

文章编号:1004-4213(2011)08-1225-6

# 基于广义正交迭代算法的立体视觉定位

许允喜<sup>1,2</sup>, 蒋云良<sup>1</sup>, 陈方<sup>1,3</sup>

(1 湖州师范学院 信息与工程学院, 浙江 湖州 313000)

(2 浙江大学 信息与电子工程系, 杭州 310027)

(3 南京航空航天大学 导航研究中心, 南京 210016)

**摘 要:**提出了一种新的基于广义正交迭代算法的立体视觉定位. 该算法通过提取 CenSurE 局部特征和相应的 U-SURF 描述符, 采用 SAD 方法进行子像素立体匹配, 并利用 U-SURF 描述符匹配进行前后帧图像特征跟踪. 在 RANSAC 框架下对匹配点进行 3D-3D 运动估计获得了运动参量的初始值. 由于 3D-3D 运动估计使 3D 点集间欧式距离误差最小, 而 3D 特征点坐标受噪音影响很大, 因此运动估计误差也较大. 本文把广义正交迭代算法应用到立体视觉定位方法中, 得到使立体相机目标空间共线性误差最小的运动估计参量, 由于共线性误差比 3D-3D 运动估计算法中的共点误差受噪音影响更小, 从而大大减少了运动估计误差. 仿真实验和户外真实实验表明: 本文算法获得了较高的计算准确度、鲁棒性和实时性, 优于传统方法.

**关键词:**视觉定位; 广义正交迭代算法; 立体视觉; 视觉导航

**中图分类号:** TP391.41

**文献标识码:** A

**doi:** 10.3788/gzxb20114008.1225

## 0 引言

机器人自主导航是目前世界范围内人工智能、机器人学、自动控制等学科领域内的研究热点. 自主导航移动机器人在自动驾驶、军事侦察、外星球探测、抢险救灾等领域具有广泛的应用. 自主车辆在其长距离漫游过程中, 实时地获得其行驶距离、姿态等参量, 可以为智能车的定位与导航提供重要的位置和方向信息<sup>[1-2]</sup>. 对一般的自主智能车, 通常采用包含车轮编码器和惯导设备的航位推算方法获得这一信息. 这种方法的最大缺点是存在车轮滑动引起编码器读数错误及惯导漂移等问题, 尤其是在复杂战场环境中, 特别是土质疏松程度及车轮滑动不可预知的地形环境下存在较大误差<sup>[1-3]</sup>.

目前, 随着先进机器视觉算法和图像处理硬件的快速发展, 基于视觉的导航定位技术受到了人们的广泛关注, 成为陆地<sup>[4-6]</sup>、水下<sup>[7]</sup>、空中<sup>[8]</sup>自主导航的一种鲁棒且精确的定位方法.

基于视觉的定位方法, 也称视觉里程计, 仅依赖视觉输入进行运动估计, 其关键是在二维图像序列中追踪视觉特征点进行机器人自身的运动估计. 由于单目视觉定位<sup>[6]</sup>假定三维世界坐标系中的高度信

息为 0, 即假定路面平坦. 因此, 目前广泛采用立体视觉的定位方法<sup>[1,4-5,7,9-10]</sup>.

立体视觉定位通过特征提取、特征匹配、运动估计得出移动机器人运动的距离和方向信息. 目前, 绝大多数立体视觉定位都采用 3D-3D 的运动估计方法<sup>[1,4-5,7,10]</sup>, 但由于 3D 特征点坐标受噪音影响较大, 得到的运动参量不够精确.

文献[11]提出了正交迭代位姿估计算法, 具有迭代次数少、计算准确度高、快速且全局收敛的优点, 广泛应用于单摄像机位姿估计中<sup>[12-14]</sup>. 文献[15]提出了应用于多摄像机系统位姿估计的广义正交迭代算法. 本文利用广义正交迭代算法位姿估计的精确性和鲁棒性等优点, 将其应用到立体视觉定位方法中, 由于广义正交迭代算法中的共线性误差比 3D-3D 运动估计算法中的共点误差受噪音影响更小, 从而大大提高了运动估计的准确度和鲁棒性. 仿真实验和户外智能车实验表明: 本文算法是实时的、鲁棒的、精确的, 优于传统的立体视觉定位方法.

## 1 特征提取与匹配

立体视觉导航定位的关键工作之一就是得到一些稳定鲁棒的特征点, 这些特征点必须能够在因场

基金项目: 国家自然科学基金目(No. 60872057)和浙江省自然科学基金(No. R1090244, No. Y1101237, No. Y1110944)资助

第一作者: 许允喜(1978-), 男, 讲师, 博士, 主要研究方向为视觉导航、实时三维重建、智能视频监控等. Email: xuyunxi@hutc.zj.cn

收稿日期: 2011-01-18; 修回日期: 2011-04-07

景变化而引起亮度、旋转和尺度等变换时仍能保持较好的不变性,而且必须能够在立体匹配后得到较为精确的三维世界坐标系下的坐标.最常用的特征点主要是 Harris 和 SIFT 特征. Harris 角点计算简单快速<sup>[16]</sup>,但在大尺度变换的情况下, Harris 角点的可重复率(repeatability)变得很低. SIFT 特征虽然能在大尺度变换下准确地找到匹配点但却存在计算量太大,在 CPU 性能有限的情况下无法实时或快速实现的缺点.本文采用多尺度的 CenSurE 特征<sup>[17]</sup>,它结合了 Harris 和 SIFT 的各自优点,计算快速,在大尺度变换下保持良好的性能.

左右相机的立体匹配采用 SAD 方法进行子像素立体匹配,窗口尺寸为  $7 \times 7$  像素大小.采用如下二次函数对视差图进行拟合,从而达到亚像素准确度

$$d_{\text{subpixel}} = d + \frac{C_{d-1} - C_{d+1}}{2(C_{d-1} - 2C_d + C_{d+1})} \quad (1)$$

式中  $d$  为立体匹配求得的整数视差,  $d_{\text{subpixel}}$  为通过估计得到的亚像素级视差,  $C_d$  为求得的匹配代价.

由于前后帧图像特征的跟踪匹配和立体匹配不同,其搜索空间大,且前后帧运动较大时,传统的基于灰度的相似性测度匹配方法会产生许多误匹配,从而导致运动估计失败.本文采用多尺度的 U-SURF 描述符<sup>[18]</sup>进行前后帧特征匹配,该描述符在因场景变化而引起亮度、旋转和尺度等变换时仍能保持较好的不变性,计算效率和鲁棒性都较高.

## 2 基于 RANSAC 的 3D-3D 初始运动估计

运动估计是在三维欧氏空间中利用奇异值分解的方法进行求解.首先利用三角关系获取前后帧匹配点的三维坐标

$$x = \frac{B(u_1 - u_0)}{d}, y = \frac{B(v_1 - v_0)}{d}, z = \frac{Bf}{d} \quad (2)$$

$B$  为立体摄像机的基线长度,  $f$  为立体摄像机的焦距,  $d = u_1 - u_2$  为视差.

设  $X_i, Y_i$  分别为运动前和运动后的 3D 匹配点集,则智能车运动参量由式(3)给出

$$Y_j = RX_j + t \quad (3)$$

误差函数为  $\|Y_j - (RX_j + t)\|^2$ ,这是一个绝对定向问题,可由基于奇异值分解的线性最小二乘算法解决.为了进行鲁棒估计,利用 3 点 RANSAC 和 3D-3D 运动估计算法进行若干次计算,有效地去除外点,得到一个由  $n'$  个内点组成的匹配点集,并得到旋转和平移的初始值.3D-3D 运动估计算法计算

步骤为:

1) 计算点集的质心

$$\mu_x = \frac{1}{n} \sum_{i=1}^n X_i, \mu_y = \frac{1}{n} \sum_{i=1}^n Y_i \quad (4)$$

$$2) \Sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_y)(X_i - \mu_x)^T \quad (5)$$

3) 对  $\Sigma_{xy}$  进行奇异值分解,即  $\Sigma_{xy} = USV^T$

4) 计算式(6)

$$S = \begin{cases} I & \text{if, } \det(U) \cdot \det(V) = 1 \\ \text{diag}(1, 1, \dots, -1) & \text{if, } \det(U) \cdot \det(V) = -1 \end{cases} \quad (6)$$

5) 计算旋转矩阵和平移

$$R = USV^T, t = \mu_y - r\mu_x \quad (7)$$

3D-3D 运动估计算法的特征点坐标受噪音影响较大,特别在深度方向上不确定度更大,无法得到全局最小的精确运动参量.

## 3 基于广义正交迭代算法的立体视觉定位

正交迭代算法已经成为目前性能最优的实时位姿估计算法之一,被广泛应用,具有迭代次数少、计算准确度高、快速且全局收敛.文献[15]在正交迭代算法基础上推导出了多摄像机系统位姿估计的广义正交迭代算法,和常用的多摄像机位姿估计算法进行了对比,获得了更高的准确度和鲁棒性.本文利用广义正交迭代算法在位姿估计上的优点,提出了基于广义正交迭代算法的立体视觉定位.广义正交迭代算法应用在立体相机运动估计中,使立体相机目标空间共线性误差最小,共线性误差比 3D-3D 运动估计算法中的共点误差受噪音影响更小,所以得到的运动估计参量更鲁棒、更精确.

### 3.1 立体相机位姿估计的目标空间误差

立体相机和移动机器人之间为刚体,因此,在立体相机坐标系下计算出的位姿参量即为在移动机器人坐标系下的位姿参量.如图 1,立体相机的 2 个相机坐标系分别为  $O_1 X_1 Y_1 Z_1$  和  $O_2 X_2 Y_2 Z_2$ ,立体相机的基线长度为  $B$ .选定相机 1 坐标系为立体相机坐标系,则相机 2 相对于相机 1 的旋转为单位矩阵  $I$ ,平移为  $t_c = (B \ 0 \ 0)^T$ .前一帧立体相机坐标系特征点  $p$  相对于当前帧立体相机坐标系为  $q_i^1$ .  $p$  投影在当前帧立体相机 1 和 2 归一化图像平面得到的像点分别为  $v_i^1 = (u_i^1, v_i^1, 1)^T$  和  $v_i^2 = (u_i^2, v_i^2, 1)^T$ .立体相机的目标空间共线性误差如图 1.

特征点  $p_i$  的坐标  $q_i^1 = R p_i + t$ ;

$$O_1 q_i^1 \text{ 向量坐标 } q_i^1 = [q_i^1 - (0, 0, 0)^T] = q_i^1;$$

$$O_1 v_i^1 \text{ 向量坐标 } v_i^1 = [v_i^1 - (0, 0, 0)^T] = v_i^1;$$

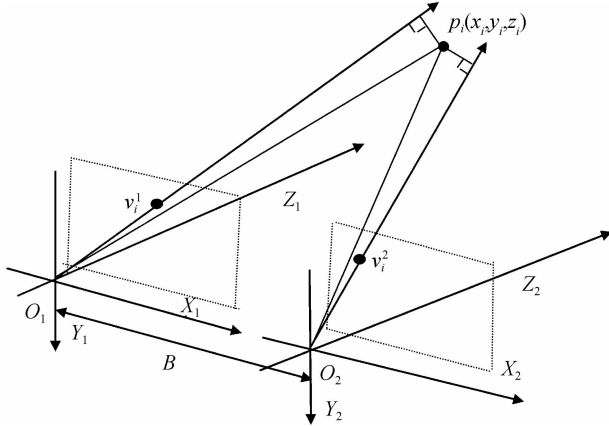


图1 立体相机位姿估计的目标空间共线性误差  
Fig. 1 Object-space errors of MCS for pose estimation

$O_2 q_i^1$  向量坐标

$$q_i^2 = q_i^1 - [I * (0, 0, 0)^T + t_c] = R p_i + t - t_c;$$

$O_2 v_i^2$  向量坐标

$$v_i^2 = (I * v_i^1 + t_c) - [I * (0, 0, 0)^T + t_c] = v_i^1;$$

则沿  $O_j v_j^i$  视线方向的投影矩阵

$$\hat{V}_i^j = \frac{v_i^j v_i^{jT}}{v_i^{jT} v_i^j} = \frac{v_i^j v_i^{jT}}{v_i^{jT} v_i^j} \quad (j=1, 2)$$

### 3.2 位姿估计算法

设有  $n$  个特征点  $p_i$ . 设  $T^j = \begin{cases} 0, j=1 \\ t_c, j=2 \end{cases}$ , 则

$$q_i^j = R p_i + t - T^j \quad (j=1, 2) \quad (8)$$

则立体相机观测到的特征点总数  $N=2 \times n$ .

算法把立体相机观测到的全部特征点的目标空间共线性误差平方和作为立体相机的目标误差函数, 则立体相机位姿估计迭代求解过程如下:

1) 相机  $j$  获取的特征点  $p_i$  的目标空间共线性误差为

$$e_i^j = (I - \hat{V}_i^j) q_i^j$$

建立目标函数

$$E(R, t) = \sum_{j=1}^2 \sum_{i=1}^n \| e_i^j \|^2 = \sum_{j=1}^2 \sum_{i=1}^n \| (I - \hat{V}_i^j) q_i^j \|^2 = \sum_{j=1}^2 \sum_{i=1}^n \| (I - \hat{V}_i^j) (R p_i + t - T^j) \|^2 \quad (9)$$

2) 给定  $R$  的值, 式(9)的目标函数关于  $t$  是二次方程,  $t$  可以由闭环形式计算

$$t^{(k)} = t(R^{(k)}) = \frac{1}{N} (I - \frac{1}{N} \sum_{j=1}^2 \sum_{i=1}^n \hat{V}_i^j) \cdot \sum_{j=1}^2 \sum_{i=1}^n (\hat{V}_i^j - I) (R^{(k)} p_i - T^j) \quad (10)$$

其中,  $k$  表示第  $k$  次迭代.

3) 由式(8)得  $q_i^{j(k)} = R^{(k)} p_i + t^{(k)} - T^j$ , 计算出  $q_i^{j(k)}$ .

4) 根据误差函数, 则

$$R^{(k+1)} = \operatorname{argmin}_R \sum_{j=1}^2 \sum_{i=1}^n \| R p_i + t - T^j - \hat{V}_i^j q_i^{j(k)} \|^2 \quad (11)$$

$$\text{设 } Q_i^j(R^{(k)}) = (\hat{V}_i^j q_i^{j(k)} + T^j)$$

$$\text{定义: } \bar{p} = \frac{1}{N} \sum_{j=1}^2 \sum_{i=1}^n p_i,$$

$$\bar{Q}_i^j(R^{(k)}) = \frac{1}{N} \sum_{j=1}^2 \sum_{i=1}^n Q_i^j(R^{(k)}),$$

$$M(R^{(k)}) = \sum_{j=1}^2 \sum_{i=1}^n Q_i^j(R^{(k)}) p_i', \quad p_i' = p_i - \bar{p},$$

$$Q_i^j(R^{(k)}) = Q_i^j(R^{(k)}) - \bar{Q}_i^j(R^{(k)})$$

$$\text{则 } R^{(k+1)} = \operatorname{argmin}_R \operatorname{tr}(R^T M(R^{(k)})) \quad (12)$$

式(12)可以使用奇异值分解(SVD)来求解.

5) 判断是否收敛, 不收敛则返回步骤 2), 否则结束, 输出位姿估计结果  $R$  和  $t$ .

### 3.3 立体视觉定位算法

本文的立体视觉定位算法计算步骤为:

1) 提取当前帧立体相机左图像的 CenSurE 特征;

2) 计算左图像 CenSurE 特征的 U-SURF 描述符;

3) 采用 SAD 区域匹配方法沿立体相机的极线进行子像素立体匹配;

4) 采用三角测量法得到特征点的 3D 坐标;

5) 利用当前帧和前一帧的 U-SURF 描述进行前后帧特征匹配跟踪;

6) 采用三点 RANSAC 的 3D-3D 运动估计算法得到内点和旋转初始值  $R_0$ ;

7) 对所有内点利用前一帧的特征点、当前帧左右图像特征点投影和广义正交迭代算法进行位姿估计. 广义正交迭代算法的旋转初始值为  $R_0$ . 位姿估计得到的旋转和平移参量即为当前帧和前一帧之间的运动参量;

8) 利用前一帧相对于第一帧的旋转和平移和当前帧的运动参量计算得到当前帧相对于第一帧的旋转和平移, 从而得到 6DOF 的 3D 定位信息.

## 4 实验

### 4.1 仿真实验

为了验证本文算法的优越性, 建立了立体视觉定位仿真实验. 实验条件为: 立体相机的等效焦距为 490, 基线长度为 0.12, 图像大小为  $640 \times 480$ . 在立体匹配和前后帧匹配中都加入高斯噪音. 实验在本文算法框架下进行, 但由于使用仿真数据, 略去了特征点检测和匹配步骤. 实验的高斯噪音方差为 0.2 像素, 行进路线为一直线, 每帧行进 0.2 m, 共行进

了 100 m, 真实行进轨迹和二种算法的估计轨迹如图 2, OI 表示广义正交迭代算法. 从图 2 可以看出, 广义正交迭代运动估计计算算法明显优于 3D-3D 运动估计计算法, 本文算法的计算准确度高, 定位误差小.

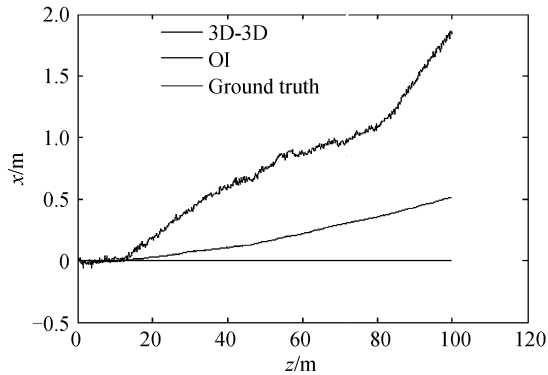


图 2 本文算法和传统方法的运动估计轨迹  
Fig. 2 Estimated motion trajectories of our algorithm and traditional method

4.2 室外真实实验

本文采用“New College Dataset” (感谢牛津大学移动机器人研究小组提供) 进行了真实智能车立体视觉定位实验. 立体相机的基线长度为 0.120 005 m, 焦距为 389.956 085, 图像大小为 512×384 像素, 立体相机相对于车体的俯仰角为 -13°. 实验采用数据集中的二个闭环路线 (New College 1 和 New College 2), 如图 3. 图 4 给出了该数据集中一组采样立体图像. 2 个数据集的实验结果如图 5 和表 1. 从图 5 和表 1 可以看出, 本文算法的定位准确度明显高于传统的 3D-3D 运动估计算法, 对二个数据集



图 3 牛津大学数据集的航拍位置  
Fig. 3 Oxford data: aerial view of location

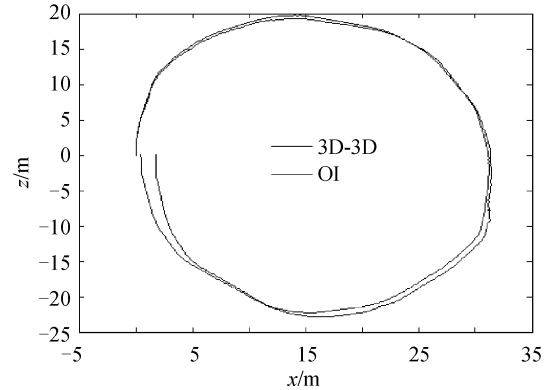


(a) Sample 1

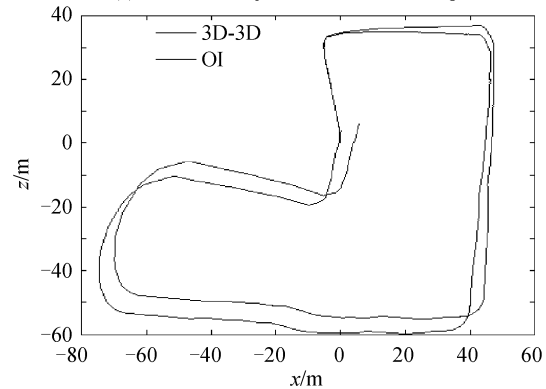


(b) Sample 2

图 4 牛津大学数据集的一对采样立体图像  
Fig. 4 Oxford data: sample stereo pair



(a) Estimated trajectories for new college 1



(b) Estimated trajectories for new college 2

图 5 本文算法和传统方法的运动估计轨迹  
Fig. 5 Estimated motion trajectories of our algorithm and traditional method

表 1 闭环误差统计

Table 1 Loop closure error statistics in meters and percent of trajectory

	Method	RMS error in XYZ	Max error in XYZ
New College 1	3D-3D	1.16(1.0%)	2.55(2.2%)
	OI	0.69(0.6%)	0.93(0.81%)
New College 2	3D-3D	6.75(1.6%)	11.41(2.8%)
	OI	4.21(1.0%)	8.15(2.0%)

表 2 本文算法各部分的运行时间

Table 2 Run-time of all components of our algorithm

	CenSurE	U-SURF	Stereo matching	Frame matching	RANSAC	OI
Run-time	12 ms	12 ms	8 ms	8 ms	10 ms	1 ms

的均方根闭环误差分别只有 0.6% 和 1.0%, 计算准确度高, 鲁棒性强。

整个算法的一帧运行时间约为 51 ms, 计算速度快, 可在线实时运行, 每个步骤的运行时间如表 2, 其中 OI 算法的运行时间仅为 1 ms。

## 5 结论

本文提出了一种新的基于广义正交迭代算法的立体视觉定位方法。本文采用广义正交迭代算法作为运动估计算法, 利用了其迭代次数少, 快速且全局收敛的优点。实验结果表明, 算法具有高准确度、鲁棒性强、实时性高的优点, 优于传统的基于 3D-3D 的运动估计算法。

### 参考文献

- [1] PENG Bo, ZHOU Wen-hui, LIU Ji-lin. Harris corner detection-based stereo visual odometry[J]. *Acta Armament Er*, 2007, **28**(12): 1498-1502.  
彭勃, 周文晖, 刘济林. 基于 Harris 角点检测的立体视觉里程计[J]. *兵工学报*, 2007, **28**(12): 1498-1502.
- [2] RODRIGUEZ F S A, FREMONT V, BONNIFAIT P. An experiment of a 3D real-time robust visual odometry for intelligent vehicles[C]. Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, MO, USA, 2009: 226 - 231.
- [3] MAIMONE M, CHENG Y, MATTHEIS L. Two years of visual odometry on the mars exploration rovers[J]. *Journal of Field Robotics*, 2007, **24**(3): 169-186.
- [4] SUNDERHAUF N, PROTZEL P. Towards using bundle adjustment for robust stereo odometry in outdoor terrain[C]. Proc. of Towards Autonomous Robotic Systems, Guildford, UK, 2006: 206-213.
- [5] JOHNSON A E, GOLDBERG S B, CHENG Y, et al. Robust and efficient stereo feature tracking for visual odometry[C]. 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA, 2008: 39-46.
- [6] SCARAMUZZA D, FRAUNDORFER F, SIEGWART R. Real-Time monocular visual odometry for on-road vehicles with 1-point RANSAC [C]. IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009.
- [7] BOTELHO S, OLIVEIRA G, DREWS P, et al. Visual odometry and mapping for underwater autonomous vehicles [C]. 2009 6th Latin American Robotics Symposium LARS, Valparaiso, 2009: 1-6.
- [8] CONTE G, DOHERTY P. Vision-based unmanned aerial vehicle navigation using geo-referenced information [J]. *EURASIP Journal on Advances in Signal Processing*, 2009, **1**(1): 1-18.
- [9] CUMANI A, GUIDUCCI A. Selecting feature detectors for accurate visual odometry [J]. *WSEAS Transactions on Circuits and Systems*, 2009, **8**(10): 822-831.
- [10] PARRA I, SOTELO M A, VIACIC L. Robust visual odometry for complex urban environments[C]. 2008 IEEE Intelligent Vehicles Symposium, Eindhoven, 2008: 440-445.
- [11] LU C P, HAGER D G, MJOLSNESS E. Fast and globally convergent pose estimation from video images [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2000, **22**(5): 610-622.
- [12] ZHANG Xiao-hu, WANG Kun-peng, ZHANG Zheng, et al. A new line-based orthogonal iteration pose estimation algorithm [C]. In Proceeding of International Conference on Information Engineering and Computer Science, Wuhan, 2009.
- [13] ZHANG Zhi-yong, ZHANG Jing, ZHU Da-yong. A fast convergent pose estimation algorithm and experiments based on vision images [J]. *Acta Aeronautica et Astronautica Sinica*, 2007, **28**(4): 943-947.  
张志勇, 张靖, 朱大勇. 一种基于视觉成像的快速收敛的位姿测量算法及实验研究[J]. *航空学报*, 2007, **28**(4): 943-947.
- [14] XU Yun-xi, JIANG Yun-liang, CHEN Fang. An extended orthogonal iteration algorithm for pose estimation from point and line segment correspondences [J]. *Acta Photonica Sinica*, 2010, **39**(10): 1881-1888.  
许允喜, 蒋云良, 陈方. 基于点线对应的多摄像机全局位姿估计迭代算法[J]. *光子学报*, 2010, **39**(10): 1881-1888.
- [15] XU Yun-xi, JIANG Yun-liang, CHEN Fang. A generalized orthogonal iterative algorithm for pose estimation of multiple camera systems[J]. *Acta Optical Sinica*, 2009, **29**(1): 72-77.  
许允喜, 蒋云良, 陈方. 多摄像机系统位姿估计的广义正交迭代算法[J]. *光学学报*, 2009, **29**(1): 72-77.
- [16] WANG A-ni, MA Cai-wen, LIU Shuang, et al. Infrared-visual image automatic registration based on corners[J]. *Acta Photonica Sinica*, 2009, **38**(12): 3328-3332.  
王阿妮, 马彩文, 刘爽等. 基于角点的红外与可见光图像自动配准方法[J]. *光子学报*, 2009, **38**(12): 3328-3332.
- [17] AGRAWAL M, KONOLIGE K. CenSurE: Center surround extremas for realtime feature detection and matching[C]. In proceeding of 10th European Conference on Computer Vision, Marseille, France, 2008: 102-115.
- [18] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up robust features [C]. In proceeding of 10th European Conference on Computer Vision, Graz, Austria, 2006: 404-417.

## Stereo Visual Localization Based on Generalized Orthogonal Iterative Algorithm

XU Yun-xi<sup>1,2</sup>, JIANG Yun-liang<sup>1</sup>, CHEN Fang<sup>1,3</sup>

(1 *Institute of Information & Engineering, Huzhou Teachers College, Huzhou, Zhejiang 31300, China*)

(2 *Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310027, China*)

(3 *Navigation Research Center, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*)

**Abstract:** A new stereo visual localization method was proposed based on generalized orthogonal iterative algorithm. Firstly, CensurE features and U-SURF descriptors were extracted, sub-pixel stereo matching were performed based on SAD method, and features between two consecutive image frames were matched using U-SURF descriptor. Then, 3D-3D motion estimation was carried out to obtain initial motion parameters in the framework of RANSAC. 3D-3D motion estimation could obtain the minimum error of Euclidean distance between 3D points. The 3D coordinates of feature points were greatly affected by noise, so the motion estimation error was large. In this paper, generalized orthogonal iterative algorithm was applied to visual stereo localization to obtain motion estimation parameters by minimising object-space collinearity error of points sensed by stereo cameras. The motion estimation error was greatly reduced because Euclidean distance error between 3D points was more affected by noise than collinearity error of points. Simulation experiment and outdoor intelligent vehicle experiment show that the proposed method can be run at real-time, and achieves a high accuracy and robustness, better than traditional methods.

**Key words:** Visual localization; Generalized orthogonal iterative algorithm; Stereo vision; Visual navigation