)。

在该实验中,使用本文方法和 DeepVO、CNN-LSTM-VO-cons 进行了对比,可以看出,本文所提方法可以直接应用于具有畸变鱼眼图像上,并且鱼眼图像更适用于相机姿态估计测量,证明了本文所提方法的可行性。因此为了实现实际场景下鱼眼相机相对姿态估计,将本文所用网络框架的训练模型参数作为实景测量的初始参数,然后进行 fine-tuning 来提高实景下测量的精度。

2.3 应用到真实鱼眼相机实验分析

为了验证本文所提出利用合成鱼眼图像 kitti_FE 训练的网络模型迁移到实景下鱼眼相机姿态估计方法的鲁棒性和精确性,使用得图 Dokicam 全景相机(它是由两个 190°FOV、2 048×2 048 分辨率的鱼眼镜头组成)集成在学校自主研制的自动驾驶小车上,在中北大学校园里采集了

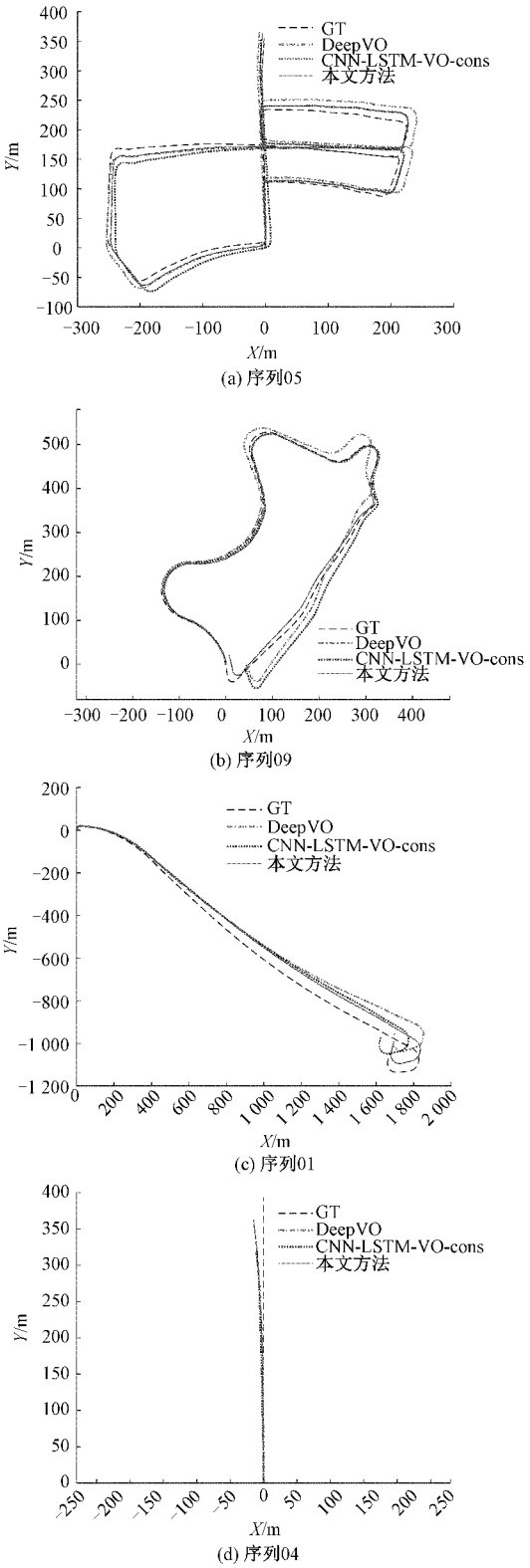
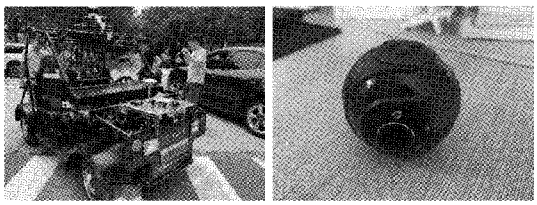


图 5 在序列 05、09 的训练结果和序列 01、04 上的测试结果

一组序列图片作为数据基础。实验过程中使用的设备如图 6 所示。



(a) 自动驾驶小车　　(b) Dokicam全景相机

图 6　自动驾驶小车和全景相机

由于合成图像与真实图像的分辨率和纵横比之间的差异，先前的训练对于真实的鱼眼镜头来说是不够的。因此本文将上一节训练好的模型参数作为初始权重，将前面的卷积层冻结，在第 5、6 层卷层进行 fine-tuning，然后再输入到 SLAM 层中重新训练时序特征，最后实现真实鱼眼相机姿态估计。

1)鱼眼相机位姿估计实验分析

为了对本文所提出算法在实际应用中的精度进行评估，在相同的场景下分别对本文方法和 DeepVO、CNN 和 CNN-LSTM-VO-cons 等主流开源框架计算出的估计位姿与真实位姿间的平均绝对误差和欧氏距离均方根误差(root mean square error，RMSE)为标准进行计算，RMSE 表达式如式(10)所示。

$$RMSE(s)=\sqrt{(\sum_{i=1}^{z}\| s_i^e-s_i^g\|^2)/z} \tag{10}$$

式中：s_i^e 表示第 i 帧的 3D 空间坐标估计值，s_i^g 表示第 i 帧的 3D 空间坐标真实值。实验测量数据结果如表 4 所示。

表 4　位姿估计误差统计

模型	Z/m	X/m	RMSE(3D)
CNN	2.685 6	3.022 5	4.126 5
DeepVO	2.5256	2.845 6	3.972 5
CNN-LSTM-VO-cons	2.414 7	2.635 3	3.785 8
本文方法	**1.936 2**	**2.046 5**	**2.827 6**

由表 4 可知，本文所提的宽视角相机相对姿态估计方法比现有开源框架的测量精度整体上有了明显地提高。

2)高速运动下位姿估计的性能稳定性分析

为了验证使用鱼眼相机的相对姿态估计方法在快速运动下的鲁棒性，通过减少数据集序列下的图像帧数来模拟快速运动，分别模拟了 2 倍速度到 5 倍速度下的跟踪。对采用迁移学习各方法的 RMSE 进行了记录，如表 5 所示。

表 5 中对比了本文算法与一些开源框架在不同运动速度下运行的结果，可以看出 CNN 和 DeepVO 在 4 倍运动速度下 RMSE 的上升速度开始明显增加，CNN-LSTM-VO-cons 和本文方法在 5 倍运动速度下 RMSE 也开始增加，但是本文所提方法的 RMSE 上升速度要比 CNN-LSTM-VO-cons 慢 38%，证明本文方法拥有更强的鲁棒性。且本文方法在相同场景下与现有框架 CNN、DeepVO 和 CNN-LSTM-VO-cons 进行了对比，相对于 CNN、DeepVO 和 CNN-LSTM-VO-cons 的方法分别提升了 32%、29%和 25%，证明了该方法的有效性。

表 5　高速位姿估计误差统计

模型	Seep (×2)	Seep (×3)	Seep (×4)	Seep (×5)
CNN	4.162 3	4.232 3	4.675 2	5.011 6
DeepVO	4.030 0	4.162 5	4.534 2	4.876 4
CNN-LSTM-VO-cons	3.763 5	3.814 5	3.876 2	4.063 2
本文方法	**2.860 2**	**2.924 7**	**2.976 0**	**3.091 4**

3　结　　论

针对现有普通镜头存在视场局限性的问题，本文提出了一种使用鱼眼镜头作为视觉传感器的宽视角图像相对姿态测量方法。结合深度学习和鱼眼相机的非线性光学特性直接对鱼眼图像进行相对姿态估计，使其能够在自动驾驶中更好的解析场景和实现自身位置测量。在实现过程中发现由于样本少，无法进行实际鱼眼相机的姿态测量，因此本文提出使用合成鱼眼数据集结合 CNN+LSTM 双向循环网络进行迁移学习的姿态估计方法，取得了较好的测量效果。本文方法虽然在一定程度上优于现有方法，但是还存在不足。在之后的研究中考虑加入位置和感兴趣区域限定条件来进一步提高姿态估计精度。为之后实现宽视角范围目标的位置与距离同时测量奠定基础，使无人驾驶更具有安全性。

参考文献

[1] HEIMBERGER M，HORGAN J，HUGHES C，et al. Computervision in automated parking systems：Design，implementation and challenges[J]. Image and Vision Computing，2017，68(12)：88-101.

[2] DAHAL A，HOSSEN J，SUMANTH C，et al. DeepTrailerAssist：Deep learning based trailer detection，tracking and articulation angle estimation on automotive rear-view camera[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop(ICCVW)，IEEE，2019：2339-2346.

[3] HORGAN J，HUGHES C，MCDONALD J，et al. Vision-based driver assistance systems：Survey，taxonomy and advances[C]. 2015 IEEE 18th International Conference on Intelligent Transportation Systems，IEEE，2015：2032-2039.

[4] DRULEA M，SZAKATS I，VATAVU A，et al. Omnidirectional stereo vision using fisheye lenses[C].

2014 IEEE 10th International Conference on IntelligentComputer Communication and Processing (ICCP), IEEE, 2014: 251-258.

[5] CARUSO D, ENGEL J, CREMERS D. Large-scale direct slam for omnidirectional cameras[C]. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), IEEE, 2015: 141-148.

[6] 季晓明,文怀海.基于非线性终端滑模的码垛机械臂轨迹跟踪控制[J].电子测量与仪器学报,2021,35(9):105-111.

[7] 林立雄,郑佳春,黄国辉,等.基于卷积神经网络与扩展卡尔曼滤波的单目视觉惯性里程计[J].仪器仪表学报,2021,42(10):188-198.

[8] 宋玉琴,熊高强,曾贺东,等.多平面点优化的单目SLAM方法[J].国外电子测量技术,2021,40(10):40-45.

[9] 张裕,徐熙平,张宁,等.基于折反射全景相机的视觉里程计研究[J].光子学报,2021,50(4):190-198.

[10] 李海滨,褚光宇,张强,等.基于优化的鱼眼镜头成像模型的空间点定位[J].光学学报,2015,35(7):247-253.

[11] 王永仲.鱼眼镜头光学[M].北京:科学出版社,2006.

[12] 牛泽,李成静,白慧敏,等.鱼眼图像的目标物体角点检测方法[J].电子测量技术,2020,43(1):147-151.

[13] KHOMUTENKO B, GARCIA G, MARTINET P. An enhanced unified camera model[J]. IEEE Robotics and Automation Letters, 2015, 1(1): 137-144.

[14] BLOTT G, TAKAMI M, HEIPKE C. Semantic segmentation of fisheye images[C]. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018: 181-196.

[15] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]. 2012 IEEE conference on computer vision and pattern recognition, IEEE, 2012: 3354-3361.

[16] 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), IEEE, 2015: 2938-2946.

[17] WALCH F, HAZIRBAS C, LEAL-TAIXÉ L, et al. Image-based localization with spatial LSTMs[J]. ArXiv Preprint, 2016, ArXiv: 1611.07890.

[18] CLARK R, WANG S, MARKHAM A, et al. VidLoc: 6-DoF video-clip relocalization[J]. ArXiv Preprint, 2017, ArXiv: 1702.06521.

[19] WANG S, CLARK R, WEN H, et al. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]. 2017 IEEE International Conference on Robotics and Automation(ICRA), IEEE, 2017: 2043-2050.

[20] 张林箭. 基于深度学习的相机相对姿态估计[D].杭州:浙江大学,2018.

作者简介

李成静,博士研究生,主要研究方向为计算机视觉。
E-mail:Xianjing_Li@163.com

王鉴,副教授,主要研究方向为信号与信息处理。
E-mail:9036944 @qq.com

郭锦铭,硕士研究生,主要研究方向为深度学习。
E-mail:1606014624@qq.com

张璇,博士研究生,主要研究方向为光谱测温。
E-mail:1094398770@qq.com

韩焱(通信作者),教授,博士生导师,主要研究方向为通信技术、信号处理和识别,数字图像处理与信息重建等。
E-mail:hanyan@nuc.edu.cn