JIG
JOURNAL OF IMAGE AND GRAPHICS
主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
09
VOL.25
ISSN1006-8961
CN11-3758/TB

遥感边缘智能技术 *P1719*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第9期（总第293期）

2020年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

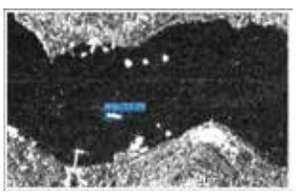
All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

2020年9月 第25卷 第9期（总第293期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

第一人称视角下的社会力优化多行人跟踪(第1869页)

分形曲线生成的频域方法(第1904页)



双向特征融合的数据自适应SAR图像舰船目标检测模型(第1943页)

学者观点

遥感边缘智能技术研究进展及挑战
孙显，梁伟，刁文辉，曹志颖，冯瑛超，王冰，付琨 …… 1719

综述

红外弱小目标检测算法综述
李俊宏，张萍，王晓玮，黄世泽 …… 1739

目标检测中的尺度变换应用综述
申奉璨，张萍，罗金，刘松阳，冯世杰 …… 1754

图像处理和编码

分层特征融合注意力网络图像超分辨率重建
雷鹏程，刘丛，唐坚刚，彭敦陆 …… 1773

适用小样本的无参考水下视频质量评价方法
宋巍，刘诗梦，黄冬梅，王文娟，王建 …… 1787

图像分析和识别

结合面料属性和触觉感测的织物识别
邢寅初，刘骊，付晓东，刘利军，黄青松 …… 1800

结合GAN的轻量级模糊车牌识别算法
段宾，符祥，江毅，曾接贤 …… 1813

非贪婪的鲁棒性度量学习算法
曾凡霞，张文生 …… 1825

不规则像素簇显著性检测算法
李明旭，翟东海 …… 1837

全局区域相异度阈值驱动构建的稀疏尺度集模型
王浩宇，沈占锋，柯映明，许泽宇，李硕，焦淑慧 …… 1848

图像理解和计算机视觉

分裂合并运动分割的多运动视觉里程计方法
王晨捷，张云，赵青，王伟，尹露，罗斌，张良培 …… 1859

第一人称视角下的社会力优化多行人跟踪
杨廷召，刘骊，付晓东，刘利军，黄青松 …… 1869

引入辅助损失的多场景车道线检测
陈立潮，徐秀芝，曹建芳，潘理虎 …… 1882

计算机图形学

旋转骨架法在二值图像分形维数计算中的应用
纪佑军，何杰，韩海水，程忠钊，曾保全 …… 1894

分形曲线生成的频域方法
陈伟，乔洁雯，周晨 …… 1904

医学图像处理

融合残差注意力机制的UNet视盘分割
侯向丹，赵一浩，刘洪普，郭鸿湧，于习欣，丁梦园 …… 1915

深层聚合残差密集网络的超声图像左心室分割
吴宣言，缑新科，朱子重，魏域林，王凯 …… 1930

遥感图像处理

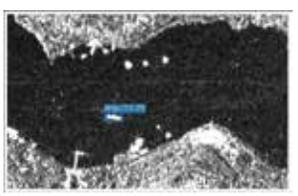
双向特征融合的数据自适应SAR图像舰船目标检测模型
张筱晗，姚力波，吕亚飞，简涛，赵志伟，藏洁 …… 1943

Volume 25, Number 9
Published September 16, 2020

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Multi-pedestrian tracking optimized by social force model under first-person perspective(P1869)

Fractal curve generation method based on the frequency domain (P1904)



Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images (P1943)

Scholar View
Progress and challenges of remote sensing edge intelligence technology
Sun Xian, Liang Wei, Diao Wenhui, Cao Zhiying, Feng Yingchao, Wang Bing, Fu Kun ····· 1719

Review
Infrared small-target detection algorithms: a survey
Li Junhong, Zhang Ping, Wang Xiaowei, Huang Shize ····· 1739
Scale changing in general object detection: a survey
Shen Fengcan, Zhang Ping, Luo Jin, Liu Songyang, Feng Shijie ····· 1754

Image Processing and Coding
Hierarchical feature fusion attention network for image super-resolution reconstruction
Lei Pengcheng, Liu Cong, Tang Jiangang, Peng Dunlu ····· 1773
Non-reference underwater video quality assessment method for small size samples
Song Wei, Liu Shimeng, Huang Dongmei, Wang Wenjuan, Wang Jian ····· 1787

Image Analysis and Recognition
Cloth recognition based on fabric properties and tactile sensing
Xing Yinchu, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong ····· 1800
Lightweight blurred car plate recognition method combined with generated images
Duan Bin, Fu Xiang, Jiang Yi, Zeng Jiexian ····· 1813
Robust distance metric learning algorithm based on nongreedy solution
Zeng Fanxia, Zhang Wensheng ····· 1825
Significance detection method with irregular pixel clusters
Li Mingxu, Zhai Donghai ····· 1837
Sparse scale set model based on global regional dissimilarity threshold
Wang Haoyu, Shen Zhanfeng, Ke Yingming, Xu Zeyu, Li Shuo, Jiao Shuhui ····· 1848

Image Understanding and Computer Vision
Multi-motion visual odometry based on split-merged motion segmentation
Wang Chenjie, Zhang Yun, Zhao Qing, Wang Wei, Yin Lu, Luo Bin, Zhang Liangpei ····· 1859
Multi-pedestrian tracking optimized by social force model under first-person perspective
Yang Tingzhao, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong ····· 1869
Multi-scenario lane line detection with auxiliary loss
Chen Lichao, Xu Xiuzhi, Cao Jianfang, Pan Lihu ····· 1882

Computer Graphics
Application of skeleton rotating method in fractal dimension calculate of binary image
Ji Youjun, He Jie, Han Haishui, Cheng Zhongzhao, Zeng Baoquan ····· 1894
Fractal curve generation method based on the frequency domain
Chen Wei, Qiao Jiewen, Zhou Chen ····· 1904

Medical Image Processing
Optic disk segmentation by combining UNet and residual attention mechanism
Hou Xiangdan, Zhao Yihao, Liu Hongpu, Guo Hongyong, Yu Xixin, Ding Mengyuan ····· 1915
Left ventricular segmentation on ultrasound images using deep layer aggregation for residual dense networks
Wu Xuanyan, Gou Xinke, Zhu Zizhong, Wei Yulin, Wang Kai ····· 1930

Remote Sensing Image Processing
Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images
Zhang Xiaohan, Yao Libo, Lyu Yafei, Jian Tao, Zhao Zhiwei, Zang Jie ····· 1943

©中国图象图形学报版权所有

中图法分类号:P227　　**文献标识码**:A　　**文章编号**:1006-8961(2020)09-1859-10

论文引用格式:Wang C J, Zhang Y, Zhao Q, Wang W, Yin L, Luo B and Zhang L P. 2020. Multi-motion visual odometry based on split-merged motion segmentation. Journal of Image and Graphics,25(09):1859-1868(王晨捷,张云,赵青,王伟,尹露,罗斌,张良培. 2020. 分裂合并运动分割的多运动视觉里程计方法. 中国图象图形学报,25(09):1859-1868)[DOI:10.11834/jig.200132]

分裂合并运动分割的多运动视觉里程计方法

王晨捷,张云,赵青,王伟,尹露,罗斌,张良培

武汉大学测绘遥感信息工程国家重点实验室, 武汉　430079

摘　要:**目的**　视觉里程计(visual odometry,VO)仅需要普通相机即可实现精度可观的自主定位,已经成为计算机视觉和机器人领域的研究热点,但是当前研究及应用大多基于场景为静态的假设,即场景中只有相机运动这一个运动模型,无法处理多个运动模型,因此本文提出一种基于分裂合并运动分割的多运动视觉里程计方法,获得场景中除相机运动外多个运动目标的运动状态。**方法**　基于传统的视觉里程计框架,引入多模型拟合的方法分割出动态场景中的多个运动模型,采用 RANSAC(random sample consensus)方法估计出多个运动模型的运动参数实例;接着将相机运动信息以及各个运动目标的运动信息转换到统一的坐标系中,获得相机的视觉里程计结果,以及场景中各个运动目标对应各个时刻的位姿信息;最后采用局部窗口光束法平差直接对相机的姿态以及计算出来的相机相对于各个运动目标的姿态进行校正,利用相机运动模型的内点和各个时刻获得的相机相对于运动目标的运动参数,对多个运动模型的轨迹进行优化。**结果**　本文所构建的连续帧运动分割方法能够达到较好的分割结果,具有较好的鲁棒性,连续帧的分割精度均能达到近100%,充分保证后续估计各个运动模型参数的准确性。本文方法不仅能够有效估计出相机的位姿,还能估计出场景中存在的显著移动目标的位姿,在各个分段路径中相机自定位与移动目标的定位结果位置平均误差均小于6%。**结论**　本文方法能够同时分割出动态场景中的相机自身运动模型和不同运动的动态物体运动模型,进而同时估计出相机和各个动态物体的绝对运动轨迹,构建出多运动视觉里程计过程。

关键词:多运动视觉里程计;多模型拟合;运动分割;量化残差偏好;交替采样与聚类

Multi-motion visual odometry based on split-merged motion segmentation

Wang Chenjie, Zhang Yun, Zhao Qing, Wang Wei, Yin Lu, Luo Bin, Zhang Liangpei

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: **Objective** With the continuous development and popularization of robotics and autonomous driving technology, the demand for high-precision localization and navigation in dynamic scenarios continues to increase. Visual localization only requires a common camera to achieve a localization function with considerable accuracy, and the obtained video data can be used for 3D scene reconstruction, scene analysis, target recognition, target tracking, and other tasks. Among these tasks, visual odometry (VO) has become a hotspot in autonomous localization research and has been widely applied in the localization and navigation of robots and unmanned vehicles. VO can estimate the camera pose relative to a static background. However, the current research and application of VO are based on static scenes where most objects are stationary. When multiple moving targets are present in a scene, the camera ego-motion generates a large error. Therefore, eliminating

收稿日期:2020-04-18;**修回日期**:2020-06-14;**预印本日期**:2020-06-21

基金项目:中央高校基本科研业务费专项资金项目

Supported by:Fundamental Research Funds for the Central Universities

the interference of moving targets in a scene (even though they occupy most of the field of view), accurately calculating the camera pose, and estimating the motion model of each moving object are practical problems that need to be solved in moving target trajectory estimation and modeling analysis. **Method** This paper proposes a multi-motion VO based on split and merged motion segmentation that applies the general method for estimating the motion model parameters based on traditional VO. We also apply the multi-model fitting method on multiple data in the motion estimation process. The motion model is estimated to fit multiple motion model parameter instances. Afterward, multiple motion models are mapped in time series to complete a continuous frame motion segmentation, and the absolute pose of each moving target at the current time is obtained. Local bundle adjustment is then applied to directly correct the camera and absolute poses of each moving target and to complete the multi-motion VO process. The main contents and innovations of this article are summarized as follows: 1) the motion segmentation method based on multi-model fitting is applied to the traditional VO framework, and a multi-motion VO framework based on multi-model fitting motion segmentation is proposed. In a dynamic scene with multiple moving rigid body targets, the trajectory of multiple moving objects and the ego-motion of the camera are simultaneously estimated. 2) This paper combines multi-model fitting with VO. The preference analysis method of quantized residuals is also combined with alternating sampling and clustering strategies to improve the performance of the existing multi-model fitting method in segmenting the motion and dynamic object motion models of the camera in dynamic scenes. 3) In this paper, the motion segmentation strategy is optimized through motion segmentation to achieve a continuous frame motion segmentation and to obtain multi-motion model parameters. Furthermore, the absolute pose of multi-motion targets (including camera motion) in the same coordinate system can be obtained to realize a complete multi-motion VO. First, the oriented feature from aclelerated segments test (FAST) and rotated binary robust independent elementary features (BRIEF) (ORB) method was used to extract the feature points of the stereo images of the current and previous frames, and then a stereo matching of the left and right images of these frames was performed by matching the feature points of these images and obtaining the associated 3D information in the current and previous frames. Second, the preference analysis method of quantized residuals was combined with alternating sampling and clustering strategies to improve the existing multi-model fitting method, and an inlier segmentation of the current frame multi-motion model was performed to achieve a single-step motion segmentation. Third, a continuous frame motion segmentation was performed based on the results of the multi-motion segmentation at the previous moment. Fourth, based on the ego-motion estimation results obtained by a camera in each moment and the estimation results for other moving targets in a scene, the inliers of multiple target movement models in a scene at each moment were obtained as a time series. Fifth, by using the inliers of each motion model obtained via motion segmentation, random sample consonsus (RANSAC) robustness was used to estimate the motion parameters of each model, and the motion results of the camera relative to each motion target were estimated. Sixth, the estimation result was converted into a unified global coordinate to determine the absolute pose of each moving target at the current time. Finally, local bundle adjustment was used to directly correct the camera pose and the absolute pose of each moving target in each moment. The inliers of the camera motion model and the motion parameters of multiple motion models across various periods were used to optimize the trajectories of multiple moving targets. **Result** Compared with the existing methods, the proposed continuous frame motion segmentation method can achieve better segmentation results, show higher robustness and continuous frame segmentation accuracy, and guarantee an accurate estimation of each motion model parameter. The proposed multi-motion VO method not only effectively estimates the pose of the camera but also that of a moving target in a scene. The results for the self-localization of the camera and the localization of the moving target show small errors. **Conclusion** The proposed multi-motion VO method based on the split-merged motion segmentation method can simultaneously segment the motion model of the camera in dynamic scenes and the moving object motion models. The absolute motion trajectories of the camera and various moving objects can also be estimated simultaneously to build a multi-motion VO process.

Key words: multi-motion visual odometry; multi-model fitting; motion segmentation; quantization of residual preferences; alternating sampling and clustering

0 引 言

随着机器人和自动驾驶技术的不断发展和应用普及，动态场景下的高精度定位导航需求不断提升。视觉定位只需要普通相机即可实现精度可观的定位功能，并且获得的视频数据能够用于场景的 3 维重建、场景分析、目标识别和目标跟踪等任务，其中，视觉里程计（visual odometry，VO）已经成为自主定位的研究热点，并且广泛应用于机器人及无人车的定位导航之中。视觉里程计可以估计出摄像机相对于静态背景的运动（即摄像机的自我运动），例如 SVO（semi-direct visual odometry）方法（Forster 等，2014）。但是当前研究及应用基本是基于场景为静态的假设，即场景中所包含的大部分物体均为静止目标，然而当场景中存在多个运动目标时，相机自定位会存在较大的误差。因此针对场景中存在多个运动目标时，如何排除其他运动目标的干扰（即使运动目标占据绝大部分的视场），实现相机自身准确的姿态解算，同时进一步估计出各个运动目标的运动模型，并对各个运动目标进行轨迹估计及建模分析，这是一项亟待解决的研究和实际应用问题。

针对多运动视觉里程计的研究还较少，更多的是针对运动分割的研究，虽然关于多目标运动估计的相关研究已逐步开展（Sabzevari 和 Scaramuzza，2016），但还是在较短视频序列的运动分割基础之上实现的多运动模型参数估计。Zappella 等人（2008）将目前主要的运动分割分成 6 类：图像差分方法（Li 等，2007）、基于统计的方法（Shen 等，2007）、基于光流的方法（Zhang 等，2007；Wiskott，1997）、小波变换的方法（Wiskott，1997）、基于分层的方法（Kumar 等，2008）和基于因式分解的方法（Costeira 和 Kanade，1998）。然而以上运动分割算法并不是对每相邻两帧图像进行运动分割，而是针对一小段视频序列的跟踪点进行处理。所以现有的基于视频序列的针对跟踪的特征点进行运动分割的算法难以实际用于视觉里程计中多运动模型的分割和估计。Sabzevari 和 Scaramuzza（2016）提出的方法虽然能够实现多目标运动估计，但是采用的运动模型较为简单，采用的实验数据与运动分割方法相同，难以扩展到实际的视觉里程计之中，实现多运动目标的视觉里程计。基于此，与 Sabzevari 和 Scaramuzza（2016）提出方法的思路不同，本文不采用传统运动分割方法的思路，而是参照视觉里程计中利用随机采样一致性（random sample consensus，RANSAC）算法用于鲁棒性运动参数估计的方法，采用 RANSAC 这种单模型拟合方法对应的多模型拟合方式，对相邻两帧图像进行运动分割以及多运动模型参数估计，在实现多运动模型参数估计的基础之上，获得多运动视觉里程计。

本文提出一种基于分裂合并运动分割的多运动视觉里程计，借鉴传统视觉里程计采用 RANSAC 进行运动模型参数估计的一般方法，引入多模型拟合方法在运动估计过程中对数据中存在的多个运动模型进行估计，拟合出多个运动模型的参数实例；接着在时间序列上将多个运动模型对应，完成连续帧运动分割，进而获得当前时刻各个运动目标的绝对姿态；最后采用局部窗口光束法平差直接对相机的姿态以及各个运动目标的绝对姿态进行校正，完成多运动视觉里程计过程。主要的研究内容以及创新之处如下：

1）将基于多模型拟合的运动分割方法应用于传统的视觉里程计框架中，提出了基于多模型拟合运动分割的多运动视觉里程计框架，在多个运动刚体目标的动态场景中，估计出相机自运动的同时，估计出场景中多个运动物体的轨迹。

2）从多模型拟合方法以及视觉里程计方法两方面结合，将量化残差的偏好分析方法结合交替采样与聚类策略，改进现有多模型拟合方法，分割出场景中的相机自身运动模型和不同运动的动态物体运动模型。

3）通过运动分割关联情况进行运动分割策略的优选，实现连续帧运动分割，获得多运动模型参数，进而获得同一坐标系下包括相机运动在内的多运动目标的绝对位姿，实现完整的多运动视觉里程计。

1 系统概述

提出一种基于分裂合并运动分割的多运动视觉里程计方法，如图 1 所示。首先，采用 ORB（oriented FAST（featrues-from accelerated segmonts test）and rotated BRIEF（binary robust independent elementary features））方法（Rublee 等，2011）分别提取出前后帧

立体图像的特征点，然后分别对前后帧左右图进行立体匹配，通过前后帧左图的特征点匹配获得前后帧3维点的关联信息；之后将量化残差的偏好分析方法结合交替采样与聚类策略，来改进现有多模型拟合方法，进行当前帧多运动模型内点分割，实现单步运动分割，接着根据前一时刻的多运动分割结果，与前一时刻估计出来的多个运动模型进行对应，进行连续帧运动分割，对应关联到各个时刻相机的自运动估计结果以及场景中其他运动目标的估计结果，获得场景中多个目标运动模型在时序序列上各时刻的模型内点。利用运动分割得到的各个运动模型内点，采用RANSAC鲁棒性估计出各个模型的运动参数，进而估计出相机相对于各个运动目标的运动结果。接着对估计结果进行坐标系转换，将其转换到统一的全局坐标空间中，获得当前时刻各个运动目标的绝对姿态。最后采用局部窗口光束法平差直接对相机的姿态以及各个运动目标各个时刻的绝对姿态进行校正，利用相机运动模型的内点和各个时刻多个运动模型的运动参数，对多个运动目标的轨迹进行优化。

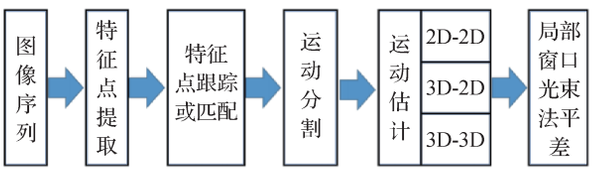


图1　多运动视觉里程计流程图

Fig. 1　Multi-motion visual odometry flow chart

2　基于多模型拟合的运动分割方法

本文将量化残差的偏好分析方法结合交替采样与聚类策略(Zhang等,2018)，改进现有多模型拟合方法，实现场景中多运动模型的分割。采用的多模型拟合方法大致包括2个关键步骤：基于量化残差偏好表示的外点去除和交替采样与聚类策略的多模型内点分割，其方法流程如图2所示。

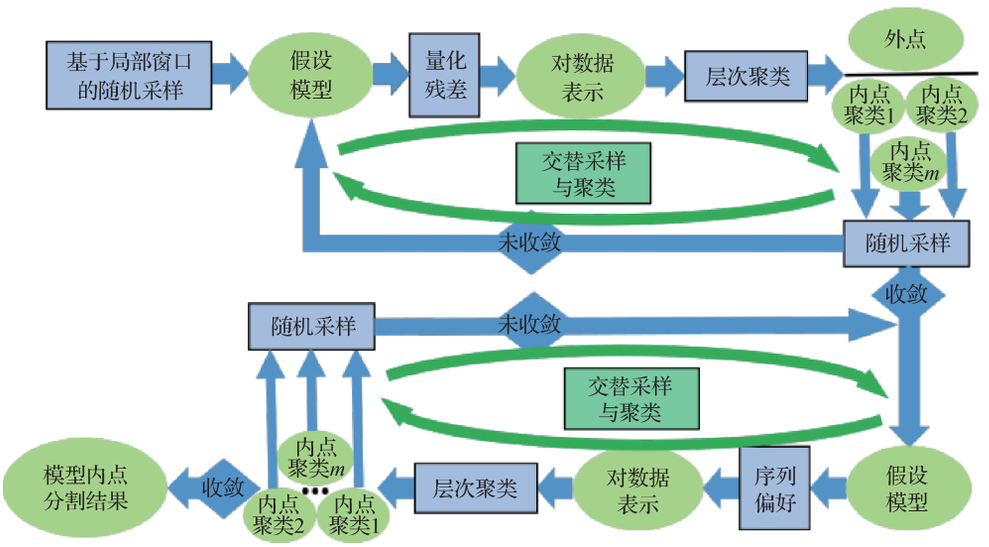


图2　基于多模型拟合的运动分割方法

Fig. 2　Motion segmentation method based on multi-model fitting

在基于量化残差偏好表示的外点去除中，首先进行基于局部窗口的随机采样，获得假设模型，然后计算假设模型对所有数据点的残差值，构成残差矩阵，利用量化残差的数据点表示方法计算出数据点的量化残差偏好表示，通过量化残差方法对残差进行量化，可以减小属于同一模型内点偏好的差异，增大不同模型内点偏好之间的差异，从而改善偏好。接下来利用量化残差的偏好表示进行典型的层次聚类，聚类之后获得多个模型，按照式1)—4)计算各个类别与外点聚类接近程度的 ξ_{Ci} 值，选择 ξ_{Ci} 值最大的那一组聚类结果设定为外点聚类。接下来就是循环到采样过程，在排除掉外点聚类的剩余其他聚类类别中分别进行随机采样，获得假设模型，以此尽量排除掉外点的干扰，获得假设模型之后计算残差矩阵，进而计算数据点的量化残差偏好表示，应用层次聚类获得聚类类别，计算 ξ_{Ci} 值来判别外点聚类。依

次不断往复运行，直至外点聚类区域收敛为止

$$\xi_i = \frac{\sum_{j=1}^{M} \hat{q}_{i,j}}{M} \tag{1}$$

$$\hat{q}_{i,j} = \begin{cases} q_{i,j} & q_{i,j} \neq 0 \\ \lambda & q_{i,j} = 0 \end{cases} \tag{2}$$

式中，$q_{i,j} \in \boldsymbol{q}_i$，$\boldsymbol{q}_i$ 为数据点 x_i 的偏好表示，M 表示生成的假设模型个数，λ 表示量化残差的等级截断阈值。得到的 ξ_i 值越大，表示数据点 x_i 更有可能是外点。

$$\xi_{C^t} = \frac{\sum_{h=1}^{T} \xi_h}{T},\ x_h \in \boldsymbol{C}^t \tag{3}$$

$$T = card\{\boldsymbol{C}^t\} \tag{4}$$

式中，$\boldsymbol{C}^t$ 表示第 t 个聚类，$card$ 表示该聚类内点的个数，接着在剔除掉外点之后的较为纯净的数据点中进行各个模型内点的同时聚类，即进行交替采样与聚类策略的多模型内点分割，由于经过了外点剔除，数据中绝大部分的外点都被成功剔除，剩余的数据中仅包含各个模型的内点。因此通过初始的局部窗口随机采样方法获得假设模型后，采用假设模型计算数据点的残差矩阵，经过序列偏好计算出剩余各模型内点的序列偏好表示，采用层次聚类获得各个聚类类别，由于其中不包含外点的聚类，所有聚类获得的类别均将参与到随机采样之中，在每个类别上分别进行随机采样，获得新的假设模型用于计算序列偏好进行新一轮的层次聚类。依次交替循环运行采样与聚类过程，直至聚类生成的各个类别不变，即为聚类结果收敛，输出得到的聚类类别即为模型内点的分割结果，用各个模型聚类类别来进行最小二乘的模型参数估计，即可获得最终的多模型参数拟合结果。

3 单步运动分割

在单步运动分割中，对当前帧采用基于多模型拟合的运动分割方法对数据点进行聚类，获得属于不同模型内点的聚类结果，由于匹配精度和误匹配点带来的影响，属于同一运动模型的内点分割结果往往会掺杂较少量其他模型内点。因此本文利用模型内点的相邻特性，即属于同一运动模型内点会落在相邻区域，通过 K 近邻投票的方法对分割结果进一步优化，去掉掺杂其间的少量伪外点。K 近邻投票方法主要是借鉴 K 近邻（K-nearest neighbor，KNN）分类算法的原理，即数据点 x_i 在图像空间中 K 个最相邻的点中，属于模型 m 的内点集合 $\boldsymbol{C}_m$ 的点的个数大于 $K/2$，则数据点 x_i 是模型 m 的内点，即 $x_i \in \boldsymbol{C}_m$；否则数据点 x_i 所属模型不变。分别对每一类模型内点均进行 K 近邻投票，更正每类内点中包含的伪外点。循环运行以上步骤，即每次对各个模型内点进行 K 近邻投票，优化内点分割结果，直至最终整个内点分割结果稳定不变，获得较为准确的内点分割结果，即可实现单步运动分割，过程如图 3 所示。

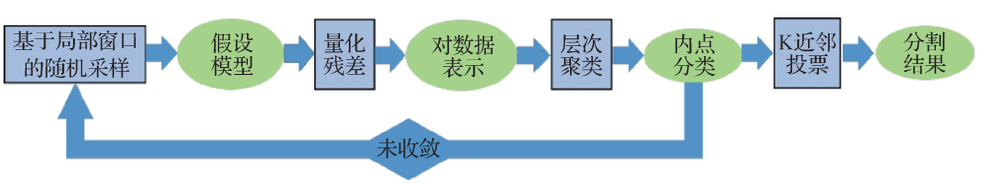


图 3　单步运动分割方法

Fig. 3　One-step motion segmentation method

4 连续帧运动分割

将单步运动分割方法用于连续帧运动分割，考虑前一时刻的场景内运动分割的结果，来实施合适的运动分割策略，实现连续帧运动分割方法。在连续帧中（如图 4 所示），首先将前后帧图像特征点进行关联，当前时刻 T_k 经过运动分割之后可以获得多个运动模型的内点分割结果，在计算下一时刻相机姿态时，将当前时刻拍摄的图像与下一时刻 T_{k+1} 进行匹配，获得特征点关联结果，根据特征关联结果，即可找到当前时刻对应的各模型中内点与下一时刻数据点的关联情况，进而根据当前时刻各运动模型内点的分割结果，对应地将下一时刻关联到的数据点进行归类分割，即可获得初步的运动分割结果。

接着采用 K 近邻投票的方法来获取未被确定

类别归属的新匹配点的分割类别，此时的判定条件更改为：当前点在图像空间中 K 个最近邻点中属于运动模型 m 的内点集合的点的个数最多，则当前点判定为运动模型 m 的内点。而当运动分割关联的内点个数较少或者没有关联到时，则判定运动模型跟踪丢失，直接进行单步的运动分割。经过以上连续帧运动分割处理后，可以得到各个时刻场景中匹配点按照运动模型的不同而分割成的内点集合。

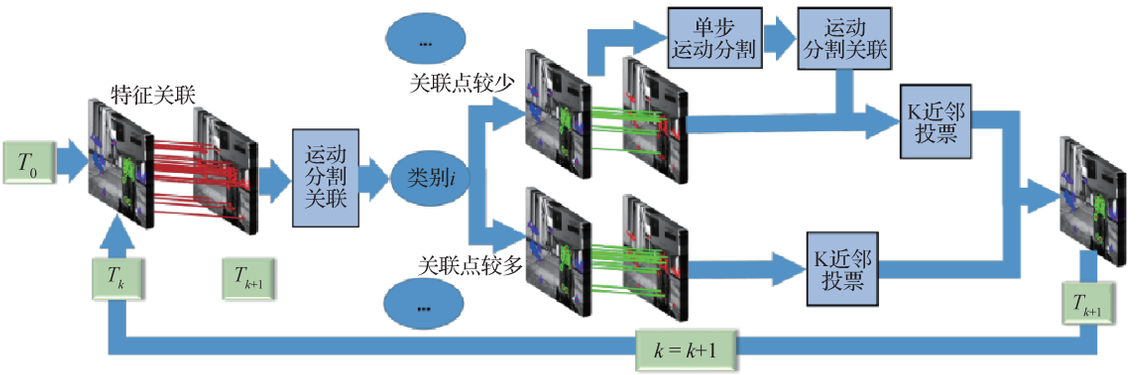


图 4　连续帧运动分割方法

Fig. 4　Continuous frame motion segmentation method

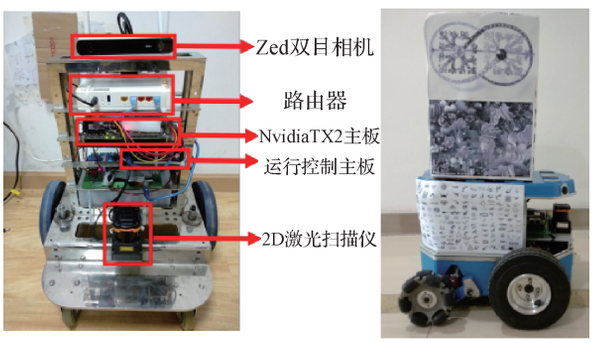
5　统一坐标系变换

经过多运动分割求解得到的运动模型参数均是相机相对于各个运动目标的运动参数，需要转换到统一的坐标系空间之中，才能估计各个运动目标绝对位姿。本文将初始时刻的相机空间坐标系作为统一的空间坐标系，通过变换得出相机当前的姿态以及各运动目标当前的姿态信息；最后通过局部窗口光束法平差，采用局部时刻窗口内相机各个时刻的运动模型内点以及各时刻计算出的相机姿态，构造局部窗口重投影误差方程，来对相机的各时刻姿态进行优化，同时进一步校正各个运动目标的运动估计结果。之后需要将计算出来的相机相对于运动目标的位姿转换成运动目标相对于初始时刻的位姿。假设获得的 T_k 时刻相机相对于运动目标的位姿为 $\boldsymbol{P}_k^{C\text{-}O}$，相机相对于初始时刻的位姿为 $\boldsymbol{P}_k^{C\text{-}T_0}$，则运动目标相对于初始时刻相机的位姿为

$$\boldsymbol{P}_k^{O\text{-}T_0} = (\boldsymbol{P}_k^{C\text{-}O})^{-1} \times \boldsymbol{P}_k^{C\text{-}T_0} \tag{5}$$

6　实　验

本文采用多运动立体视觉里程计的实验来评估方法的有效性和精度，实验所采用的相机移动平台和移动目标如图 5 所示，过程中始终保持有一个移动物体在相机的视野中，也就是场景中至多有两个运动模型。相机平台所搭载的设备如图 5(a)所示，包括采集双目图像数据的 Zed 双目相机、2D 激光扫描仪和惯性测量单元(inertial measurement unit, IMU)等。移动目标是如图 5(b)所示的移动机器人，两个运行平台所采用的底层定位导航包保持一致。



(a) 相机移动平台　　(b) 移动目标

图 5　相机移动平台和移动目标

Fig. 5　Moving platform with camera and moving target

((a) moving platform with camera; (b) moving target)

实验开始之前，利用 gmapping(Grisetti 等，2007)算法融合激光雷达、轮式里程计以及 IMU 等多传感器，构建室内场景的 2D 占据栅格地图(occupancy grid map)(岳伟韬 等，2020)，结果如图 6 所示。在实验过程中，相机平台和移动目标均利用事先生成的占据栅格地图和传感器数据进行有图环境下的即时定位，将二者各个时刻的位姿分别记录，作为真值数据，用来评估多运动视觉里程计的定位精度。所采用的 2D 激光扫描仪为北阳公司 UTM-

图 6　实验场景的 2 维占据栅格地图

Fig. 6　A 2D occupancy grid map of the experimental scene

30LX 2D 激光扫描测距产品，具有 30 m，270°测量范围，25 ms 扫描时间间隔，通过 gmapping 算法融合 IMU 与轮式里程计，可实现 5 cm 的制图精度，同时利用事先构建的地图进行定位的精度在 5 cm 以内，可作为真值数据使用，进行多运动视觉里程计定位精度的评估。

6.1　运动分割实验

实验一共拍摄 3 段双目立体视频数据，对应的视频数据编号代表该数据，其中视频 1 拍摄过程中相机平台保持静止不动，而场景中的运动目标存在运动；视频2和视频3在拍摄过程中，相机平台存在运动，并且此时场景中的移动目标也会运动。各段数据多运动分割精度（大部分内点分割正确的帧数占总视频帧数的百分比）如表 1 所示，以上 3 段视频的连续帧分割结果均能将视频中的所有帧图像中的运动模型正确分割出来，连续帧的分割精度均能达到 100%，视频中的运动目标上的大部分内点均能被正确提取。连续帧分割结果如图 7 所示，绿色圆圈标记出来的点为静止背景区域代表的运动模型内点分割结果，即用来估计相机自运动模型参数的数据点；红色圆圈标记出来的点则为场景中移动目标运动模型内点分割结果，也即用来估计移动目标运动模型参数的数据点。

表 1　连续帧运动分割正确率

Table 1　Continuous frame motion segmentation accuracy

数据编号	图像帧数	正确分割帧数	连续帧分割正确率/%
1	180	180	100
2	280	280	100
3	536	536	100

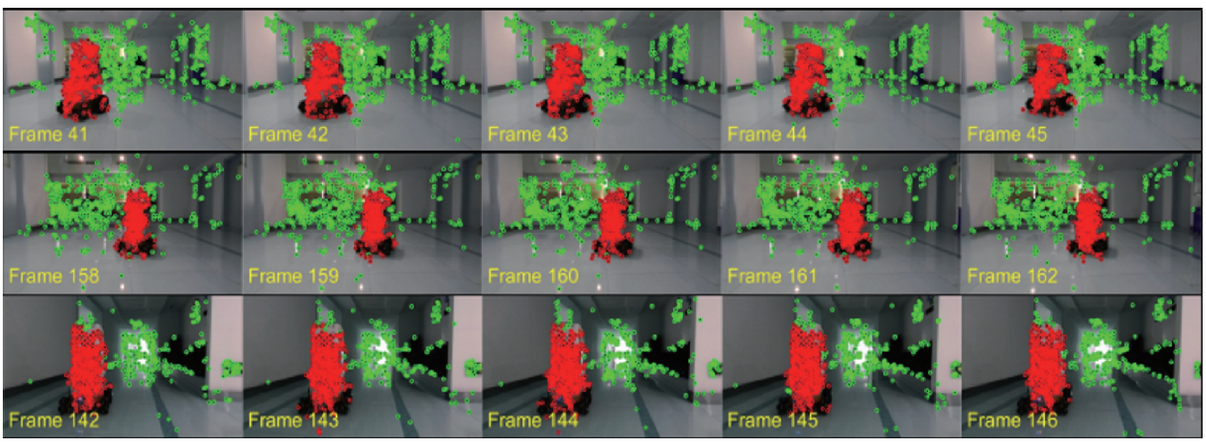


图 7　连续帧运动分割结果示例图

Fig. 7　Example diagram of continuous frame motion segmentation results

从实验展示的 3 组视频数据的运动分割结果可以明显看出，针对多运动视觉里程计中所需要的运动分割处理效果，本文所构建的连续帧运动分割方法能够达到较好的分割结果，具有较好的鲁棒性和较高的连续帧分割精度，充分保证后续估计各个运动模型参数的准确性。

6.2　多运动视觉里程计实验

多运动视觉里程计相较于传统的视觉里程计而言，主要是针对场景中存在多个运动目标时，能够将各个运动目标的运动模型参数估计出来，获得各个运动目标的位置与姿态。图 8 是本文的多运动视觉里程计估计出相机轨迹以及场景中的运动目标轨迹，与 gmapping 算法获得的真值之间的对比。

本文多运动视觉里程计精度评价是对相机以及场景中的运动目标的定位精度进行评价。定位精度计算方法采用分段计算与真实轨迹之间的方向、位置偏差，计算出偏差后除以本段的长度，获得针对路程长度的平均角度和位置误差，再将各段之间相较

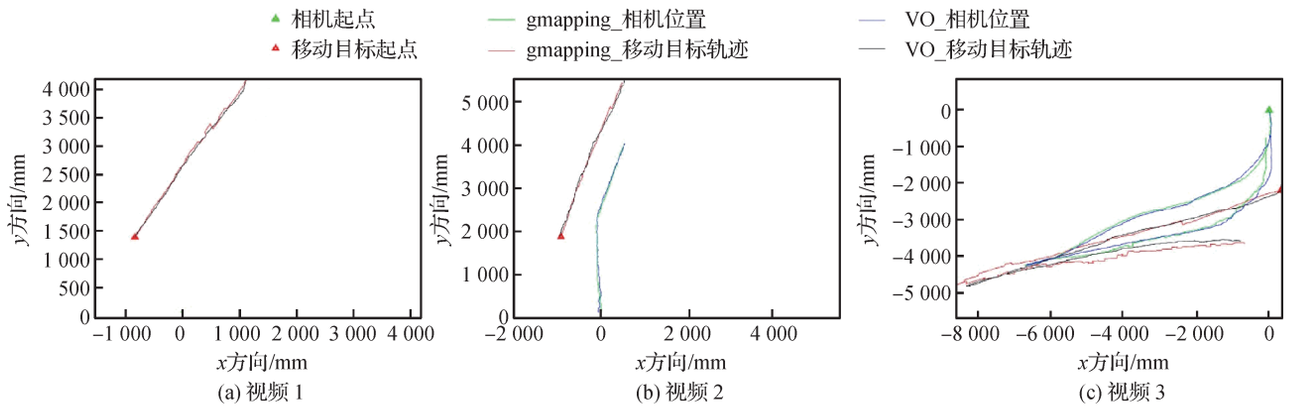


图 8　估计轨迹与 gmapping 真值对比

Fig. 8　Comparison of estimated trajectory and ground truth obtained from gmapping((a) video 1; (b) video 2; (c) video 3)

于长度的平均角度和位置偏差进行针对各分段的平均,获得最终的平均角度误差和平均位置误差,为最终的精度评价结果。

表 2 中展示了针对以上 3 组数据处理得到的定位精度,包括相机的位置误差和姿态误差、移动目标的位置误差和姿态误差。其中在评价视觉定位误差时,将视觉里程计生成的路径按照 400 mm、600 mm、800 mm、1 000 mm 和 1 200 mm 分别进行分段,计算每次相对于以上 5 种分段间隔的相机位置与姿态和对应时刻真实位置与姿态间的差异。

由于视频数据 1 在拍摄过程中相机平台保持静止不动,因此数据中的相机位置、姿态误差未统计分段误差。

表 2　多运动视觉里程计位置与姿态误差表

Table 2　Position and attitude error of multi-motion visual odometry

数据编号	对象		400 mm	600 mm	800 mm	1 000 mm	1 200 mm	平均误差
1	相机	位置误差/%	—	—	—	—	—	—
		姿态误差/(degree/mm)	—	—	—	—	—	—
	移动目标	位置误差/%	5.08	4.32	3.02	2.55	1.89	3.37
		姿态误差/(degree/mm)	0.004 1	0.003 5	0.002 5	0.002 1	0.001 7	0.002 8
2	相机	位置误差/%	4.17	3.21	2.52	2.12	1.63	2.73
		姿态误差/(degree/mm)	0.002 2	0.001 3	0.001	0.000 9	0.000 7	0.001 2
	移动目标	位置误差/%	5.38	4.32	3.66	2.99	2.47	3.76
		姿态误差/(degree/mm)	0.003 7	0.002 8	0.002 1	0.002 1	0.001 7	0.002 5
3	相机	位置误差/%	5.63	4.38	3.52	2.83	2.29	3.73
		姿态误差/(degree/mm)	0.005 2	0.004	0.003 2	0.002 8	0.002 8	0.003 6
	移动目标	位置误差/%	7.61	6.45	5.52	4.36	3.85	5.56
		姿态误差/(degree/mm)	0.005 7	0.004 9	0.004	0.003 5	0.003 3	0.004 3
平均误差	相机	位置误差/%	4.75	3.55	2.61	2.17	1.78	2.97
		姿态误差/(degree/mm)	0.003 7	0.002 8	0.002 2	0.001 9	0.001 7	0.002 4
	移动目标	位置误差/%	5.57	4.37	3.39	2.81	2.32	3.69
		姿态误差/(degree/mm)	0.004 8	0.003 6	0.002 8	0.002 6	0.002 3	0.003 2

注:“—”表示未统计分段误差。

可以看出，本文提出的多运动视觉里程计方法不仅能够有效估计相机的位置与姿态，还能估计出场景中存在的显著移动目标的位置与姿态，相机自定位与移动目标的定位结果与真实的定位结果较为接近，仅在很小的局部区域会存在一些偏差。

7 结 论

提出一种基于分裂合并运动分割的多运动视觉里程计方法，不同于通常的只处理单个运动模型（即相机运动模型）的视觉里程计方法，本文将多模型拟合方法引入通常视觉里程计方法中，分割出动态场景中的相机自身运动模型和不同运动的动态物体运动模型，进而同时估计出相机和各个动态物体的绝对运动轨迹，构建出多运动视觉里程计过程，初步解决了所提出的多运动视觉里程计的问题。

然而目前的实验场景中，除了相机平台以外只存在一个运动物体，即场景中只有两个运动模型，没有进行多于两个的运动模型存在情况下的实验；同时目前实验场景设置较为简单，没有在更加复杂的室外环境，且物体进行更复杂、更大范围运动情况下进行实验；并且目前还未考虑运动物体轨迹与相机轨迹的联合平差优化，轨迹估计精度还可以进一步提升。

未来的工作将立足于在更复杂、更多数量运动模型存在的场景中进行实验；同时构建运动物体轨迹与相机轨迹联合优化的平差方程，进一步提高轨迹估计的精度；并且在动态场景中，除了相机和各个运动物体轨迹外，研究同时获取静态背景密集 3 维地图，并对各个动态物体进行 3 维建模，进而构建一个动态场景 4D 密集地图（3 维加时间，即 3D 地图 + 运动目标时序 3D 模型）。

参考文献（References）

Costeira J P and Kanade T. 1998. A multibody factorization method for independently moving objects. International Journal of Computer Vision, 29(3): 159-179 [DOI: 10.1023/A:1008000628999]

Forster C, Pizzoli M and Scaramuzza D. 2014. SVO: fast semi-direct monocular visual odometry//Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong IEEE: 15-22 [DOI: 10.1109/ICRA.2014.6906584]

Grisetti G, Stachniss C and Burgard W. 2007. Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE Transactions on Robotics, 23(1): 34-46 [DOI: 10.1109/TRO.2006.889486]

Kumar M P, Torr P H S and Zisserman A. 2008. Learning layered motion segmentations of video. International Journal of Computer Vision, 76(3): 301-319 [DOI: 10.1007/s11263-007-0064-x]

Li R J, Yu S Y and Yang X K. 2007. Efficient spatio-temporal segmentation for extracting moving objects in video sequences. IEEE Transactions on Consumer Electronics, 53(3): 1161-1167 [DOI: 10.1109/TCE.2007.4341600]

Rublee E, Rabaud V, Konolige K and Bradski G. 2011. ORB: an efficient alternative to SIFT or SURF//Proceedings of 2011 International Conference on Computer Vision. Barcelona: IEEE: 2564-2571 [DOI: 10.1109/ICCV.2011.6126544]

Sabzevari R and Scaramuzza D. 2016. Multi-body motion estimation from monocular vehicle-mounted cameras. IEEE Transactions on Robotics, 32(3): 638-651 [DOI: 10.1109/TRO.2016.2552548]

Shen H F, Zhang L P, Huang B and Li P X. 2007. A MAP approach for joint motion estimation, segmentation, and super resolution. IEEE Transactions on Image Processing, 16(2): 479-490 [DOI: 10.1109/TIP.2006.888334]

Wiskott L. 1997. Segmentation from motion: combining Gabor- and Mallat-wavelets to overcome aperture and correspondence problem//Proceedings of International Conference on Computer Analysis of Images and Patterns. Berlin: Springer: 329-336 [DOI: 10.1007/3-540-63460-6_134]

Yue W T, Su J, Gu Z M, Yu C Y and Ge T. 2020. Best grid size of the occupancy grid map and its accuracy. Robot, 42(2): 199-206. 岳伟韬, 苏婧, 谷志珉, 余传义, 葛彤. 2020. 占据栅格地图的最佳栅格大小与地图精度. 机器人, 42(2): 199-206 [DOI: 10.13973/j.cnki.robot.190335]

Zappella L, Lladó X and Salvi J. 2008. Motion segmentation: a review//Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence. Sant Martí d'Empúries, Spain: IOS Press: 398-407 [DOI: 10.3233/978-1-58603-925-7-398]

Zhang J, Shi F H, Wang J H and Liu Y C. 2007. 3D motion segmentation from straight-line optical flow//Proceedings of 2007 International Workshop on Multimedia Content Analysis and Mining. Berlin: Springer: 85-94 [DOI: 10.1007/978-3-540-73417-8_15]

Zhang Y, Luo B and Zhang L P. 2018. Permutation preference based alternate sampling and clustering for motion segmentation. IEEE Signal Processing Letters, 25(3): 432-436 [DOI: 10.1109/LSP.2017.2777997]

作者简介

王晨捷,1996 年生,男,硕士研究生,主要研究方向为视觉 SLAM 和智能机器人。
E-mail: wangchenjie@ whu. edu. cn

罗斌,通信作者,男,教授,主要研究方向为摄影测量与遥感。
E-mail:luob@ whu. edu. cn

张云,男,博士, 主要研究方向为智能机器人和多运动分割。
E-mail: zhangyun198902@ sina. com

赵青,男,博士研究生,主要研究方向为智能机器人和视觉里程计。E-mail: zhaoqing@ whu. edu. cn

王伟,男,博士研究生,主要研究方向为智能机器人和多传感器融合 SLAM。E-mail: kinggreat24@ whu. edu. cn

尹露,男,博士研究生,主要研究方向为智能机器人和多传感器标定。E-mail: ccyinlu@ whu. edu. cn

张良培,男,教授,主要研究方向为高光谱遥感和模式分析与机器学习。E-mail: zlp62@ whu. edu. cn