

基于车载双目立体视觉的运动障碍检测

孙浩¹, 周石琳¹, 邹焕新¹, 王程²

- (1. 国防科学技术大学 电子科学与工程学院, 湖南 长沙, 410073;
2. 厦门大学 信息科学与技术学院, 福建 厦门, 361005)

摘要: 提出一种基于多视几何约束的车载双目立体视觉运动障碍检测算法。研究结果表明: 多视极线约束有效地解决了当相机和障碍同方向运动时对极几何无法检测运动障碍的问题。在粒子滤波框架下, 以多视极线约束为状态观测, 可同时检测和跟踪多个运动障碍。不同场景实际数据集上的实验结果证明: 该方法具有鲁棒性和有效性。

关键词: 无人驾驶; 运动障碍检测; 立体视觉; 多视几何约束

中图分类号: TP391.41

文献标志码: A

文章编号: 1672-7207(2013)S2-0113-04

Surrounding moving obstacles detection using onboard binocular stereo vision

SUN Hao¹, ZHOU Shilin¹, ZOU Huanxin¹, WANG Cheng²

- (1. School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410083, China;
2. School of Information Science and Technology, Xiamen University, Xiamen 361005, China)

Abstract: A multiview epipolar constraint based algorithm was proposed for moving obstacles detection using onboard binocular stereo vision. The results show that a multiview epipolar constraint is derived from the relative camera positions in pairs of consecutive stereo views for independent motion detection. A dimensional variable particle filter for joint detection and tracking of multiple moving obstacles is presented. Experimental results on real-world driving sequences demonstrate that the method is effective and robust.

Key words: smart vehicles; moving obstacles detection; stereo vision; multiview geometric constraint

近年来, 学者们开始研究在城市交通环境下的自动驾驶汽车, 独立运动障碍物(行人、机动车、汽车等)检测是其中的核心模块^[1-6]。现有的解决方案主要关注基于多传感器融合的方法, 但系统造价昂贵。视觉传感器相对于其他传感器而言, 能够提供最为丰富的感知信息, 是一种被动的传感器, 最接近人类的环境感知方式, 且其价格也较为适中。车载运动成像条件下, 相机在世界坐标系中的三维运动导致图像中的每个像素点产生二维运动, 其中, 运动障碍像素点的二维运动是由障碍的运动和相机的运动联合产生的, 而静态背景像素点的二维运动仅仅由相机的运动引起。相机

的运动会产生多视几何约束, 满足约束的像素为背景像素, 不符合约束的像素为运动障碍像素。用于运动障碍检测的多视几何约束大致可以分为 2 类: 第一类是由多个视角的相机相互姿态关系导出的约束, 又称为“相对姿态”约束, 此类约束不对观测的场景做任何假设; 第二类约束又称为“平面-视差”约束, 假设场景中有一个三维的参考平面^[1]。一般来讲, 相对姿态约束在稀疏特征匹配点时可获得比较好的效果, 但其需要准确的特征匹配并且相邻图像间的相机基线要大。平面-视差约束对图像的噪声不敏感, 更适合于相机运动较小情况下的稠密特征匹配, 但其只能应用

收稿日期: 2013-03-01; 修回日期: 2013-05-02

基金项目: 教育部高等学校博士学科点专项科研基金资助项目(20124307120013)

通信作者: 孙浩(1984-), 男, 陕西三原人, 博士, 讲师, 从事计算机视觉与智能信息处理研究; 电话: 0731-84574724; E-mail: clhaosun@gmail.com

于场景可以近似的表示为一个平面的情形。针对双目会聚式立体视觉,本文作者提出了一种基于 4 个视角的多视极线约束,用以检测运动障碍。算法首先在输入的相邻时刻立体图像对中检测 SIFT 特征,然后采用闭环特征匹配方法^[6]得到立体特征对和运动特征对,最后在粒子滤波体系下采用自适应状态预测和多视极线约束状态观测对系统状态进行迭代更新,同时检测和跟踪的多个运动障碍。

1 多视极线约束

给定同一静态场景的 2 个不同视角 C_1 和 C_2 下的 2 幅投影图像,假设三维坐标点 P 在第 1 幅图像中的投影点表示为 p_1 ,那么其在第 2 幅图像中的投影点 p_2 一定位于由 p_2 导出的极线上。对极几何约束描述了 2 个相机姿态间的相对位置关系,包含了 2 个光学中心间的相机运动信息,可用于 2 个视角中的独立运动检测。在相机未校准的情况下,对极几何约束可以通过基本矩阵 F_{21} 表示。基本矩阵把同一个三维点在 2 个不同视角中的投影点通过下式联系起来:

$$p_2^T F_{21} p_1 = 0 \quad (1)$$

令 $l_2 = F_{21} p_1$ 表示投影点 p_1 在视角 2 中产生的极线,如果三维点静止,那么理想情况下, p_2' 应该位于直线 l_2 上。类似的, p_1 应该位于极线 $l_1' = F_{21}^T p_2'$ 上。但当投影像素点沿极线运动时,对极几何约束无法检测出独立运动目标,这种情况经常发生在相机和目标沿相同方向运动时。为了解决像素沿极线运动无法被检测的情形,需要引入更多约束。

令 $C_{v,n}(v=r, l; n=1, 2, \dots, t)$ 表示 n 时刻的双目立体相机系统(r 表示右视相机, l 表示左视相机),在相邻帧的立体图像对中,存在 4 组对极极线约束, $\{F_{r,r(t+1)}, F_{r,l(t+1)}, F_{l,r(t+1)}, F_{l,l(t+1)}\}$ 。单组的 $F_{r,r(t+1)}$ 或者 $F_{l,l(t+1)}$ 不能用于检测图像中的独立运动。第一,当相机运动方向与独立运动目标运动方向相同时,极线约束不能检测独立运动;第二,由于单目相机相邻图像间的相机运动幅度较小,基本矩阵 $F_{r,r(t+1)}$ 或者 $F_{l,l(t+1)}$ 的估计对噪声和特征匹配中的误差比较敏感^[7-10]。由于立体相机的相对几何关系和相机的运动, $C_{r,t}$ 和 $C_{l,t+1}$, $C_{l,t}$ 和 $C_{r,t+1}$ 之间的相机运动比较大,因此可以稳定的估计基本矩阵 $F_{r,l(t+1)}$ 和 $F_{l,r(t+1)}$ 。而且,当同时使用时,由于双目相机的光轴是相交的,因此理论上讲,不存在独立运动向量同时满足基本矩阵 $F_{r,l(t+1)}$ 和 $F_{l,r(t+1)}$ 的约束关系。 $\{F_{r,l(t+1)}, F_{l,r(t+1)}\}$ 本质上是两组交

错的极几何约束,称其为多视极线约束。多视极线约束包含了会聚式双目相机之间的相互位置关系和相机的运动信息,可以稳定地检测出独立运动障碍。定义多视像素直线距离用来衡量投影像素点偏离立体交错极线的程度:

$$d = \alpha(|l_{r,t} \cdot p_{r,t}| + |l_{l,t+1} \cdot p_{l,t+1}|) + \beta(|l_{l,t} \cdot p_{l,t}| + |l_{r,t+1} \cdot p_{r,t+1}|) \quad (2)$$

其中: $|l_{v,n} \cdot p_{v,n}|(v=r, l; n=t, t+1)$ 表示投影点 $p_{v,n}$ 到其对应的交错极线 $l_{v,n}$ 的垂直距离; α 和 β 为归一化的加权因子。运动障碍的投影点不满足多视极线约束,投影像素点到对应极线的距离比较大,而静止场景的投影像素点到对应极线的距离比较小, d 可以作为独立运动的一个度量。

2 基于粒子滤波的运动障碍检测

运动障碍的动态特性通常是非线性、非高斯的,无法获得解析的表达形式,采用粒子滤波框架进行建模,系统状态向量 $x_t = [u_t, v_t, o_t]^T$ 为运动障碍投影点的位置坐标(u_t, v_t)和跟踪标记 o_t 。由于障碍的运动与相机的运动是独立的,无法直接预测障碍的运动模式,一个直接的解决方法是根据历史光流向量进行预测状态的传递,系统状态传递模型定义为:

$$x_t = \begin{cases} 0, o_t < 0 \\ x_{t-1} + s_{t-1,t} + w_t, o_t > 0 \\ \Phi_b + w_t, o_t = 0 \end{cases} \quad (3)$$

其中, $s_{t-1,t} = [du, dv, 0]^T$ 表示图像坐标系中投影点的运动向量,由闭环匹配获得; Φ_b 表示没有后向运动匹配向量的投影点集,这些投影点属于刚刚进入视野的运动障碍; w_t 表示离散的白噪声序列,服从零均值的标准正态分布。系统状态传递采用自适应的策略进行系统状态进化,随着运动障碍进入或离开视野,粒子滤波的维数自适应的变化,进入视野时 $o_t=0$ 投影点自动地加入到系统粒子集合中,离开视野时 $o_t < 0$ 投影点被去除。已经检测到的运动障碍采用历史观测运动向量进行预测跟踪。

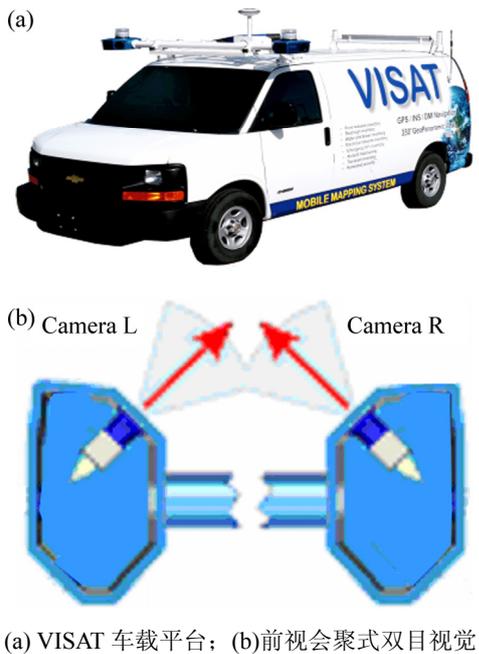
以似然函数对观测过程进行建模,由于是投影点属于运动障碍的概率,而多视极线距离函数是独立运动信息的一个度量,因此定义似然函数为:

$$p(y_t | x_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (4)$$

其中, d 为归一化的多视极线距离, 由式(2)计算。观测过程受到零均值方差为 σ^2 的正态噪声影响。在粒子滤波基础上, 要想从离散的粒子表示方式获得运动目标, 需要进一步处理。提出了基于空间距离和权重的二维聚类算法, 对空间粒子进行聚类, 设定相应的门限后, 提取出运动障碍的各种属性, 其中包含目标的外接矩形框、中心位置、特征点对数目、独立运动平均概率。

3 实验结果

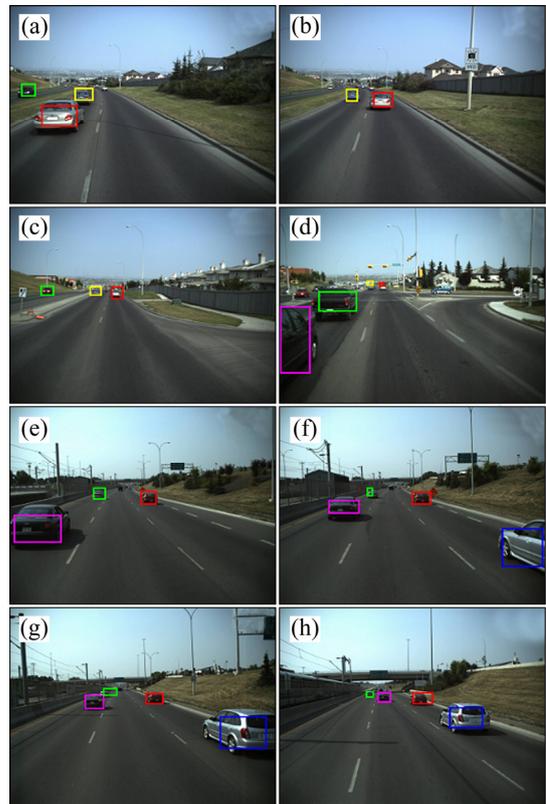
为了验证提出的方法, 采用 VISAT™ 车载移动成像平台(图 1(a))获得的视频数据。VISAT™ 视觉^[11]系统为多目的数字相机系统, 视野覆盖 0~360°, 实验中, 只利用具有会聚式结构的双目前视相机采集到的视频数据, 如图 1(b)所示。



(a) VISAT 车载平台; (b)前视会聚式双目视觉
图 1 实验平台与配置
Fig. 1 Experimental platform and setup

实验中, 进行闭合回路匹配后, 获得了匹配特征对, 采用最大后验概率采样一致性方法(maximum a posterior sample consensus, MAPSAC)^[9]估计立体极线约束对应的基本矩阵。为了验证方法在不同场景中的有效性, 采集了 2 类双目视频, 一类为 on-road 视频数据, 另一类为 intersection 视频数据, 其中, intersection 视频中运动障碍较多、而且运动模式复杂, 更具挑战性。图 2 所示为 on-road 视频中不同时刻图

像(右视)中的独立运动障碍目标检测跟踪结果。从图 2 可以看出: 独立运动障碍都被正确地检测与跟踪。



f/帧: (a) 17; (b) 33; (c) 67; (d) 103; (e) 266;
(f) 272; (g) 276; (h) 284

图 2 on-road 运动障碍检测
Fig. 2 Moving obstacles detection in on-road sequences

图 3 所示为 intersection 视频中不同时刻图像(右视)中的独立运动障碍目标检测跟踪结果, 为了更清楚地显示检测结果, 对结果的局部图像进行了放大。



图 3 Intersection 运动障碍检测
Fig. 3 Moving obstacles detection intersection sequences

为了定量评估运动障碍检测算法的性能, 计算检测率 x_j 和正确率 x_c :

$$x_j = N_c / (N_c + N_m), \quad x_c = N_c / (N_c + N_f) \quad (5)$$

其中, N_c 为正确检测的运动障碍数; N_m 为漏检的运动障碍数; N_f 为误检的运动障碍数。在不同场景中采集了 6 段 on-road 双目视频和 10 段 intersection 双目视频对算法进行了测试, 结果见表 1。

表 1 运动障碍检测统计结果

Table 1 Statistics results of moving obstacles detection

视频类别	图像 数/帧	正确检测 目标	漏检 目标	误检 目标	检测率	正确率
On-road	420	829	67	48	0.925 2	0.945 3
Intersection	160	406	55	42	0.880 7	0.906 3

由表 1 可以看出: interseciton 视频数据中算法的检测率和正确率都比 on-road 视频数据低。在复杂场景及动态遮挡情况下, 算法的检测率仍需要进一步提升。算法的处理时间跟场景的复杂度相关性较大, 因为不同场景中的目标数及复杂度不同会导致特征数目的不同, 算法的大部分处理时间为闭合回路特征匹配和多视极线几何基本矩阵的估计。VISATTM 数据图像原始分辨率为 $1\ 600 \times 1\ 200$, 采用一台奔四 2.66 GHz, 2 G 内存的电脑, Matlab 算法平均每 3 s 可处理 5 帧的 on-road 视频数据。算法的实时性需要进一步优化和提升。

4 结论

提出了一种基于车载双目立体相机的运动障碍检测算法, 多视极线约束能有效地解决当相机和障碍沿相同方向运动时对极几何约束无法检测独立运动障碍的问题。采用粒子滤波框架可有效地处理运动平台下多个运动障碍在多个时刻进入或离开视野的情况。实验结果证明: 提出的方法具有鲁棒性和有效性。

参考文献:

- [1] Chang Y, Gerard M, Jinman K, et al. Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(9): 1627-1641.
- [2] Torr P. Geometric motion segmentation and model selection[J]. Philosophical Transactions, 1998, 356(1740): 1321-1340.
- [3] Hartley R, Vidal R. The multibody trifocal tensor: Motion segmentation from 3 perspective views[C]//IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE Computer Society, 2004: 769-775.
- [4] Vidal R, Ma Y, Soatto S, et al. Two-view multibody structure from motion[J]. International Journal of Computer Vision, 2006, 68(1): 7-25.
- [5] Lourakis M I, Argyros A A, Orphanoudakis S C. Independent 3D motion detection using residual parallax normal flow fields[C]//International Conference on Computer Vision. Bombay: IEEE Computer Society, 1998: 1012-1017.
- [6] Sun H, Wang C, Elsheimy N. Joint detection and tracking of independently moving objects in stereo sequences using scale-invariant feature transform features and particle filter[J]. Optical Engineering, 2010, 49: 037006.
- [7] Armangue X, Salvi J. Overall view regarding fundamental matrix estimation[J]. Image and Vision Computing, 2003, 21: 205-220.
- [8] Zhang Z. Determining the epipolar geometry and its uncertainty: A review[J]. International Journal of Computer Vision, 1998, 27: 161-198.
- [9] Torr P H S. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting[J]. International Journal of Computer Vision, 2002, 50: 35-61.
- [10] Crisan D, Doucet A. A survey of convergence results on particle filtering methods for practitioners[J]. IEEE Transactions on Speech and Audio Processing, 2002, 10(3): 173-185.
- [11] El-Sheimy N, Schwarz K. Navigating urban areas by VISAT: A mobile mapping system integrating GPS/INS/Digital cameras for GIS application[J]. Navigation, 1999, 45: 275-286.

(编辑 赵俊)