

文章编号: 2095-4980(2015)06-0897-07

基于 SiftGPU 特征匹配方法的实时视觉里程计系统

张 云, 尹 露, 王雨婷, 罗 斌

(武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079)

摘 要: 采用尺度不变特征变换(SIFT)特征匹配方法对双目相机图像进行立体匹配, 同时匹配相邻两时刻的三维点, 求解运动方程进行运动估计, 得到机器人 2 个时刻坐标变换的旋转和平移参数; 使用每 2 个时刻的旋转和平移结果进行机器人的路径反演, 采用 GPU 加速 SIFT 特征提取与匹配, 实现实时的视觉里程计系统, 并采用 RANSAC 算法用于运动估计剔除误匹配点干扰。实验结果表明, 具有仿射变换较强不变性的 SIFT 特征匹配算法能够得到较为精确的路径反演结果, 采用 GPU 加速 SIFT 特征提取与匹配能达到实时的视觉定位效果。

关键词: 视觉里程计; SIFT 特征匹配; 运动估计

中图分类号: TN911.73

文献标识码: A

doi: 10.11805/TKYDA201506.0897

Real-time visual odometry system based on SiftGPU

ZHANG Yun, YIN Lu, WANG Yuting, LUO Bin

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

Abstract: The visual odometry method is adopted based on Scale Invariant Feature Transform(SIFT) to get stereo corresponding points for motion estimation between every two binocular frames. Then the rotation and translation matrix is employed to invert the moving path of the robot. SiftGPU is applied to accelerate computing so that real-time visual odometry system can be achieved. RANSAC algorithm is adopted to eliminate mismatching interference. Experimental results show that SIFT algorithm with strong scale and affine transform invariance can get more accurate path inversion results; and GPU graphics acceleration will achieve real-time visual positioning.

Key words: visual odometry; Scale Invariant Feature Transform; motion estimation

视觉里程计是仅采用移动车辆或者机器人上的相机所获得的视觉信息来实现车辆或者机器人移动的定位, 即通过移动车体或者机器人车载相机拍摄运行过程中周围场景的图像或者视频中提取出车体或者机器人运行的情况和运行环境信息对移动车体或者机器人进行实时定位。

早在 1980 年, Matthies 等人^[1]提出了视觉里程计的概念, 设计了从特征提取、特征匹配与跟踪到运动估计的理论框架。2004 年, Nister 等人^[2]设计了一种实时的视觉里程计系统, 真正意义上实现了机器人室外运动导航, 同时还提出了视觉里程计的 2 类实现途径和流程, 即单目视觉和立体视觉的方法。单目视觉是一种相对定位方式, 双目视觉信息进行运动估计则是一种绝对定位方法。尽管目前研究比较多的还是采用单目相机进行视觉定位^[3-6], 但单目视觉定位方式一方面是相对定位方式, 必须结合额外的传感器反演出比例参数才能实现绝对定位; 另一方面单目视觉定位特征匹配或者特征跟踪的鲁棒性较差, 特别是对于存在较大仿射变换的情况下, 特征匹配与跟踪的定位精确度不高。基于双目立体视觉的定位方式^[7-10]由于每个时刻需要对左右图像进行立体匹配, 并且每 2 个相邻时刻间需要对立体匹配点进行跟踪或者匹配, 所以在实时性上远不如单目视觉里程计, 这也是目前双目视觉定位乃至多相机视觉定位方式最大的缺陷, 因此目前大多实时的双目视觉里程计或者 SLAM(Simultaneous Localization And Mapping)基本上都是采用简单的特征提取和匹配算法。Harris^[11], Fast^[12], CenSurE^[13]等简单的角点提取方法在帧率较高的情况下, 连续 2 帧图像的旋转和形变较小, 经过相应的限制和优化后也得到较好的结果^[7-10], 且能满足实时的基本要求, 在视觉里程计中用得较为广泛, 但这些方法不具有尺度和仿射变换的不变性, 在相机

拍摄帧率较低、平台运动较快的情况下定位精确度会出现严重偏差。文献[14]中较为系统地比较了各种角点、特征点的提取方法、特征描述方法以及匹配方法在视觉跟踪中的性能。

本实验采用 PointGray BumblebeeX3 立体相机,校正后的输出图像帧率为每秒 10 帧,帧率较低。由于相机的镜头小,经过实际的双目相机三维重建精确度测试,有效、准确的三维重建精确度范围在 20 m 以内,所以相机拍摄的基本上是 10 m 左右的近景,针对小型机器人移动平台而言,微小的运动(特别是相机旋转)容易造成前后帧影像在视角上的较大差异,因此本文采用 SIFT^[15]特征匹配方法、SiftGPU^[16]显卡加速达到实时的计算速度。

1 提取同名点

同名点的提取主要是获得前后 2 个时刻左右 4 幅图像的匹配点,以此获得前后相邻 2 时刻的立体匹配点。具体的匹配策略方法描述如下:

- 1) 提取 2 帧双目图像 4 幅图像的 SIFT 特征点,并对 SIFT 特征点生成 SIFT 特征描述;
- 2) 匹配第 1 帧左右图 SIFT 特征点,得到立体匹配点(PL1,PR1);
- 3) 匹配第 2 帧左右图 SIFT 特征点,得到立体匹配点(PL2,PR2);
- 4) 匹配第 1 帧左相机图像和第 2 帧左相机图像 SIFT 特征点(LL1,LL2);
- 5) 找出步骤 4)中得到的第 1 帧左图匹配点 LL1 与步骤 2)中得到的第 1 帧左图匹配点 PL1 相同的特征点作为第 1 帧左图最终的匹配点;同理得到第 2 帧左图匹配点;
- 6) 根据步骤 5)中得到的左图匹配点,由步骤 2)中的匹配点对找到对应的右图匹配点;同理找到第 2 帧右图匹配点,即完成 2 帧 4 幅图的匹配过程。

由于采用 SIFT 特征匹配算法, SIFT 特征匹配非常耗时,难以达到实时的计算速度,因此采用 SiftGPU,即采用显卡加速计算来达到实时的计算速率,针对 400×300 大小的图像同名点提取可达到每秒 4 帧的速度,基本上保证实时的定位效果。

2 运动估计

提取到同名点后,采用三角测量的方法计算前后 2 帧左右图像同名点对的三维坐标,得到三维坐标后代入运动方程(式(1))中,进行求解,得出相机的 6 个自由度参数和。其中采用的像空间辅助坐标系(S-XYZ)如图 1 所示。

$$P_i = RP_i' + T \quad (1)$$

式中: P_i 表示当前帧某一点的三维坐标; P_i' 表示后一帧与之匹配的同名点三维坐标; R, T 分别表示旋转、平移矩阵。

由得到的前后 2 帧 4 幅图像中的对应匹配点,根据三角测量的方法计算同一时刻对应同名点在像空间辅助坐标系下的三维坐标点。

由图 2 中的三角形 $S1$ 与 $S2$ 的相似以及 $S1'$ 与 $S2'$ 的相似得出以下计算公式:

$$X_o = \frac{dx_1}{x_r - x_l} \quad (2)$$

$$Y_o = \frac{dy_1}{x_r - x_l} \quad (3)$$

$$Z_o = \frac{df}{x_r - x_l} \quad (4)$$

式中: $(x_l, y_l), (x_r, y_r)$ 为同一帧左右图像匹配点相对于图像中心的坐标; d 为双目相机的基线; f 为相机焦距。

得出两时刻同名点的三维坐标后代入式(1)中,其误差方程如下:

$$E(R, T) = \sum_{i=1}^N \|P_i - (RP_i' + T)\|^2 \quad (5)$$

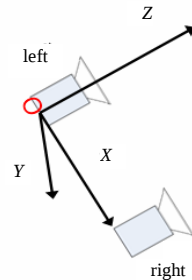


Fig.1 3D coordinate space for binocular camera
图1 双目相机三维坐标空间

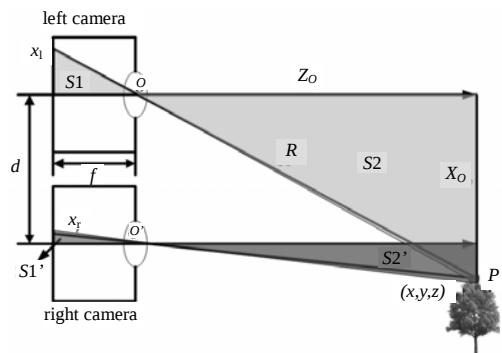


Fig.2 Diagram of parallel binocular stereo vision principle
图2 平行双目立体视觉原理图

即要求找到某一组 $\mathbf{R}$ 和 $\mathbf{T}$ 使得式(5)中的 E 的值最小, 这是参数最优化求解问题, 可通过奇异值分解的方法进行估计, 具体的求解算法在文献[17]中有详细的算法流程。但奇异值分解方法进行线性拟合求得的平移、旋转结果会受到使用的匹配点集 $P_{1,2,3,\dots,N}$ 和 $P'_{1,2,3,\dots,N}$ 中匹配点误差较大的影响, 即点集中存在匹配点误差较大的点时, 最终计算的平移旋转结果误差也很大, 线性拟合结果产生较大的偏差。所以在使用奇异值分解的方法求解运动估计时, 采用了 RANSAC 方法(随机抽样一致性方法)^[18], 每次随机选择 3 个点(求解运动方程最少需要 3 对匹配点才能得到稳定的解), 进行奇异值分解计算, 将所有点代入式(5)计算 $E(\mathbf{R}, \mathbf{T})$, 统计 $E(\mathbf{R}, \mathbf{T})$ 值小于某一阈值的点(内点)的个数, 这样经过若干次选择后, 取内点数最多的那一组结果为最终的计算结果, 在很大程度上避免了匹配误差较大的点的干扰, 提高了计算结果精确度。

3 路径反演

路径反演是通过运动估计得到相邻两时刻车体的运动(旋转和平移), 以初始位置为坐标原点, 仅考虑车体在以第一帧时刻相机的像空间辅助坐标系下, 车体在 XOZ 平面的运动情况, 不考虑垂直方向 Y 方向上的位移, 反演出平面运动轨迹。因为每次计算的是相邻 2 帧之间的平移和旋转, 是以前一帧时刻相机的像空间坐标系为标准的平移和旋转, 所以需将平移和旋转矢量转换到以起始时刻左相机的像空间坐标系为基准建立的像空间辅助坐标系中。具体的处理方法是绘制点作为矢量点, 其方向为前一帧的旋转角的累加和, 得出下一点时在当前点方向上平移 $\mathbf{T}$, 确定其坐标, 姿态是前一帧方向上乘以旋转矩阵 $\mathbf{R}$, 具体的路径反演中点的确定如式(6):

$$\begin{cases} P_1 = T_1 + P_0 \\ P_i = \prod_{k=1}^{i-1} R_k * T_i + P_{i-1}, \quad 2 \leq i \leq N \end{cases} \quad (6)$$

式中: P_0 为初始时刻相机在 XOZ 平面的位置坐标, 设为 $(0,0)$; P_i 为第 i 时刻相机在 XOZ 平面的位置坐标。

由于提取特征点精确度的限制、车体行进过程中的抖动、场景光线的变化等因素造成运动估计结果存在误差, 或者计算得出的 $\mathbf{R}, \mathbf{T}$ 存在偶然性的跳变误差, 造成最终绘制路径的误差较大, 路径不连续, 所以一般运动估计得到每相邻两帧之间的运动估计后, 一般需要对运动估计结果进行平滑处理, 通常使用到的限制条件是车体运动过程中旋转和平移速度或者加速度的连续性、限制性, 将误差较大的估计结果使用邻域均值或者中值代替, 较为复杂的可以使用卡尔曼滤波或者扩展卡尔曼滤波进行平滑, 使得到的路径连续平滑。为尽量减少整个过程的时间消耗, 使用的是前者较为简单的滤波处理, 即剔除误差较大的 $\mathbf{R}, \mathbf{T}$, 并用邻域中值代替, 平滑效果较为理想。

4 实验及结论

根据以上所阐述的实时视觉里程计系统各个关键部分的内容, 首先采用 SiftGPU 特征匹配方法, 提取前后 2 帧 4 幅图像之间公共的匹配点, 然后将同一帧左右 2 幅图像的匹配点, 根据三角测量原理转换到左相机像平面坐标系中, 得到其对应的三维坐标值。根据前后 2 帧之间相机运动造成坐标系的运动继而造成同名点三维坐标值的改变, 建立相机的运动方程, 通过第 2 节中介绍 RANSAC 方法估计每 2 个相邻时刻相机的旋转和平移位姿参数。最后对每 2 个相邻时刻的平移和旋转进行滤波处理, 减小误差干扰, 设定初始相机的位姿, 即可连接成相机运动过程中的视觉里程计, 绘制出相机的运行轨迹。

实验一与实验二均由实验室自己配备的机器人视平台获取的序列帧图像进行的实时视觉定位结果与 LibViso2^[19]视觉里程计系统的定位结果进行对比试验, 实验三所采用的是 LibViso2 网站^[20]上所提供的实验数据截取的前 1 000 帧图像序列所做出的对比试验。LibViso2 采用的是以 Sobel 边缘响应极值点为特征点, 邻域 Sobel 响应值为特征描述的点匹配方法进行的密集点立体匹配和前后帧特征点跟踪。其中实验所采用的平台是两驱履带式机器人平台, 如图 3 所示。

立体相机采用的是 PointGray 公司的 Bumblebee XB3 型号的三目相机, 图像获取只使用了两端的镜头组成的基线为 240 mm 的双目相机, 相机自带的图像校正程序直接输出已校正好的灰度图。

采用 Thinkpad T420 笔记本对机器人平台的运动进行操作, 在机器人行进过程中通过笔记本电脑实时获取相机拍



Fig.3 Robot platform
图3 机器人平台

摄的图像数据,相机的拍摄帧率约为每秒 10 帧,图像大小为 400×300 ,输出图像为校正后的图像。

此次 2 组实验场景选在武汉大学信息学部教二东向停车场,停车场路宽约为 5 m,在 GoogleEarth 上的区域截图如图 4 所示,实验一与实验二的相机拍摄的场景图像实例如图 5 所示。



Fig.4 Test site on GoogleEarth satellite map
图 4 试验场地 GoogleEarth 卫星地图



Fig.5 Example image for Experiment 1 and 2
图 5 实验一、二图像数据示例

实验一中机器人在 GoogleEarth 卫星地图上所走的 GPS 路径如图 6 所示。实时视觉里程计反演的路径结果如图 7、图 8 所示。从图中明显看出,本文的方法得出的路径与 GPS 参考路径非常相近,LibViso2 方法得到的结果偏移较大,特别是在转弯处角度的估算上存在较大的误差。实时立体视觉里程计针对 400×300 大小图像计算速度可达到每秒 4 帧,LibViso2 处理 800×600 分辨力的图像速度为每秒 4 帧。



Fig.6 GPS GroundTruth for Experiment 1
图 6 实验一机器人 GPS 参考路径图

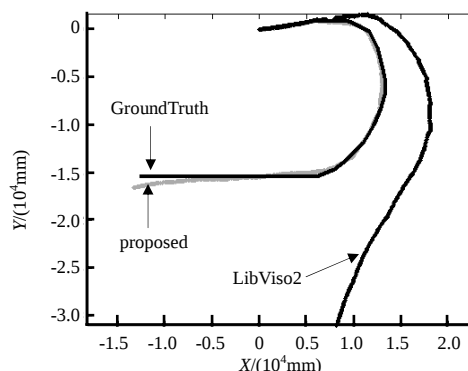


Fig.7 Comparison with GPS GroundTruth and LibViso2 results
图 7 与 GPS 参考数据以及 LibViso2 视觉里程计结果对比

第 2 个实验也是在这一场地进行,主要是测定 SIFT 特征匹配点进行运动估计的角度估计的准确性。机器人所运行的 GPS 参考路径图如图 9 所示。

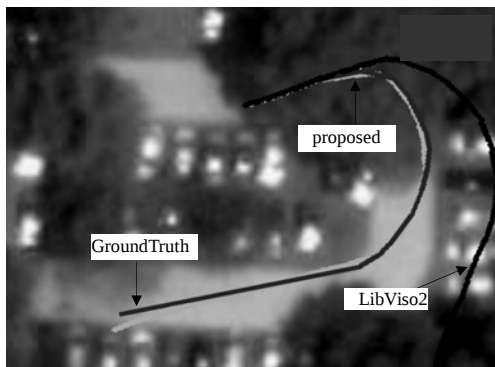


Fig.8 Comparison result on satellite map
图 8 对比路径叠加到 GoogleEarth 卫星地图



Fig.9 GPS GroundTruth for Experiment 2
图 9 实验二机器人 GPS 参考路径图

实时视觉里程计反演的路径结果如图 10、图 11 所示。从图中明显看出本文的方法得出的路径与 GPS 参考路径相近,而 LibViso2 计算结果存在较大偏差。尤其是在 2 个大角度转弯处,采用 SIFT 特征匹配估计的偏差较小,而 LibViso2 估计的误差较大,这也是 SIFT 特征匹配所带来的效果。

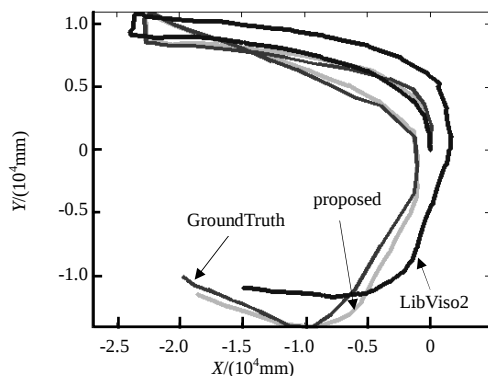


Fig.10 Comparison with GPS GroundTruth and LibViso2 results
图 10 与 GPS 参考数据以及 LibViso2 视觉里程计结果对比

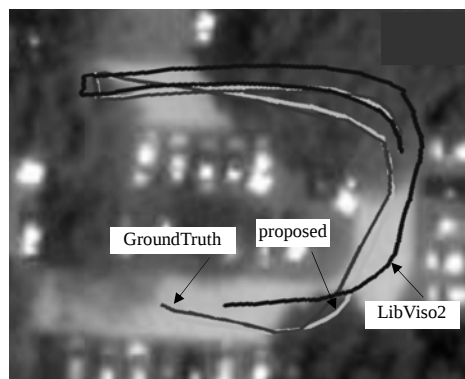


Fig.11 Comparison result on satellite map
图 11 对比路径叠加到 GoogleEarth 卫星地图

实验三采用的是 LibViso2 网站上提供的数据,图 12 为其中某一帧图像,图 13 为反演路径与 LibViso2 反演结果比较。图像帧之间的变化平缓、流畅,相机视场宽,并且是装配在汽车顶的,在车体转弯时不会造成较大的图像仿射变化,简单的特征点匹配方法也能取得较好效果。

从实验三中可明显看出,在拍摄流畅、视场宽、相机镜头大(有效双目重建距离远)的情况下,车体在转弯时造成的图像仿射变化较小,简单的不具备仿射变换不变性的特征匹配方法即可实现效果较好的视觉定位,但是具有仿射变换不变性的 SIFT 特征点匹配会获得较高的定位精确度。



Fig.12 Example image for Experiment 3(the first frame of the left camera)
图 12 实验三数据示例图像(左相机第 1 帧)

上述实验中, SIFT 特征匹配算法最终获取的视觉里程计路径与参考的 GPS 路径吻合度较高。SIFT 特征匹配算法的估计结果在各方向的位移估计结果误差均很小,在相机拍摄帧率较低而机器人平台的运动表现为骤然跳变性运动的情况下, SIFT 特征匹配算法能很好地克服因相机变化带来的拍摄图像剧烈的尺度变化和仿射变换,得到准确的特征匹配点,继而获得精确的运动估计结果,最终得到精准的视觉里程计结果。并且机器人的角度估计也是相当准确的,特别是在 2 个 90° 旋转的地方。而且在实际的测试实验中 SIFT 特征匹配算法提取的匹配点按照 RANSAC 运动估计方法,进行直线距离的估计是相当准确的,误差一般在 50 cm 以内。而使用简单的特征匹配算法虽然在计算速度上有一定的优势,但是在定位精确度上不如 SIFT 算法,特别是在低帧率、视场窄、镜头小的情况下,平台旋转带来的相机拍摄图像存在较大仿射变换的情况下, SIFT 算法具有绝对优势。

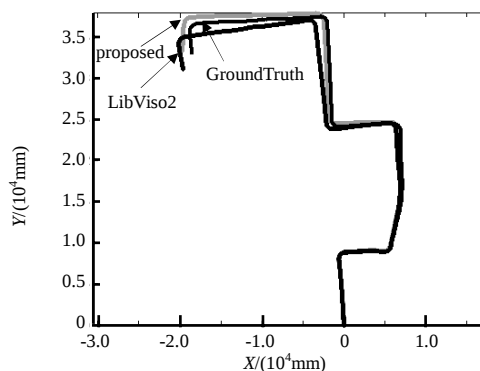


Fig.13 Comparison with LibViso2 in Experiment 3
图 13 实验三与 LibViso2 的对比试验

5 结论

本文在采用 SIFT 特征匹配算法准确获得匹配点的基础之上获得了较为准确的视觉里程计结果,并且采用

GPU 显卡加速使算法效率大大提高。通过 SiftGPU 在视觉里程计实时计算中的试验,可明显看出在拍摄帧率较低、平台运行频率较快的情况下宜采用对仿射变换具有较强不变性的特征匹配算法进行匹配,能够有效保证姿态估计的准确性。

目前视觉里程计总体上来说具有较大的不稳定性(光照条件、运动模糊等影响),可靠性较差;而惯性导航的实时性和可靠性较好,但在打滑、颠簸、松软的土地上行进时,存在较大误差,并且误差积累较为严重,致使精确度较低。以视觉导航与惯性导航相结合的方式来弥补视觉导航的实时性、可靠性的同时,采用视觉导航所获得的精确定位数据对惯性导航进行校准,将是移动平台自主导航的一个更具前景的研究方向。

参考文献:

- [1] Matthies L, Shafer S. An error modeling in stereo navigation[J]. IEEE Journal of Robotics Automation, 1987, 3(3): 239–248.
- [2] Nister D, Naroditsky O, Bergen J. Visual odometry[C]// Proc of IEEE Conference on Computer Vision and Pattern recognition(CVPR). Washington, DC, USA: [s.n.], 2004: 652–659.
- [3] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM: Why filter?[C]// Proc. IEEE Int. Conf. on Robotics and Automation. Anchorage, Alaska: [s.n.], 2010: 2657–2664.
- [4] Kneip L, Chli M, Siegwart R. Robust real-time visual odometry with a single camera and an IMU[C]// Proc. British Machine Vision Conference. Dundee, UK: [s.n.], 2011: 1–11.
- [5] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[C]// Proc. IEEE Intl. Conf. on Robotics and Automation. Hong Kong: [s.n.], 2014: 15–22.
- [6] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]// In European Conference on Computer Vision(ECCV). Zurich, Switzerland: [s.n.], 2014: 834–849.
- [7] Andrew H. Real-time stereo visual odometry for autonomous ground vehicles[C]// Intelligent Robots and Systems, 2008. IEEE/RSJ International Conference on. Nice, France: [s.n.], 2008: 3946–3952.
- [8] Johnson A E, Goldbert S B, CHENG Y, et al. Robust and efficient stereo feature tracking for visual odometry[C]// In IEEE International Conference on Robotics and Automation(ICRA). Pasadena, USA: [s.n.], 2008: 39–46.
- [9] Bernd K, Geiger A, Lategahn H. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme[C]// Intelligent Vehicles Symposium(IV), 2010 IEEE. San Diego, CA: IEEE, 2010: 486–492.
- [10] Kurt K, Agrawal M, Sola J. Large-scale visual odometry for rough terrain[C]// The 13th International Symposium Robotics Research. Hiroshima, Japan: Springer Berlin Heidelberg, 2011: 201–212.
- [11] Harris C, Stephens M. A combined corner and edge detector[C]// In Alvey Vision Conference. Manchester, UK: [s.n.], 1988: 147–151.
- [12] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]// In European Conference on Computer Vision, Graz, Austria: [s.n.], 2006: 430–443.
- [13] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features(SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346–359.
- [14] Gauglitz S, Höllerer T, Turk M. Evaluation of interest point detectors and feature descriptors for visual tracking[J]. International Journal of Computer Vision, 2011, 94(3): 335.
- [15] Agrawal M, Konolige K, Blas M. CenSurE: Center surround extremas for realtime feature detection and matching[C]// European Conference on Computer Vision 2008. [S.l.]: Springer Berlin, 2008: 102–115.
- [16] WU Changchang. SiftGPU: A GPU implementation of scale invariant feature transform(SIFT)[EB/OL]. (2011–04–19). <http://www.cs.unc.edu/~ccwu/siftgpu/>.
- [17] 吴功伟, 周文晖, 顾伟康. 基于视差空间的双目视觉里程计[J]. 传感技术学报, 2007, 20(6): 1432–1436. (WU Gongwei, ZHOU Wenhui, GU Weikang. Disparity space based binocular visual odometry[J]. Chinese Journal of Sensors and Actuators, 2007, 20(6): 1432–1436.)
- [18] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381–395.
- [19] Geiger A, Ziegler J, Stiller C. StereoScan: Dense 3D reconstruction in real-time[C]// In IEEE Intelligent Vehicles Symposium. Alcalá de Henares, Spain: [s.n.], 2011: 963–968.
- [20] Andreas Geiger[EB/OL]. (2012)[2014–10]. http://www.cvlibs.net/datasets/kitti/eval_odometry.php.