

# 动态场景下基于运动物体检测的立体视觉里程计

林志林, 张国良, 姚二亮, 徐 慧

火箭军工程大学控制科学与工程系, 陕西 西安 710025

**摘要** 为了提升动态场景中视觉里程计的稳健性和精确度, 提出一种基于运动物体检测的立体视觉里程计算法。首先, 建立考虑相机位姿的场景流计算模型, 用于表示物体的运动矢量。其次, 提出构造虚拟地图点的方法, 一方面结合场景流进行运动物体检测, 另一方面使运动物体在图像中占比较大时仍有足够匹配点对用于位姿估计。最后, 通过局部地图点及虚拟地图点与当前帧特征点的匹配结果, 构建考虑虚拟点的非线性优化模型进行相机位姿估计, 既保证静态地图点不与运动物体的特征点形成错误匹配, 又避免因有效匹配点对过少而导致视觉里程计失效。数据集实验和实际场景在线实验结果表明, 本文算法提升了视觉里程计在动态场景中的稳健性和精确度。

**关键词** 机器视觉; 立体视觉里程计; 动态场景; 运动物体检测; 场景流; 虚拟地图点

**中图分类号** TP242.6 **文献标识码** A

**doi:** 10.3788/AOS201737.1115001

## Stereo Visual Odometry Based on Motion Object Detection in the Dynamic Scene

Lin Zhilin, Zhang Guoliang, Yao Erliang, Xu Hui

*Department of Control Science and Engineering, Rocket Force Engineering University,  
Xi'an, Shaanxi 710025, China*

**Abstract** In order to improve the robustness and accuracy of the visual odometry in the dynamic scene, a stereo visual odometry based on moving object detection is proposed. Firstly, a scene flow calculation model considering the camera pose is established to represent the motion vector of the objects. Secondly, the method of constructing virtual map points is proposed. On the one hand, the motion object detection can be complied according to the virtual map points and the scene flow, on the other hand, the virtual map points ensure that there are still enough matched point pairs to estimate the pose when the proportion of the moving objects in the image is too large. Finally, the feature points in the current frame will be matched with the local map points and the virtual map points, and according to the matching results, the nonlinear optimization model considering the virtual points is constructed to estimate the camera pose. It can not only ensure that the static map points do not match with the feature points of the motion object, but also avoid the failure of the visual odometry when the effective point pairs are too few. The results of the dataset experiment and the online experiment in the actual scene show that the proposed method improves the robustness and accuracy of the visual odometry in the dynamic scene.

**Key words** machine vision; stereo visual odometry; dynamic scene; motion object detection; scene flow; virtual map point

**OCIS codes** 150.0155; 330.1400; 110.2970; 330.4150; 110.4153

## 1 引 言

视觉里程计<sup>[1-2]</sup>(VO)是一种利用单个或多个视觉传感器的输入对运载体(如汽车、人、机器人)进行准确位姿估计的方法,因其使用成本低、功耗小、获取的信息丰富和定位精度高等优点,广泛应用于机器人学、增强现实和自动驾驶等领域<sup>[3]</sup>。立体视觉里程计<sup>[4-6]</sup>是其中一种被广泛研究的算法,能够获得视野的景深信息,不存在尺度因子模糊性的问题,解决了尺度二义性<sup>[7]</sup>,对环境和地形的适应性更好。然而,目前大部分视觉里程计需要假设环境中的特征是静态的或者特征的变化只依赖于相机自身的运动。而实际场景(如广场、

**收稿日期:** 2017-05-22; **收到修改稿日期:** 2017-06-26

**作者简介:** 林志林(1993—),男,硕士研究生,主要从事视觉 SLAM 方面的研究。E-mail: 15091281454@163.com

**导师简介:** 张国良(1970—),男,博士,教授,主要从事先进控制理论与应用方面的研究。

E-mail: lzl13468619594@163.com(通信联系人)

街道等)中难以避免地存在运动物体,若这些运动物体出现在相机采集的图像中且占比较大时,则难以通过随机采样一致性(RANSAC)算法将这些动态特征滤除,易将其作为内点计算里程计信息,从而引入误差,造成视觉里程计估计失效。

为了解决动态场景下的视觉里程计问题,必须引入运动物体检测(MOD)方法<sup>[8]</sup>将运动物体与静态特征区分开来。Bibby等<sup>[9]</sup>的SLAMIDE系统及Hahnel等<sup>[10]</sup>的研究利用期望最大化(EM)方法估计场景特征的运动状态,并通过可逆的模型选择来允许在视觉里程计中出现动态特征。这类方法对特征建立了概率分布模型,有助于对多个运动物体的检测。文献[11-12]将光流或场景流计算引入视觉里程计中,通过判断特征的光流或场景流进行运动物体检测。虽然上述方法能够剔除部分动态特征,但是容易在静态特征较少时,因用于位姿估计的匹配点对太少而造成位姿估计误差大。Choi等<sup>[13]</sup>根据运动物体的深度变化不同于静态特征,统计双目图像中特征的深度置信地图来进行运动物体检测;康轶非等<sup>[14]</sup>通过旋转平移解耦降低运动特征的影响。这类方法虽然不依赖于相机位姿初始估计,但是对图像信息使用过少,检测结果容易受图像噪声及匹配误差的影响。

本文在视觉里程计中引入场景流,并构造出图像特征的高斯混合模型进行运动物体检测,保证了对图像中存在多个运动物体的情况的稳健性。提出了构造虚拟地图点的方法,一方面进一步严格筛选运动物体的过程,减少将静态特征误认为运动特征的情况,另一方面将运动特征点与虚拟地图点形成的匹配点对纳入位姿估计的优化框架中,避免了图像中静态特征较少时用于位姿估计的匹配点对过少的情况。生成了由静态特征对应地图点构成的局部地图,而后根据图像帧中的特征与局部地图及虚拟地图点集合所形成的匹配点对构造投影误差函数,通过最小化该误差函数可以估计得到相机位姿。实验表明,本文算法可以保证立体视觉里程计在动态场景中仍具有良好的稳健性及较高的定位精度。

## 2 基于场景流的运动物体检测

### 2.1 考虑帧间位姿的场景流计算模型

为了能够在相机运动的条件下进行场景流的计算,将帧间位姿引入场景流的计算模型中。根据相机运动模型可以获取当前帧较为精确的位姿估计,因此可以将上一帧图像经过位姿变换再与当前帧进行场景流的计算,经位姿变换后即可类似于在相机静止条件下进行场景流的计算。

假设基于相机运动模型估计得到相邻两帧之间的旋转变换为 $\mathbf{R}$ ,平移变换为 $\mathbf{t}$ ,这两帧之间形成匹配点对 $\{(u, v) \rightarrow (u', v')\}$ ,特征点对应的深度分别为 $h$ 和 $h'$ ,此处的相机由深度相机直接获取或根据双目相机模型<sup>[15]</sup>获取,那么,可以获得对应的点在当前帧坐标系中的坐标分别为

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = h \cdot \mathbf{R} \cdot \mathbf{K}^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + \mathbf{t}, \quad (1)$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = h' \cdot \mathbf{K}^{-1} \cdot \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}, \quad (2)$$

式中 $\mathbf{K}$ 为相机内参矩阵,可由相机焦距 $(f_x, f_y)$ 及光学投影中心坐标 $(c_x, c_y)$ 表示为

$$\mathbf{K} = \begin{bmatrix} \frac{1}{f_x} & 0 & -\frac{c_x}{f_x} \\ 0 & \frac{1}{f_y} & -\frac{c_y}{f_y} \\ 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

那么,该特征点对应的3D运动矢量即场景流为

$$\mathbf{M} = \begin{pmatrix} X' - X \\ Y' - Y \\ Z' - Z \end{pmatrix}. \quad (4)$$

由于存在匹配误差、投影误差等的影响,部分静态物体对应的特征点也会存在非零的运动矢量,所以这一运动矢量不能直接用于区分场景中的运动物体及静态物体。

### 2.2 虚拟地图点的构造

为了能够充分利用前一帧运动物体检测的结果,本文构造虚拟地图点(VP)。如图 1 所示,根据运动物体的运动模型,估计出运动物体对应的地图点(称为运动点)及运动状态未知的地图点(称为状态未知点)在下一时刻的 3D 位置,由估计得到的地图点组成虚拟地图点集合。

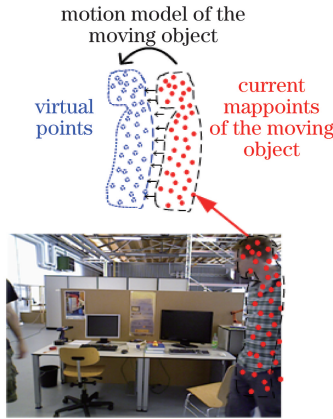


图 1 虚拟地图点生成原理图

Fig. 1 Schematic of generating the virtual map points

在相机位姿估计结束后,根据估计结果将运动特征点集合  $\mathcal{P}$  和状态未知特征点集合  $\mathcal{Q}$  中已知深度为  $z_c$  的特征点  $(u, v)$  投影为地图点  $(x_k, y_k, z_k)$ :

$$\mathbf{P}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = \mathbf{R}_k^w \cdot \begin{bmatrix} (u - c_x) \cdot z_c / f_x \\ (v - c_y) \cdot z_c / f_y \\ z_c \end{bmatrix} + \mathbf{t}_k^w, \quad (5)$$

式中  $\mathbf{R}_k^w$  和  $\mathbf{t}_k^w$  分别为  $k$  时刻相机在师姐坐标系中的旋转矩阵和平移变换矩阵。

根据前一时刻及当前时刻的地图点位置,可以得到运动物体的运动模型  $\mathbf{m}_k$  为

$$\mathbf{m}_k = \mathbf{P}_k - \mathbf{P}_{k-1}. \quad (6)$$

则可构建虚拟地图点为

$$\mathbf{P}_{k+1}^{(v)} = \mathbf{m}_k + \mathbf{P}_k. \quad (7)$$

### 2.3 基于高斯混合模型的运动物体检测

假设属于同一运动物体的运动矢量的分布服从高斯模型,那么同一场景中多个运动物体的运动矢量则服从高斯混合模型(GMM)。

当一个物体运动时,它所对应的运动矢量角度[图 2(a)]是趋于一致的,因此运动矢量的角度服从高斯混合模型,使用 EM 方法<sup>[16-17]</sup>求解对应的高斯混合模型具体参数。如图 2(b)所示,根据求得的高斯混合模型,可以确定静态物体及不同运动物体所对应的高斯模型。为了确定某一特征点是否属于运动物体,以该特征点在静态物体对应的高斯模型(均值较小的高斯模型)中的高斯值作为阈值  $\psi_{th}$ ,计算该特征点在运动物体对应的高斯模型中的高斯值  $\psi_p$ ,将满足  $\psi_p > \psi_{th}$  的特征点列入待选运动点集合  $\mathcal{T}$  中。

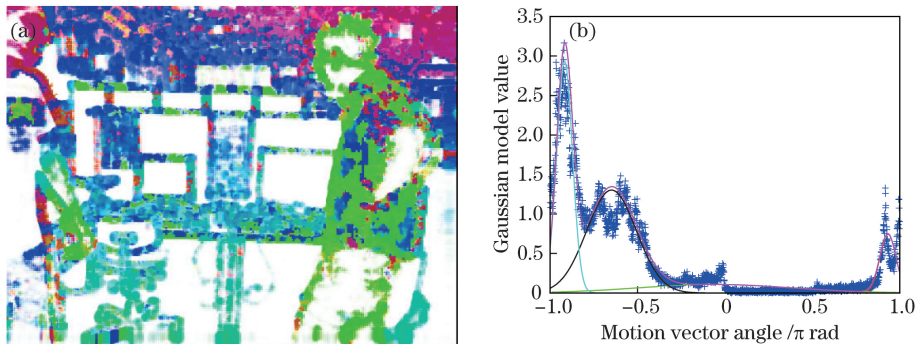


图 2 (a)物体运动矢量角度示意图;(b)对应的 GMM 结果

Fig. 2 (a) Schematic of the object motion vector angle; (b) corresponding result of GMM

存在匹配误差时,场景流的计算也服从高斯分布,因此为了排除匹配误差等对运动物体检测的影响,本文将待选运动点集合  $\mathcal{T}$  与虚拟地图点集合  $\mathcal{V}$  进行特征匹配,以  $\phi_p$  表示待选运动点  $p \in \mathcal{T}$  的匹配结果:

$$\phi_p = \begin{cases} \text{true, success} \\ \text{false, fail} \end{cases} \quad (8)$$

将匹配成功的待选运动点列入运动点集合  $\mathcal{D}$  中,否则将其列入状态未知点集合  $\mathcal{Q}$  中,以  $\varphi_p$  表示点  $p$  的运动状态:

$$\varphi_p = \begin{cases} \text{moving,} & \psi_p > \psi_{th} \text{ and } \phi_p \text{ is true} \\ \text{unknown,} & \psi_p > \psi_{th} \text{ and } \phi_p \text{ is false.} \\ \text{static,} & \psi_p \leq \psi_{th} \end{cases} \quad (9)$$

### 3 基于运动物体检测的立体视觉里程计

图 3 为本文算法框架图,本文算法共分为两部分:运动物体检测和相机位姿跟踪。在进行了运动物体检测后,运动点已经与虚拟地图点形成了匹配,而对于静态特征点,借鉴文献[18]构造局部地图的方法,使静态特征点与局部地图点建立匹配关系。一方面,匹配局部地图的方法使得位姿估计时有更多的约束来提高相机位姿估计结果的准确性,另一方面,匹配虚拟地图点的方法使得运动物体在图像帧中占比较大时,仍能够有足够的匹配点对来保证相机位姿估计的正常进行。

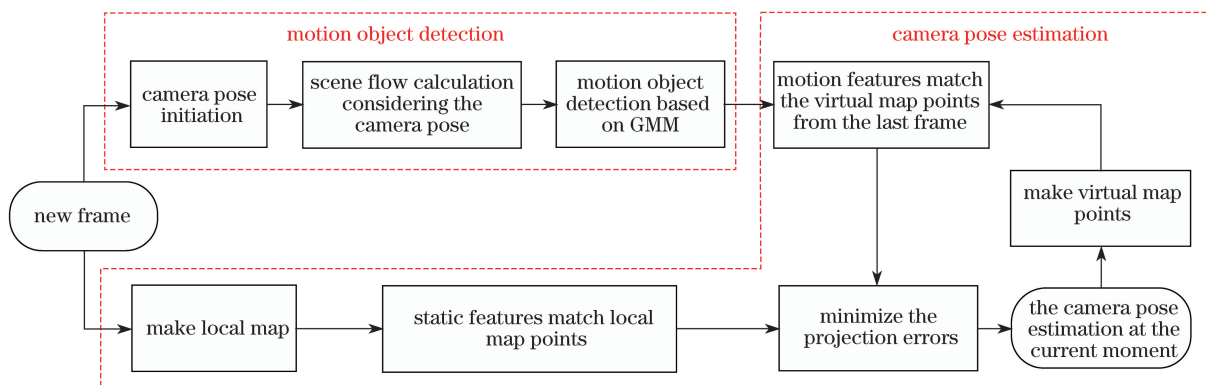


图 3 算法框架图

Fig. 3 Diagram of the algorithm

#### 3.1 局部地图的构造

图 4 为运动物体检测的结果,其中红点表示属于运动物体的特征点。如果将所有的地图点与当前帧特征点进行匹配,计算量大且时间消耗多,因此本文只构建局部地图与当前帧特征点进行匹配,这既能提升位姿估计的精确度,又能保证耗时不会过多。如图 5 所示,局部地图是由当前帧共视地图点的关键帧集合  $\mathcal{H}_1$  及其共视地图点的关键帧集合  $\mathcal{H}_2$  组成。局部地图点即为被  $\mathcal{H}_1$  和  $\mathcal{H}_2$  观测到的地图点集合  $\mathcal{L}$ 。



图 4 运动物体检测结果

Fig. 4 Results of motion object detection

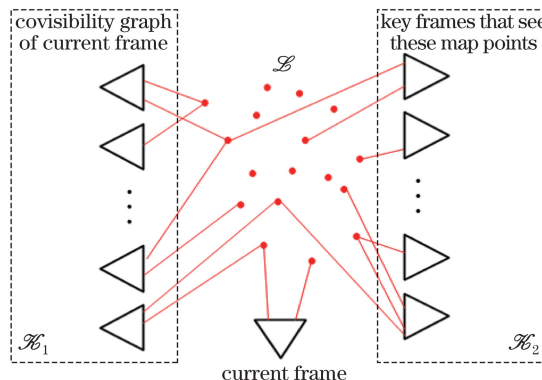


图 5 局部地图示意图

Fig. 5 Schematic of local map

### 3.2 考虑虚拟点的相机位姿估计

将局部地图点集合  $\mathcal{L}$  和虚拟地图点集合  $\mathcal{V}$  与当前帧特征点集合  $\mathcal{F}$  进行匹配,对于所形成的匹配点对  $\{P_i \rightarrow p_i | P_i \in \mathcal{L} \cup \mathcal{V}, p_i \in \mathcal{F}\}$ , 计算匹配之后的投影误差:

$$\boldsymbol{\varepsilon}_i = \pi(\mathbf{R} \cdot P_i + \mathbf{t}) - p_i, \quad (10)$$

式中  $\pi(\cdot)$  为相机投影函数,表示将地图点投影到相机投影平面获得像素坐标。

通过最小化如下误差函数可以获取相机的位姿估计:

$$F = \sum_i \|\boldsymbol{\varepsilon}_i\|_{\boldsymbol{\Omega}} = \sum_i \|\pi(\mathbf{R} \cdot P_i + \mathbf{t}) - p_i\|_{\boldsymbol{\Omega}}, \quad (11)$$

式中  $\boldsymbol{\Omega}$  为信息矩阵,  $\|\boldsymbol{\varepsilon}_i\|_{\boldsymbol{\Omega}} = \boldsymbol{\varepsilon}_i^T \cdot \boldsymbol{\Omega} \cdot \boldsymbol{\varepsilon}_i$ 。将当前帧划分为  $10 \times 10$  的图像块, 分别在每个图像块  $B_i$  中, 计算与虚拟地图点构成的匹配点对在该图像块中所有匹配点对中所占的比例  $r_i$ , 也即  $r_i = n_i / \sum_j n_j$ , 那么信息矩阵中的对角元素正比于  $1/r_i$ , 也即  $\boldsymbol{\Omega}_i \sim 1/r_i$ 。信息矩阵的设置表征了区域中虚拟点占比越多, 就越不信任其对误差函数的贡献。这是因为虚拟点是由运动点根据运动物体的运动模型计算得到的, 而运动物体的运动并不是绝对满足运动模型的, 所以与虚拟点形成正确匹配的概率要小于与局部地图点形成正确匹配的概率。另外, 一个区域内虚拟地图点匹配点对越多, 说明该区域属于运动物体的概率越大, 该区域内的点被误认为静态点的可能性越高, 因此应该将其信任度降低。

## 4 实验与分析

实验所用电脑配置为: CPU 为 i7 处理器, 主频为 2.5 GHz, 内存为 4G, 不使用图形处理器 (GPU) 加速, 系统为 Ubuntu 14.04。本文使用 TUM 数据集<sup>[19]</sup> 进行实验, 与 ORB-SLAM2<sup>[4]</sup> 方法进行对比, 检验本文算法在动态场景下的运行效果。并且使用 Bumblebee2 双目相机进行在线实际场景实验, 该相机采用全局快门曝光方式, 帧率为 30 frame/s, 图像分辨率为 640 pixel  $\times$  480 pixel, 基线为 12 cm。

### 4.1 TUM 数据集实验

TUM 数据集<sup>[19]</sup> 是慕尼黑工业大学实验场景数据集, 其中的 fr3 系列 (包括 sitting\_xyz、sitting\_halfsphere、sitting\_rpy、sitting\_static、walking\_xyz、walking\_halfsphere、walking\_rpy 和 walking\_static 等序列) 中相机始终对着桌子, 不同序列中相机运动轨迹及人的移动方式不一样, 是一个典型的动态场景序列集。在 TUM 的 fr3 系列数据集中分别应用 ORB\_SLAM2 算法和本文算法估计相机轨迹, 并与真实轨迹比较, 评价两种方法在动态场景中的稳健性及精确度。

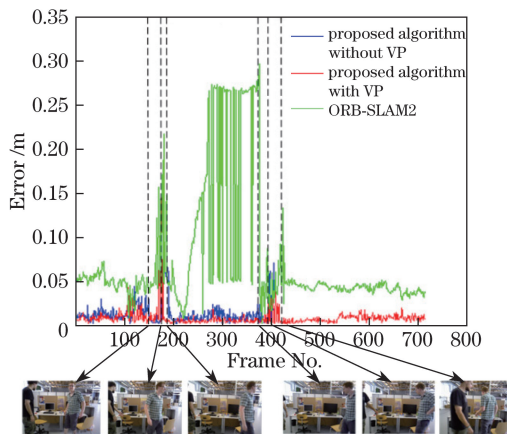


图 6 数据集对比实验结果

Fig. 6 Results of the contrast experiment on TUM dataset

图 6 为序列 walking\_static 的轨迹误差对比图。从图 6 可以看出, ORB-SLAM2 方法的结果 (绿线) 受运动物体的影响, 大约在第 160 帧到第 192 帧之间, 虽然相机基本不动, 但是估计的相机位姿随人向右侧运动发生了偏移, 且偏移量较大, 这一误差持续累积直至第 377 帧, 这是因为该算法误认为是相机运动引起场景变化, 从而错误估计了相机位姿。而在第 377 帧之后虽然误差减少, 但这只是因为此时人由右向左运动,

所引起的误差刚好与前期的累积误差相反,从而使最终误差减少。而本文算法(红线)能够在图像中有物体运动时仍保持较为准确的相机位姿估计,这是因为在本文的视觉里程计中进行了运动物体检测,同时使运动物体的特征点不再错误地与静态地图点形成匹配,而是被独立地与虚拟地图点进行匹配,这使得进行相机位姿估计时使用正确的特征点与地图点的匹配点对,所以本文算法受运动物体的影响较小。

另外,如图 6 中蓝线所示,当本文算法不加入生成虚拟地图点的方法时,约第 416 帧前视觉里程计的精度虽然低于加入生成虚拟地图点方法的本文算法,但仍高于 ORB-SLAM2 方法,说明本文算法得到的待选运动点集合  $\mathcal{T}$  已经包含了大部分运动特征点,达到了较好的运动物体检测效果。但在第 416 帧时视觉里程计失效,这是因为此时运动物体在图像中占比很大,如果运动物体的特征点既不与局部地图点形成匹配,又不与虚拟地图点匹配,那么可供位姿估计的匹配点对数量将会很少,所估计得到的误差将会很大从而造成视觉里程计失效,也即是跟丢第 417 帧,所以本文算法是有效可行的。同时,本文算法的耗时情况如表 1 所示,由表 1 可知,基于本文算法进行的数据集实验中,仅位姿估计部分的耗时约为 34.44 ms(即 30.03 Hz),满足实时定位的要求。

表 1 数据集对比实验运行时间表  
Table 1 Run time of the contrast experiments on dataset

Operation	ORB extraction	Pose initiation	MOD	Track $\mathcal{L}$	Track $\mathcal{V}$	Total
Time /ms	10.92	3.44	1.27	16.31	2.50	34.44

进一步基于 fr3 的其余数据集进行了对比实验,实验结果如表 2 所示(加粗为误差较小者),在运动物体运动幅度很小时(如 sitting 系列),本文算法与 ORB-SLAM2 方法的位姿估计误差相近,但在运动物体运动幅度很大,也即场景动态属性较强时(如 walking 系列),本文算法明显优于 ORB-SLAM2 方法,更充分说明了本文算法的稳健性强于 ORB-SLAM2 方法。

表 2 对比实验结果表  
Table 2 Results of the contrast experiment

Method	ORB-SLAM2 method	Proposed algorithm
Sitting_xyz	<b>0.009921</b>	0.010144
Sitting_halfsphere	<b>0.021115</b>	0.023901
Sitting_rpy	0.022008	<b>0.018039</b>
Sitting_static	0.008294	<b>0.008146</b>
Walking_xyz	0.426640	<b>0.225116</b>
Walking_halfsphere	0.447923	<b>0.042330</b>
Walking_rpy	0.740774	<b>0.196101</b>
Walking_static	0.102837	<b>0.012093</b>

#### 4.2 实际场景在线实验

图 7 为 Bumblebee2 双目相机。本文设置了如下的在线实验场景:固定相机位置,在相机视野内人员按照如图 8 路线行走,设置的场景中运动人员在相机近处及远处前后移动、左右移动、曲线运动,较为全面地包含了动态场景的特征。



图 7 Bumblebee2 双目相机  
Fig. 7 Bumblebee2 stereo camera



图 8 运动人员运动路线  
Fig. 8 Route of the moving person

分别基于 ORB-SLAM2 和本文算法进行实际场景在线实验,实验结果如图 9 和图 10 所示,图 9 和图 10 中(a)为所估计得到的相机运动轨迹,(b)为相机位置估计误差图,(c)~(h)为(b)中 A~F 位置标出的对应时刻的图像帧及相机位姿状态。

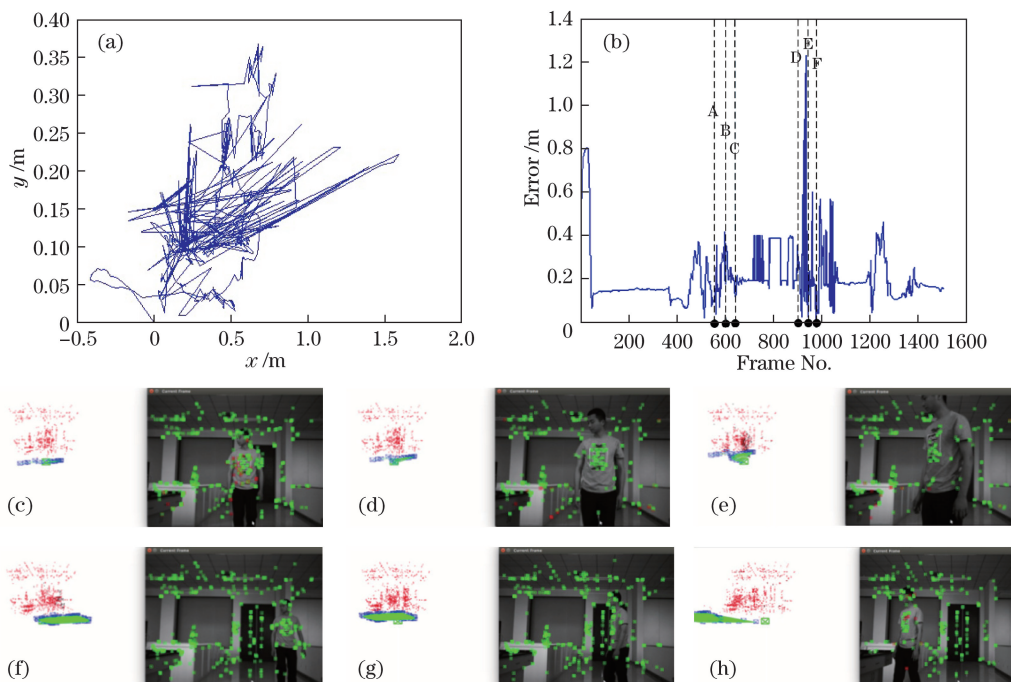


图 9 ORB-SLAM2 方法在线实验结果。(a)估计的相机轨迹;(b)相机位置估计误差;  
(c)~(h) A~F 位置对应的图像帧及相机位姿状态

Fig. 9 Results of online experiment based on ORB-SLAM2 method. (a) Estimated trajectory of camera; (b) error of the estimated position of camera; (c)-(h) image frames and poses of camera corresponding to position A-F

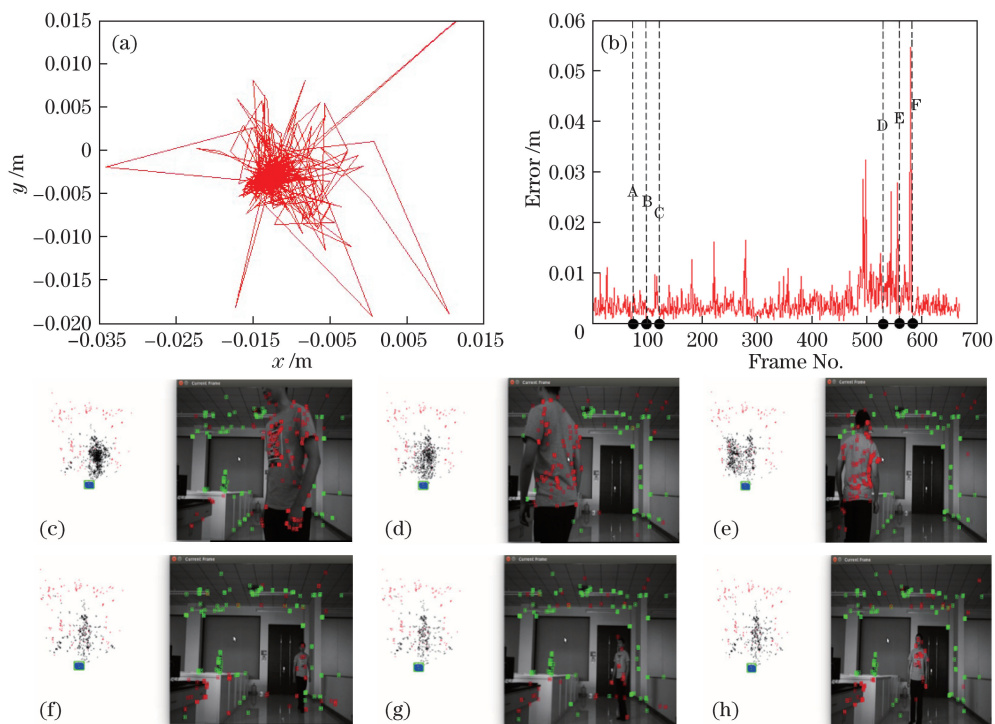


图 10 本文算法在线实验结果。(a)估计的相机轨迹;(b)相机位置估计误差;  
(c)~(h) A~F 位置对应的图像帧及相机位姿状态

Fig. 10 Results of online experiment based on proposed algorithm. (a) Estimated trajectory of camera; (b) error of the estimated position of camera; (c)-(h) image frames and poses of camera corresponding to position A-F

由图 9 可知,基于 ORB-SLAM2 方法的相机位姿估计的均方根误差(RMSE)为 0.249581 m。运动人员在相机近处(A~C)及相机远处(D~F)时相机的轨迹受场景中运动人员的影响较大,场景变化时算法误认为相机正在运动,所以估计得到的相机位姿与实际相机状态偏差较大。

由图 10 可知,基于本文算法的相机位姿估计的 RMSE 为 0.005159 m。运动人员在相机近处(A~C)及相机远处(D~F)时,相机的轨迹基本不受场景中运动人员的影响,场景变化时相机位姿仍能保持准确估计,与实际相机状态基本一致。

从实际场景在线实验中可以看出,本文算法很好地克服了动态场景中运动物体带来的干扰,当场景中存在左右移动、前后移动或曲线运动的物体时,均能够保持较强的稳健性,在线实验效果充分验证了本文算法的可行性和有效性。

## 5 结 论

动态场景中的视觉里程计问题一直是视觉里程计研究的重点和难点。通过将运动物体检测引入立体视觉里程计中,降低了动态场景对相机位姿估计的影响,使得在动态场景中视觉里程计的稳健性更好、精确度更高、实用性更强。

在立体视觉下基于场景流进行运动物体检测,提出基于运动物体检测结果建立虚拟地图点的方法,生成的虚拟地图点和其余的静态地图点共同与图像中的特征点形成匹配点对,利用所有的匹配点对进行非线性优化,得到相机的位姿估计。数据集实验和实际场景在线实验结果表明,本文算法与 ORB-SLAM2 方法相比,具有更好的稳健性。即使运动物体在图像中占比很大时,基于本文算法的立体视觉里程计也能够基本不受动态场景的影响,保持准确的相机位姿估计。

下一步计划进行如下两方面的改进:1)在开始运动物体检测之前使用惯性组合元件进行位姿估计,从而得到更准确的初始估计,使运动物体检测结果更精确;2)在图像中加入语义属性,以提升视觉里程计的实用性和可扩展性。

## 参 考 文 献

- [1] Wang Ke, Jia Songmin, Li Xiuzhi, *et al.* Mobile robot monocular visual odometry algorithm based on ground features[J]. *Acta Optica Sinica*, 2015, 35(5): 0515002.  
王可, 贾松敏, 李秀智, 等. 基于地面特征的移动机器人单目视觉里程计算法[J]. *光学学报*, 2015, 35(5): 0515002.
- [2] Zhao Zhenqing, Ye Dong, Chen Gang, *et al.* Binocular vision method of measuring pose based on perpendicular lines[J]. *Acta Optica Sinica*, 2014, 34(10): 1015003.  
赵振庆, 叶东, 陈刚, 等. 垂直直线特征的双目视觉位姿测量方法[J]. *光学学报*, 2014, 34(10): 1015003.
- [3] Fraundorfer F, Scaramuzza D. Visual odometry Part II: matching, robustness, optimization, and applications[J]. *IEEE Robotics & Automation Magazine*, 2012, 19(2): 78-90.
- [4] Mur-Artal 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): 1255-1262.
- [5] Engel J, Stueckler J, Cremers D. Large-scale direct SLAM with stereo cameras[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015: 1935-1942.
- [6] Forster C, Zhang Z, Gassner M, *et al.* SVO: semi-direct visual odometry for monocular and multicamera systems[J]. *IEEE Transactions on Robotics*, 2017, 33(2): 249-265.
- [7] Maimone M, Cheng Y, Matthies L. Two years of visual odometry on the Mars exploration rovers[J]. *Journal of Field Robotics*, 2007, 24(3): 169-186.
- [8] Mo Shaowen, Deng Xinpu, Wang Shuai, *et al.* Moving object detection algorithm based on improved visual background extractor[J]. *Acta Optica Sinica*, 2016, 36(6): 0615001.  
莫邵文, 邓新蒲, 王帅, 等. 基于改进视觉背景提取的运动目标检测算法[J]. *光学学报*, 2016, 36(6): 0615001.
- [9] Bibby C, Reid I. Simultaneous localisation and mapping in dynamic environments (SLAMIDE) with reversible data association[C]. *Proceedings of Robotics Science and Systems*, 2007: 105-112.
- [10] Hahnel D, Triebel R, Burgard W, *et al.* Map building with mobile robots in dynamic environments[C]. *IEEE International Conference on Robotics and Automation*, 2003: 1557-1563.



- [11] Alcantarilla P F, Yebes J J, Almazán J, *et al.* On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments[C]. IEEE International Conference on Robotics and Automation, 2012: 1290-1297.
- [12] Shimamura J, Morimoto M, Koike H. Robust vSLAM for dynamic scenes[C]. MVA 2011 IAPR Conference on Machine Vision Applications, 2011: 344-347.
- [13] Choi J W, Lee S H, Song H K, *et al.* Stable stereo based EKF-SLAM in dynamic situation[J]. Advanced Science and Technology Letters (ASTL), 2015, 118: 10-13.
- [14] Kang Yifei, Song Yongduan, Song Yu, *et al.* Stereovisual odometry algorithm with rotation-translation decoupling for dynamic environments[J]. Robot, 2014, 36(6): 758-768.  
康轶非, 宋永端, 宋宇, 等. 动态环境下基于旋转-平移解耦的立体视觉里程计算法[J]. 机器人, 2014, 36(6): 758-768.
- [15] Zhang Qiang, Dong Yu, Li Haibin, *et al.* Underwater curve constraint model based on non-parallel binocular vision[J]. Acta Optica Sinica, 2016, 36(3): 0315001.  
张强, 董玉, 李海滨, 等. 非平行双目视觉系统水下成像的极线约束模型[J]. 光学学报, 2016, 36(3): 0315001.
- [16] Tan W, Liu H, Dong Z, *et al.* Robust monocular SLAM in dynamic environments[C]. 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2013: 209-218.
- [17] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society Series B (Methodological), 1977: 1-38.
- [18] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [19] Sturm J, Engelhard N, Endres F, *et al.* A benchmark for the evaluation of RGB-D SLAM systems[C]. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012: 573-580.