

激光与单目视觉融合的移动机器人运动目标跟踪

苑晶^{1,2†}, 刘钢墩¹, 孙沁璇¹

(1. 南开大学 计算机与控制工程学院, 天津 300071; 2. 天津市智能机器人技术重点实验室, 天津 300071)

摘要: 针对室内环境下的移动机器人运动目标跟踪问题, 提出一种基于激光与单目视觉传感信息融合的机器人定位和目标运动估计方法. 首先, 利用激光传感信息实现对目标的检测, 并完成机器人定位与环境建图; 然后, 设计一种基于单目视觉传感器的目标位置估计算法, 获得目标的距离和角度信息; 为了实现两类传感信息的有效融合, 将激光与单目视觉进行联合标定, 得到二者的相对位姿关系, 基于此, 将激光与单目视觉提取的目标距离和角度通过具有最优重要性函数和权重的粒子滤波器进行融合, 实现对目标运动状态的准确估计. 实验结果表明该方法具有良好的跟踪性能.

关键词: 移动机器人; 目标跟踪; 激光扫描仪; 单目摄像头

中图分类号: TP24 **文献标识码:** A

Moving target tracking of mobile robots with fusion of laser scanner and monocular camera

YUAN Jing^{1,2†}, LIU Gang-dun¹, SUN Qin-xuan¹

(1. College of Computer and Control Engineering, Nankai University, Tianjin 300071, China;
2. Tianjin Key Laboratory of Intelligent Robotics, Tianjin 300071, China)

Abstract: Moving target tracking with mobile robots in the indoor environments is addressed. A method to localize the robot and estimate the moving target's states is proposed based on information fusion of a laser scanner and a monocular camera. First, moving target detection, robot self-localization and map building of the environment are simultaneously accomplished using the laser data. Then, a monocular vision-based scheme to detect the target's location is developed for measuring the range and bearing of the target. To achieve effective fusion between two kinds of sensor information, the laser scanner and the monocular camera are jointly calibrated to obtain the relative coordinate transform between them. Finally, a particle filter (PF) with the optimal importance function and importance weights is designed to fuse the measurements of the target's range and bearing from the laser scanner and the monocular camera. The experimental results show that the proposed method can achieve satisfactory tracking performances.

Key words: mobile robots; target tracking; laser scanner; monocular camera

1 引言(Introduction)

移动机器人对运动目标的跟踪是指移动机器人通过车载传感器实现对运动目标的检测、识别和状态估计, 其研究涉及到传感信息处理、移动机器人定位、运动目标的状态估计和环境建图等内容. 在家庭、办公室等室内环境, 移动机器人与人共存, 因此对人的跟踪是移动机器人在室内环境下进行目标跟踪需要解决的主要问题, 其研究将在人一机器人的交互与协作^[1]、环境监控^[2]和以人为中心的机器人导航^[3]等场合具有广泛的应用前景.

目标跟踪要解决的主要问题是目标的检测和目标

运动状态的估计. 根据机器人车载传感器的类型不同, 常见的目标跟踪方法包括基于激光信息的跟踪^[4-10]、基于视觉信息的跟踪^[11-13]和基于多传感信息融合的跟踪^[14-21].

基于激光信息的人检测包括两类: 使用几何特征的方法^[4-6]和使用运动检测的方法^[7-8]. 前者通过识别人腿特定的模式或形状对人腿进行检测; 后者则检测环境中的动态区域, 将其视为人腿, 而环境中的静态区域可用于完成机器人定位和环境建图. 基于激光信息的人体运动状态估计常用的方法包括扩展的卡尔曼滤波(extended Kalman filter, EKF)方法^[9]和粒子滤

收稿日期: 2015-01-28; 录用日期: 2015-08-20.

†通信作者. E-mail: nkyuanjing@gmail.com; Tel.: +86 22-23508106.

本文责任编辑: 苏剑波.

国家自然科学基金资助项目(61175085, 61175083, 61573196), 天津市自然科学基金项目(15JCYBJC18800)资助.

Supported by National Natural Science Foundation of China (61175085, 61175083, 61573196) and Natural Science Foundation of Tianjin (15JCYBJC18800).

波方法^[10].

常用的利用单目视觉进行人检测和识别的方法包括: 基于脸识别技术的跟踪方法^[1,11]、基于颜色模型的跟踪方法^[12]和基于轮廓匹配的跟踪方法^[13]. 利用视觉进行人的运动估计的方法包括: 最近邻搜索方法^[1]、卡尔曼滤波方法^[11]和粒子滤波方法^[13]. 然而, 由于缺少目标的深度信息, 上述基于单目视觉的跟踪只解决了图像平面上的目标位置估计, 并未考虑笛卡尔空间中的目标状态估计问题.

为解决单一传感器存在的信息量少、鲁棒性差等问题, 多传感器融合技术越来越多地被应用到人的检测与跟踪领域. 文[15]和[16]分别使用无迹卡尔曼滤波器(UKF)和协方差交互滤波器(covariance intersection, CI)将人腿的激光数据与单目视觉获得的人脸图像进行融合. 文[17]利用激光传感器锁定环境中的动态区域, 然后在该区域对应的视觉图像上使用基于神经网络的人脸识别技术检测目标. 文[18]通过激光传感器检测人, 并利用视觉图像提取的人脸进行验证. 上述使用人脸作为视觉观测对象的方法存在的主要缺点在于跟踪过程中人必须面对摄像头. 且跟踪性能严重依赖于人脸的识别结果. 而文[19]和文[20]使用视觉图像提取人的形状信息, 并与激光提取的人腿进行融合, 实现了更鲁棒的跟踪. 近年来, RGB-D摄像头由于可以同时获得观测对象的深度信息和视觉图像信息而受到人机交互和目标跟踪领域的关注^[21], 然而由于3D点云数据进行匹配时的计算复杂度过高, 给在线跟踪带来了很大的困难.

本文面向人检测与跟踪问题提出一种鲁棒的激光与单目视觉传感信息融合方法, 在目标跟踪的过程中, 机器人利用激光提取目标的深度和角度, 同时还通过单目视觉观测人脚, 计算目标的视觉深度和角度, 并与激光信息进行融合. 论文的主要创新点包括: 1) 提出一种基于运动检测与几何特征相结合的激光人腿识别方法, 实现完备的运动目标检测和有效的人腿提取. 2) 提出一种基于单目视觉的人脚检测与位置测量方法, 通过单目视觉获得目标的深度和角度信息, 相比于现有方法(例如文[19-20]), 本文提出的方法可直接用于单目视觉跟踪系统, 因此具有更强的鲁棒性和适用性. 3) 针对本文中的激光与视觉信息融合问题, 提出一种有效的激光与视觉标定算法, 同时, 在最优粒子滤波器中设计了一种粒子多样性增强策略, 获得了准确的跟踪结果.

2 基于激光信息的人腿提取与环境建图(Laser data-based human leg extraction and map building)

机器人利用激光传感器扫描环境, 区分环境中的动态区域和静态背景. 对于静态背景, 使用扫描匹配

算法完成环境建图和机器人定位; 对于动态区域, 利用几何信息提取人腿.

2.1 运动检测(Motion detection)

本文提出一种简单且有效的基于扫描匹配的运动检测算法, 其基本思想是将所有激光扫描点顺次相连, 再分别将第1个和最后一个激光扫描点与机器人位置连接, 形成闭合多边形, 位于多边形内部的区域为自由区. 若前一帧扫描数据中的自由区在当前扫描中被覆盖, 则认为当前扫描中有运动物体进入该区域; 同理, 若前一帧扫描中被覆盖的区域在当前扫描中变成了自由区, 则认为有运动物体离开了该区域. 然而文[22]仅考虑了第1种情况, 因此无法对第2种情况下的运动物体进行检测. 为此, 本文中机器人利用连续三帧激光扫描数据 S_1 , S_2 和 S_3 进行运动检测, 将 S_2 与 S_1 进行比较以判断是否有运动物体进入 S_1 的自由区; 同时将 S_2 与 S_3 进行比较以判断是否有运动物体离开 S_2 的被覆盖区域. 对于检测出的运动物体, 使用第2.2节中的方法进行人腿提取, 而对于静态的环境背景, 使用作者之前提出的PMICP算法^[23]进行机器人定位与环境建图.

2.2 人腿提取(Human leg extraction)

基于运动检测的人腿识别方法^[7-8,22]容易将环境中其它的运动物体误识别为人腿, 而基于几何特征的人腿识别方法^[4-6]难以区分人腿和某些具有类似形状的物品(例如垃圾桶、灭火器等). 为此本文将两种方法进行融合, 在第2.1节提取出运动区域的基础上, 引入几何特征进行人腿识别.

人在行进过程中, 双腿的状态包含两种: 摆动和站立, 两种状态在双腿之间进行周期性地交替. 如果一条腿处于站立状态, 运动检测算法通常无法检测出该腿, 而只能检测出处于摆动状态的另一条腿, 由于人体的实际位置通常位于双脚中间, 若机器人仅将摆动的腿视为目标并对其进行位置估计, 则会产生较大的目标定位误差. 为了同时识别出双腿, 机器人提取激光数据的几何特征. 首先将运动区域对应的激光数据点进行分段, 若连续两个数据点之间的距离小于某一阈值(实验中设为12 cm), 则它们属于同一段, 并且同一段中的数据点个数应在一个合理范围 $[S_{i\min}, S_{i\max}]$ 内(实验中 $S_{i\min} = 4$, $S_{i\max} = 12$). 然后计算每段的宽度 W_i , 即每段第一个数据点和最后一个数据点之间的欧氏距离. 使用3个阈值 W_{\min} , W_{mid} 和 W_{\max} (实验中分别设为3 cm, 20 cm和45 cm)进行判断, 若 $W_{\min} < W_i < W_{\text{mid}}$, 则该段为单腿; 若 $W_{\text{mid}} < W_i < W_{\max}$, 则该段为双腿; 否则, 删除该段. 对于单腿, 需要寻找与之配对的另一条腿, 为此给出如下策略. 根据文献[4], 人行进过程中双脚之间的间隙应满足一定的范围 $[G_{\min}, G_{\max}]$ (实验中 $G_{\min} = 15$ cm, $G_{\max} =$

55 cm), 对每一条未配对的单腿, 将其附近的静态激光数据点进行分段, 在 $[G_{\min}, G_{\max}]$ 的范围内, 使用最近邻方法, 寻找与之距离最近的一段作为另一条腿。

提取出双腿的激光数据点之后, 计算它们的COG点(center-of-gravity point), COG点的距离和角度构成了激光对人体的观测数据 $Z_L = [r_L, \phi_L]^T$, 其模型为

$$Z_L = H_L(X_L, X_P) = \begin{bmatrix} \sqrt{[x_L - x_P]^2 + [y_L - y_P]^2} \\ a \tan 2(y_L - y_P, x_L - x_P) - \theta_L \end{bmatrix} + \begin{bmatrix} n_{r_L} \\ n_{\phi_L} \end{bmatrix}. \quad (1)$$

其中: $X_L = [x_L, y_L, \theta_L]^T$ 表示激光传感器的位置和方向角, $X_P = [x_P, y_P]^T$ 表示人体的位置, $n_L = [n_{r_L}, n_{\phi_L}]^T$ 为观测噪声。

3 视觉特征提取与目标位置测量(Visual feature extraction and people's location measure)

在常用的人脸跟踪方法中, 人必须始终面对摄像头, 因此机器人无法从背后进行跟踪。另外, 若以人脸或人的上半身为视觉目标, 环境背景会包含大量复杂的物体和干扰, 且随着目标和机器人的连续运动而产生显著变化。因此, 本文以人脚为视觉目标, 机器人可以从任意方向对人进行观测和跟踪, 同时视觉背景为地面, 其颜色、纹理的分布一致, 不会出现显著变化, 因此降低了视觉特征提取的难度, 增强了特征的稳定性和计算效率。

为了降低算法对光照的敏感程度, 本文选用HSV(hue, saturation and value)颜色表示方式, 在H-S空间建立颜色直方图。机器人使用摄像头观测人脚, 计算鞋所在的区域的颜色直方图 $\rho_h = \{\rho_h^u\}_{u=1,2,\dots,m}$, 并与事先给定的颜色参考模型 $\rho_r = \{\rho_r^u\}_{u=1,2,\dots,m}$ 进行比较, 以确定所观测的区域是否为目标区域, 其中相似度标准使用Bhattacharyya距离^[24]。若确定为目标区域, 机器人将其按 10×10 像素进行栅格化, 并提取鞋与地面的接触点CP(contact points), 如图1所示。令 $T_i^1 = [u_i^1, v_i^1]^T$ 和 $T_j^2 = [u_j^2, v_j^2]^T$ 分别表示两只鞋上的CP点在图像平面的坐标, 选取 v_i^1 和 v_j^2 最小(分别位于两只鞋最下方)的两个点作为视觉目标点, 分别记为 $T^1 = [u^1, v^1]^T$ 和 $T^2 = [u^2, v^2]^T$ 。下面给出计算和位置的方法。

图2给出了摄像头的观测原理图, 其中, $O_c-X_cY_c$ 表示摄像头坐标系, f 为焦距, γ_c 为摄像头的tilt角, O_p 为摄像头焦点在地面上的投影, U 为光轴与地面交点, $T^i (i = 1, 2)$ 为视觉目标点, T^i 在图像平面 O_I-uv 和 ${}^cX_{T^i}, {}^cY_{T^i}, {}^cZ_{T^i}$ 。假设图像平面的 u 轴平行于地面, $T_1^i(u_i^i, v_i^i)$ 为目标点 T^i 在图像平面的投影, S_1^i 为 T_1^i 在图像平面 v 轴的投影, S^i 是直线 $O_cS_1^i$ 与地面的交点。向量 O_cO_1 和 $O_cS_1^i$ 夹角为 $\alpha = \arctan(|O_1S_1^i|/f)$,

其中 $O_1S_1^i = v_{T^i} - v_0$,

$$|O_cS^i| = \frac{h_c}{\sin(\gamma_c - \alpha)}, \quad (2)$$

则

$${}^cZ_{T^i} = |O_cV^i| = \frac{h_c \cos \alpha}{\sin(\gamma_c - \alpha)}. \quad (3)$$

由摄像头针孔模型可知

$$\begin{bmatrix} {}^cX_{T^i} \\ {}^cY_{T^i} \end{bmatrix} = \begin{bmatrix} u_{T^i} \cdot {}^cZ_{T^i}/f \\ v_{T^i} \cdot {}^cZ_{T^i}/f \end{bmatrix}. \quad (4)$$

根据式(3)-(4)可以确定视觉目标点在摄像头坐标系下的位置, 进而可直接计算出目标点 T^i 对于 O_p 距离和角度。

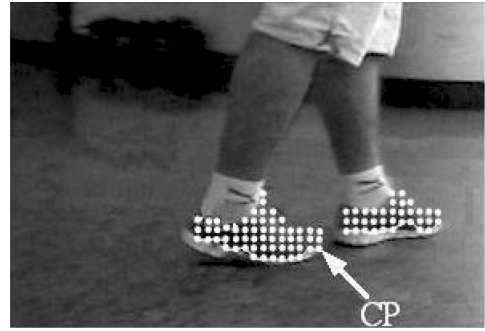


图1 鞋与地面的接触点

Fig. 1 Contact points between the shoes and the ground

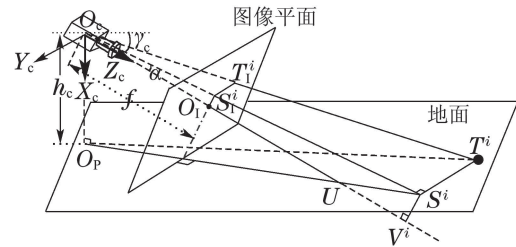


图2 摄像头的观测原理

Fig. 2 Observation principle of the camera

4 基于粒子滤波器的激光与视觉信息融合 (Fusion of laser and visual information based on particle filter)

4.1 激光与视觉的标定(Calibration of laser scanner and camera)

融合激光与单目视觉信息首先要解决的问题是二者的标定, 图3给出了系统坐标系的变换关系。将激光传感器的初始位姿定义为世界坐标系 $O_W-X_WY_WZ_W$, 将标定板竖直立于地面, 其坐标系可表示为 $O_B-X_BY_BZ_B$, 如图4所示。定义激光坐标系 $O_L-X_LY_LZ_L$ 的 X 轴和 Y 轴分别与 0° 和 90° 激光角一致, Z 轴竖直向上。简便起见, 定义其坐标原点 O_L 位于地面上, 激光坐标系相对于 $O_W-X_WY_WZ_W$ 的位姿通过PMICP算法获得。

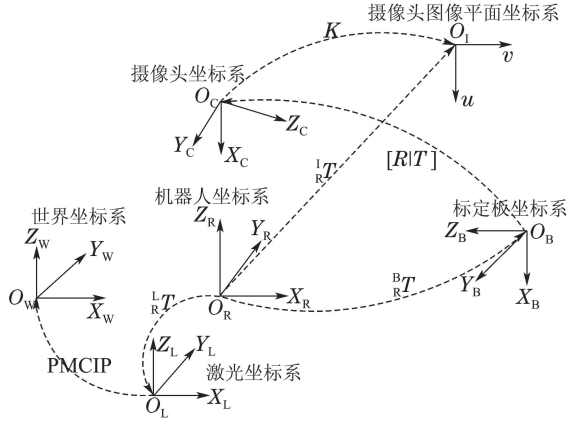


图 3 坐标系变换关系

Fig. 3 Transform among different coordinate systems

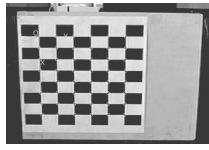


图 4 标定板

Fig. 4 Calibration board

首先将机器人面向标定板放置, 并让机器人中心近似位于平面 $O_B X_B Z_B$ 上. 定义机器人附体坐标系为 $O_R-X_R Y_R Z_R$, 其原点 O_R 位于地面上, X 轴与 $O_B-X_B Y_B Z_B$ 的 Z 轴负方向一致, Y 轴与 $O_B-X_B Y_B Z_B$ 的 Y 轴负方向一致, 而 Z 轴竖直向上. O_R 相对于 $O_B-X_B Y_B Z_B$ 的位置表示为 ${}^B r_{X_R}, {}^B r_{Y_R}, {}^B r_{Z_R}$, 因此, 从 $O_R-X_R Y_R Z_R$ 到 $O_B-X_B Y_B Z_B$ 的变换 ${}^B_R T$ 为

$${}^B_R T = \begin{bmatrix} 0 & 0 & -1 & {}^B r_{X_R} \\ 0 & -1 & 0 & {}^B r_{Y_R} \\ -1 & 0 & 0 & {}^B r_{Z_R} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

激光传感器与摄像头的标定共分3步进行.

Step 1 使用标定板标定摄像头内参矩阵 K 和外参矩阵 $[R|T]$.

Step 2 确定机器人附体坐标系 $O_R-X_R Y_R Z_R$ 到摄像头图像坐标系 O_I-uv 的变换矩阵 ${}^I_R T$.

从 $O_R-X_R Y_R Z_R$ 到摄像头坐标系 $O_c-X_c Y_c Z_c$ 的变换矩阵 ${}^C_R T$ 为

$${}^C_R T = [R|T] \cdot {}^B_R T. \quad (6)$$

则从 $O_R-X_R Y_R Z_R$ 到 O_I-uv 的变换矩阵 ${}^I_R T$ 可表示为

$${}^I_R T = K \cdot {}^C_R T = K \cdot [R|T] \cdot {}^B_R T. \quad (7)$$

Step 3 确定机器人附体坐标系与激光坐标系的变换矩阵.

本文中, 激光与视觉信息的融合在激光坐标系 $O_L-X_L Y_L Z_L$ 中完成, 为此需要计算从 $O_c-X_c Y_c Z_c$ 到 $O_L-X_L Y_L Z_L$ 的变换. 因为 ${}^L_C T = {}^L_R T \cdot {}^C_R T$, 因此需

要计算 $O_R-X_R Y_R Z_R$ 到 $O_L-X_L Y_L Z_L$ 的变换 ${}^L_R T$.

$${}^L_R T = \begin{bmatrix} \cos {}^L_R \theta & -\sin {}^L_R \theta & 0 & {}^L O_{R,X} \\ \sin {}^L_R \theta & \cos {}^L_R \theta & 0 & {}^L O_{R,Y} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

其中: ${}^L O_{R,X}, {}^L O_{R,Y}, 0$ 为 O_R 在 $O_L-X_L Y_L Z_L$ 中的坐标, ${}^L_R \theta$ 表示从 $O_R-X_R Y_R Z_R$ 到 $O_L-X_L Y_L Z_L$ 的旋转角. 根据两个坐标系的定义可知, 两个坐标系只有水平面上的平移变换和绕 Z 轴的旋转变换, 因此只需要在两个坐标系下的两组对应点, 即可确定. 为此, 机器人用激光扫描标定板, 提取到标定板的两个边界点 L_1 和 L_2 , 如图5所示. L_1 和 L_2 在 $O_L-X_L Y_L Z_L$ 中的坐标分别为 $({}^L x_{L_1}, {}^L y_{L_1})$ 和 $({}^L x_{L_2}, {}^L y_{L_2})$. 同时机器人用摄像头提取标定板的两个角点 V_1 和 V_2 在 $O_R-X_R Y_R Z_R$ 中的坐标分别为 $({}^R x_{V_1}, {}^R y_{V_1})$ 和 $({}^R x_{V_2}, {}^R y_{V_2})$, 则 L_1 和 V_1, L_2 和 V_2 组成两组对应点, 构成两个约束, 先后改变标定板的位置 q 次, 建立 $2q$ 个约束方程, 利用最小二乘法计算 ${}^L_R T$.

综上所述, 视觉目标点在激光坐标系下的位置为

$$\begin{bmatrix} {}^L X_{T^i} \\ {}^L Y_{T^i} \\ {}^L Z_{T^i} \end{bmatrix} = {}^L_C T \cdot \begin{bmatrix} {}^C X_{T^i} \\ {}^C Y_{T^i} \\ {}^C Z_{T^i} \end{bmatrix}, \quad (9)$$

则视觉目标点相对于激光坐标系的深度和角度可由式(9)直接获得.

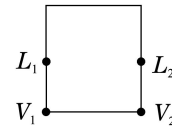


图 5 激光与机器人坐标系对应关系

Fig. 5 Correspondence between LRF and robot

4.2 激光与视觉信息的融合(Fusion of laser and visual information)

假设 t 时刻目标的状态可表示为 $X_t = [x_{P,t}, y_{P,t}, v_{P,t}, \varphi_{P,t}]^T$, 其中 $[x_{P,t}, y_{P,t}]^T$ 表示目标的位置, $v_{P,t}$ 和 $\varphi_{P,t}$ 分别表示目标的行走速度和方向角, 则目标的运动模型可表示为

$$X_t = f(X_{t-1}) + n_{P,t} = \begin{bmatrix} x_{P,t-1} + v_{P,t-1} \cos \varphi_{P,t-1} \\ y_{P,t-1} + v_{P,t-1} \sin \varphi_{P,t-1} \\ v_{P,t-1} \\ \varphi_{P,t-1} \end{bmatrix} + \begin{bmatrix} n_{P,t}^x \\ n_{P,t}^y \\ n_{P,t}^v \\ n_{P,t}^\varphi \end{bmatrix}. \quad (10)$$

其中 $n_{P,t}$ 表示均值为0, 方差为 Σ_P 的高斯运动噪声.

激光与视觉的观测信息在激光坐标系下进行融合, 激光与视觉的深度和角度观测模型为

$$Z_t = H(X_t) + n_{Z,t}, \quad (11)$$

其中: $Z_t = [Z_{L,t}^T, Z_{C,t}^T]^T$, $Z_{L,t} = [r_{L,t}, \phi_{L,t}]^T$ 为 t 时刻激光观测到的目标的距离和角度, $Z_{C,t} = [r_{C,t}, \phi_{C,t}]^T$ 为 t 时刻视觉观测到的目标的距离和角度在激光坐标系中的投影, $H = [H_L^T, H_C^T]^T$ 表示相应的激光和视觉的观测模型, $n_{Z,t}$ 表示均值为0, 方差为 Σ_Z 的高斯观测噪声. 系统完整的观测信息为 $\bar{Z}_{t+1} = [(Z_{t+1})^T, \rho_{h,t+1}]^T$, 其中 $\rho_{h,t}$ 为目标鞋的颜色直方图, 假设 Z_t 和 $\rho_{h,t}$ 条件独立, 因此

$$p(\bar{Z}_t | X_t) = p(Z_t | X_t) \cdot p(\rho_{h,t} | X_t). \quad (12)$$

式(12)表示激光和单目视觉对目标观测的概率分布由两部分组成, 其中 $p(Z_t | X_t)$ 为激光与单目视觉观

$$\begin{aligned} \tilde{w}_t &= \frac{p(X_{1:t} | Z_{1:t})}{q(X_{1:t} | Z_{1:t})} = \frac{p(X_{1:t-1} | Z_{1:t})p(X_t | X_{1:t-1}, Z_{1:t})}{q(X_{1:t-1} | Z_{1:t-1})q(X_t | X_{1:t-1}, Z_{1:t})} \stackrel{\text{Bayes}}{=} \\ &= \frac{p(X_{1:t-1} | Z_{1:t-1})p(Z_t | X_{1:t}, Z_{1:t-1})p(X_t | X_{1:t-1}, Z_{1:t-1})}{p(Z_t | Z_{1:t-1})q(X_{1:t-1} | Z_{1:t-1})q(X_t | X_{1:t-1}, Z_{1:t})} \propto \\ &= \tilde{w}_{t-1} \frac{p(Z_t | X_{1:t}, Z_{1:t-1})p(X_t | X_{1:t-1}, Z_{1:t-1})}{q(X_t | X_{1:t-1}, Z_{1:t})} \stackrel{\text{Markov}}{=} \tilde{w}_{t-1} \frac{p(Z_t | X_t)p(X_t | X_{t-1})}{q(X_t | X_{t-1}, Z_{1:t})}, \end{aligned} \quad (14)$$

其中 $p(X_t | X_{t-1})$ 和 $p(Z_t | X_t)$ 分别为目标的运动模型(10)和传感器的观测模型(11), 显然二者均服从高斯分布. 然而, 由于上述最优重要性函数 $q(X_t | X_{1:t-1}, Z_{1:t})$ 和粒子权重 \tilde{w}_t 在一般情况下没有解析解, 为此, 将观测模型进行线性化, 并使用高斯分布 $q(X_t | X_{1:t-1}, Z_{1:t}) \sim N(\mu_t, \Sigma_t)$ 对最优重要性函数进行近似^[25], 其中 μ_t 和 Σ_t 分别表示均值向量和协方差矩阵.

$$\begin{aligned} (\Sigma_t)^{-1} &= \Sigma_P^{-1} + \left[\frac{\partial H(X_t)}{\partial X_t} \Big|_{X_t=f(X_{t-1})} \right]^T \cdot \\ &\quad \Sigma_P^{-1} \left[\frac{\partial H(X_t)}{\partial X_t} \Big|_{X_t=f(X_{t-1})} \right], \quad (15) \\ \mu_t &= f(X_{t-1}) + \Sigma_t \left[\frac{\partial H(X_t)}{\partial X_t} \Big|_{X_t=f(X_{t-1})} \right]^T \cdot \\ &\quad \Sigma_Z^{-1} (Z_t - H(f(X_{t-1}))). \end{aligned} \quad (16)$$

由于机器人的观测信息除了激光和视觉获得的目标距离和角度之外, 还包含了目标鞋的颜色直方图, 且该观测与其它观测相互独立, 结合颜色直方图信息, 将最终的粒子权重更新公式扩展为

$$\tilde{w}_t = \tilde{w}_{t-1} \frac{p(Z_t | X_t)p(X_t | X_{t-1})}{q(X_t | X_{1:t-1}, Z_{1:t})} p(\rho_{h,t} | X_t), \quad (17)$$

其中 $p(\rho_{h,t} | X_t) = \frac{1}{\sqrt{2\pi}\sigma_B} \exp(-\frac{D_B^2}{2\sigma_B^2})$, D_B 为目标颜色直方图与参考颜色直方图之间的Bhattacharyya距离, σ_B 为其方差.

根据式(17), 那些同两种传感器观测的距离和角

测的目标距离和角度的概率分布, $p(\rho_{h,t} | X_t)$ 为单目视觉获得的目标鞋的颜色概率分布.

首先考虑以激光和视觉获得的目标距离和角度 Z_t 为观测的粒子滤波器. 粒子滤波器用于估计非线性非高斯系统的状态后验概率分布 $p(X_{1:t} | Z_{1:t})$ (或其边缘概率分布 $p(X_t | Z_{1:t})$), 其中, $X_{1:t} = \{X_1, X_2, \dots, X_t\}$, $Z_{1:t} = \{Z_1, Z_2, \dots, Z_t\}$. 粒子滤波器使用一组加权的粒子 $\{X_{1:t}^{(i)}, w_t^{(i)}\}_{i=1,2,\dots,N}$ 逼近 $p(X_{1:t} | Z_{1:t})$, 其最优重要性函数(importance function)^[25]为

$$q(X_t | X_{1:t-1}, Z_{1:t}) = p(X_t | X_{1:t-1}, Z_{1:t}) = \frac{p(Z_t | X_{1:t}, Z_{1:t-1})p(X_t | X_{t-1})}{p(Z_t | X_{1:t-1}, Z_{1:t-1})}. \quad (13)$$

相应的最优粒子权重为

度更接近, 且与颜色参考模型匹配程度更高的粒子将被赋予更大的权重, 这些粒子对捕捉目标的真实位置的贡献更大. 显然, 相比于单一传感信息的目标跟踪, 式(17)融合了更丰富的多源观测信息, 因此为表征粒子的重要性提供了更充足的证据, 且具有更强的鲁棒性.

在更新完粒子的权重之后, 将所有粒子的权重进行归一化

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{i=1}^N \tilde{w}_t^{(i)}}, \quad (18)$$

则目标在当前时刻的状态为

$$\hat{X}_t = \sum_{i=1}^N w_t^{(i)} X_t^{(i)}. \quad (19)$$

若有效粒子数 N_{eff} 小于事先指定的阈值 $N_{\text{threshold}}$, 则对粒子进行重采样操作^[25]. 考虑到人行走的机动性特点, 为更准确地捕捉到人的运动变化, 增强粒子的覆盖能力, 在重采样之后, 设计一种粒子多样性增强策略, 即对粒子进行扩散操作, 增大其覆盖范围, 具体如下. 首先, 计算重采样后所有粒子的位置的中心 μ 和方差 σ ; 然后确定新的采样区域 $ER = (\mu - k \cdot \sigma, \mu + k \cdot \sigma)$, 其中 $k = N_{\text{threshold}}/N_{\text{eff}}$ 表示粒子的扩散系数; 最后, 在 ER 中按均匀分布进行采样, 直至获得 N 个新的粒子.

5 实验(Experiments)

图6给出了本文实验所使用的Pioneer 3 DX移动

机器人系统, 其配备一套SICK LMS-200激光传感器和一台Canon VC-C50i PTZ摄像头. 首先进行激光与视觉标定的验证实验, 然后完成激光与视觉融合的跟踪实验, 再次, 将本文的融合算法与其他算法进行实验对比, 最后, 验证多传感器融合算法的鲁棒性.



图6 实验系统

Fig. 6 Experimental system

5.1 激光与视觉的标定实验 (Experiment of calibration between LRF and camera)

使用图4中的标定板, 按照第4.1节中的方法进行标定, 计算出摄像头坐标系到激光坐标系的变换矩阵 ${}^L_C T$.

为验证本文提出的标定算法的准确性, 进行如下实验. 将标定板先后竖直放置在8个不同的位置, 在每个位置上按图5的方式用激光提取标定板的两个边界点, 共扫描5次, 得到40组共80个激光扫描点, 记为 L_1^i 和 L_2^i ($i = 1, \dots, 40$), 它们在激光坐标系 $O_L-X_L Y_L Z_L$ 的 $X_L O_L Y_L$ 平面(即水平地面)上的投影为 ${}^L P_{L_1^i}(x_{L_1^i}, y_{L_1^i})$ 和 ${}^L P_{L_2^i}(x_{L_2^i}, y_{L_2^i})$; 同时, 机器人使用摄像头提取标定板角点, 在每个位置上拍摄5帧, 得到角点 V_1^i 和 V_2^i ($i = 1, \dots, 40$). 然后利用 ${}^L_C T$ 计算 V_1^i 和 V_2^i ($i = 1, \dots, 40$)在激光坐标系 $O_L-X_L Y_L Z_L$ 中的位置 ${}^L P_{V_1^i}(x_{V_1^i}, y_{V_1^i})$ 和 ${}^L P_{V_2^i}(x_{V_2^i}, y_{V_2^i})$. 若在 $O_L-X_L Y_L Z_L$ 中, 激光与视觉对应点的位置误差 $d_1^i = \|{}^L P_{L_1^i} - {}^L P_{V_1^i}\|_2$ 和 $d_2^i = \|{}^L P_{L_2^i} - {}^L P_{V_2^i}\|_2$ 足够小, 说明标定结果很准确. 图7给出了40组共80个位置误差组成的标定误差曲线, 其中最大误差为2.5 cm, 最小误差为0.04 cm, 平均误差为1.17 cm, 因此本文提出的标定算法具有较高的标定精度, 能够满足目标跟踪的需要.

5.2 激光与视觉融合的跟踪实验(Experiment of fusion of laser and visual information)

实验环境为两间相连的实验室, 目标行人的行走速度大约0.25 m/s, 通过手动方式控制机器人在目标的后方进行跟踪. 图8给出了一组粒子数 $N = 100$ 的实验结果, 其中虚线表示估计的目标轨迹; 实线表示通过屋顶的全景摄像头获取的实际的目标轨迹; “+”表示地图; “→”表示目标的运动方向. 可见, 跟踪过程中机器人对目标位置估计的误差始终

小于12 cm, 获得了准确的跟踪结果.

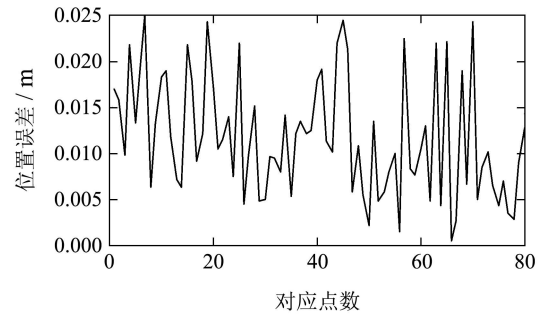
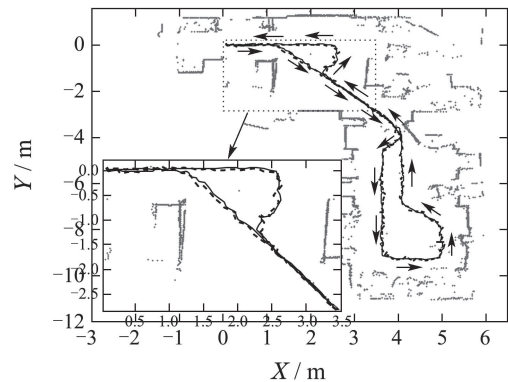
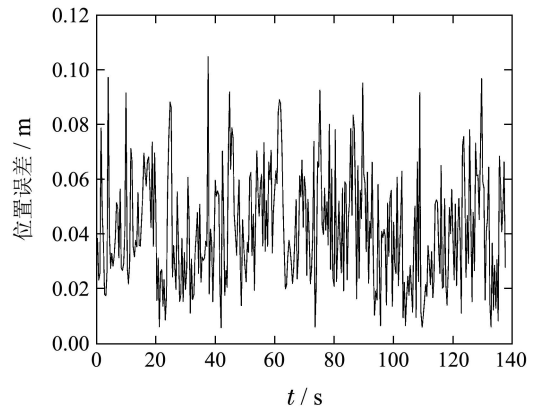


图7 激光与视觉的标定误差

Fig. 7 Errors of calibration between the LRF and the camera



(a) 目标的估计轨迹和真实轨迹



(b) 目标的位置误差

图8 激光与视觉融合的实验结果

Fig. 8 Experimental result of fusion of laser and visual information

为进一步验证两种传感信息融合的跟踪(FPF)相比于单一的激光(LPF)或视觉跟踪(VPF)的优势, 本文进行多组实验, 每组实验的粒子数目不同, 分别选为5, 10, 50, 100和200, 针对每种粒子数, 重复实验10次, 计算目标位置估计的误差和协方差, 结果见表1和表2. 由此可见, 通过两种传感信息的融合, 可以实现更准确的目标跟踪, 特别是当粒子数目很少时, FPF算法的误差均值和方差明显小于LPF和VPF, 这是因为两种传感信息融合后能够赋予同目

标实际状态更接近的粒子更高的权重,使它们能更好地表示目标状态的后验概率.此外,实验结果还表明VPF与LPF相比性能并没有显著下降,这说明所提出的基于单目视觉的目标位置测量方法具有较高的测量精度.

表 1 FPF, LPF和VPF的误差均值(单位: m)

Table 1 Mean errors of LPF, VPF and FPF

| 算法 | N | | | | |
|-----|-------|-------|-------|-------|-------|
| | 5 | 10 | 50 | 100 | 200 |
| LPF | 0.077 | 0.067 | 0.056 | 0.055 | 0.053 |
| VPF | 0.088 | 0.077 | 0.072 | 0.070 | 0.071 |
| FPF | 0.060 | 0.056 | 0.050 | 0.049 | 0.046 |

表 2 FPF, LPF和VPF的误差方差(单位: m)

Table 2 Variances of errors of LPF, VPF and FPF

| 算法 | N | | | | |
|-----|-------|-------|-------|-------|-------|
| | 5 | 10 | 50 | 100 | 200 |
| LPF | 0.071 | 0.040 | 0.033 | 0.032 | 0.032 |
| VPF | 0.061 | 0.051 | 0.043 | 0.046 | 0.042 |
| FPF | 0.039 | 0.037 | 0.029 | 0.030 | 0.028 |

5.3 不同融合算法的实验比较(Experimental comparison among different fusion algorithms)

为验证本文所提出的最优粒子滤波器的性能,这一部分将其与Bootstrap filter(BF)、EKF和CI进行实验比较,其中CI滤波器的协方差权重使用粒子群优化算法进行计算^[26],表3和表4给出了实验结果.由此可见,随着粒子数的增多,FPF滤波器的准确性最终超过了EKF和CI,并且当粒子数达到50时,FPF的性能与EKF近似,同时优于CI.与BF算法相比,5个粒子的FPF的跟踪性能优于50个粒子的BF的性能,并且随着粒子数的减少,FPF性能的下落速度明显慢于BF,说明FPF算法具有更稳定的跟踪能力.

表 3 FPF, BF, EKF和CI的误差均值(单位: m)

Table 3 Mean errors of FPF, BF, EKF and CI

| 算法 | N | | | | |
|-----|-------|-------|-------|-------|-------|
| | 5 | 10 | 50 | 100 | 200 |
| FPF | 0.060 | 0.056 | 0.050 | 0.049 | 0.046 |
| BF | 0.208 | 0.156 | 0.089 | 0.086 | 0.088 |
| EKF | | | 0.051 | | |
| CI | | | 0.063 | | |

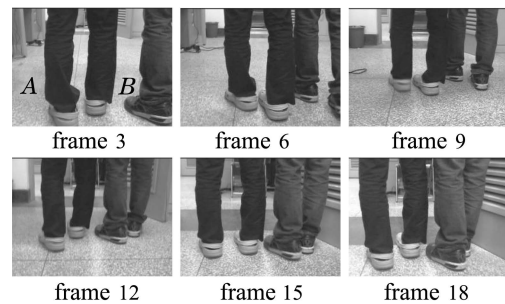
表 4 FPF, BF, EKF和CI的误差方差(单位: m)

Table 4 Variances of errors of FPF, BF, EKF and CI

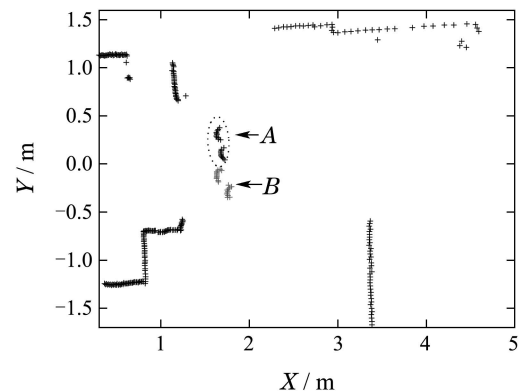
| 算法 | N | | | | |
|-----|-------|-------|-------|-------|-------|
| | 5 | 10 | 50 | 100 | 200 |
| FPF | 0.039 | 0.037 | 0.029 | 0.030 | 0.028 |
| BF | 0.125 | 0.091 | 0.052 | 0.049 | 0.050 |
| EKF | | | 0.028 | | |
| CI | | | 0.038 | | |

5.4 存在干扰时的跟踪实验(Experiment under disturbance)

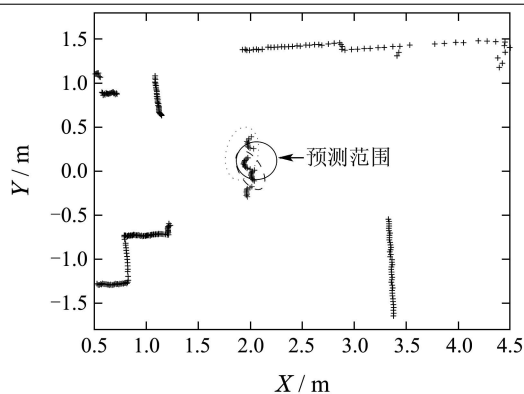
假设机器人正在跟踪目标(表示为A),此时,另一个人B出现,并在第3帧图像中从右侧进入机器人视野,随后同目标A并肩行进,如图9(a)所示.数据关联使用最近邻方法,尽管两个人行进时相距很近,但视觉特征能够很好地区分目标A与干扰B,通过融合激光和视觉的观测信息,机器人能够正确地识别并聚焦于目标A,实现对目标A的准确跟踪,跟踪误差始终小于0.08 m,如图9(d)所示.然而,若仅使用激光传感器进行跟踪,由于两个人距离过近,并且两个人腿部存在部分的相互遮挡,来自于干扰B左腿的激光数据与来自于目标A右腿的激光数据相交,如图9(b)(c)所示,机器人将这两条腿误识别为目标A的双腿,导致产生较大的跟踪误差,见图9(d).可见,基于激光与视觉融合的目标跟踪相比于单一的激光跟踪,有更强的目标识别能力和抗干扰能力.



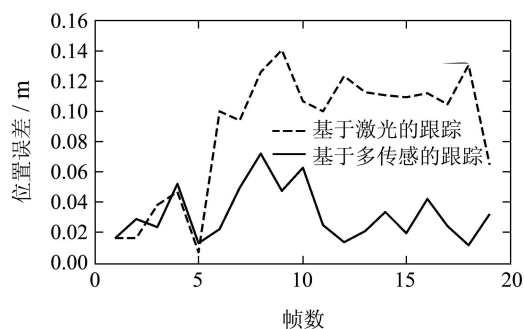
(a) 部分图像



(b) 第3帧激光数据



(c) 第6帧激光数据



(d) 目标的位置误差

图9 存在干扰时的实验结果

Fig. 9 Experimental result under disturbance

6 结论(Conclusion)

论文系统地研究了基于激光与单目视觉融合的移动机器人目标跟踪问题, 通过将运动检测和几何特征相结合, 实现了基于激光信息的运动目标检测; 同时通过观测人脚提出了一种基于单目视觉的目标位置测量方法; 然后, 利用最优粒子滤波器实现对激光信息与视觉信息的有效融合. 实验结果表明, 论文提出的跟踪算法具有很强的目标识别与跟踪能力, 同时具有很好的抗干扰能力, 并且本文设计的粒子滤波算法只需要少数目的粒子即能获得准确的跟踪结果, 这一特点对于实际应用非常重要, 因为它可以提供不同层次的跟踪精度, 并且可以根据机器人不同的计算能力和不同类型的跟踪任务, 灵活地选取粒子数目, 即使使用较少的粒子, 跟踪性能也不会有显著的下降.

参考文献(References):

- [1] SONG K T, CHEN W J. Face recognition and tracking for human-robot interaction [C] // *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. Hague, Netherlands: IEEE, 2004: 2877 – 2882.
- [2] ZHAO H, CHEN Y, SHAO X, et al. Monitoring a populated environment using single-row laser range scanners from a mobile platform [C] // *Proceedings of IEEE International Conference on Robotics and Automation*. Roma, Italy: IEEE, 2007: 4739 – 4745.
- [3] LAM C, CHOU C, CHIANG K, et al. Human-centered robot navigation-towards a harmoniously human-robot coexisting environment [J]. *IEEE Transactions on Robotics*, 2011, 27(1): 99 – 111.
- [4] CHUNG W, KIM H, YOO Y, et al. The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder [J]. *IEEE Transactions on Industrial Electronics*, 2012, 59(8): 3156 – 3166.
- [5] XU Z, FITCH R, SUKKARIEH S. Decentralised coordination of mobile robots for target tracking with learnt utility models [C] // *Proceedings of IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany: IEEE, 2013: 2014 – 2020.
- [6] XAVIER J, PACHECO M, CASTRO D, et al. Fast line, arc/circle and leg detection from laser scan data in a player driver [C] // *Proceedings of IEEE International Conference on Robotics and Automation*. Barcelona, Spain: IEEE, 2005: 3930 – 3935.
- [7] KONDAXAKIS P, BALZAKIS H, TRAHANIAS P. Learning moving objects in a multi-target tracking scenario for mobile robots that use laser range measurements [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. St. Louis, USA: IEEE, 2009: 1667 – 1672.
- [8] WANG C, THORPE C, HEBERT M, et al. Simultaneous localization, mapping and moving object tracking [J]. *International Journal of Robotics Research*, 2007, 26(9): 889 – 916.
- [9] LEE J H, TSUBOUCHI T, YAMAMOTO K, et al. People tracking using a robot in motion with laser range finder [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. Beijing, China: IEEE, 2006: 2936 – 2942.
- [10] MONTEMERLO M, THRUN S, WHITTAKER W. Conditional particle filters for simultaneous mobile robot localization and people-tracking [C] // *Proceedings of IEEE International Conference on Robotics and Automation*. Washington DC, USA: IEEE, 2002: 695 – 701.
- [11] MICHELONI C, FORESTI G L, PICIARELLI C, et al. An autonomous vehicle for video surveillance of indoor environments [J]. *IEEE Transactions on Vehicular Technology*, 2007, 56(2): 487 – 498.
- [12] ZOIDI O, TEFAS A, PITAS I. Visual object tracking based on local steering kernels and color histograms [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(5): 870 – 882.
- [13] CIELNIAK G, DUCKEET T, LILIENTHAL A J. Improved data association and occlusion handling for vision-based people tracking by mobile robots [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, USA: IEEE, 2007: 3436 – 3441.
- [14] LIAO Chuanjin, HUANG Xiyue, CHAI Yi. Target tracking and decision-making for collision avoidance based on multisensor fusion [J]. *Control Theory & Applications*, 2005, 22(1): 127 – 133. (廖传锦, 黄席越, 柴毅. 基于多传感信息融合的目标跟踪与防撞决策[J]. *控制理论与应用*, 2005, 22(1): 127 – 133.)
- [15] BELLOTTO N, HU H. Multisensor-based human detection and tracking for mobile service robots [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2009, 39(1): 167 – 181.
- [16] LUO R C, CHANG N W, LIN S C, et al. Human tracking and following using sensor fusion approach for mobile assistive companion robot [C] // *Proceedings of IEEE Annual Conference on Industrial Electronics*. Porto, Portugal: IEEE, 2009: 2235 – 2240.
- [17] BLANCO J, BURGARD W, SANZ R, et al. Fast face detection for mobile robots by integrating laser range data with vision [C] // *Proceedings of International Conference on Advanced Robotics*. Coimbra, Portugal: IEEE, 2003: 953 – 958.
- [18] KLEINEHAGENBROCK M, LANG S, FRITSCH J, et al. Person tracking with a mobile robot based on multi-modal anchoring [C] // *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication*. Berlin, Germany: IEEE, 2002: 423 – 429.

- [19] SCHULZ D. A probabilistic exemplar approach to combine laser and vision for person tracking [C] // *Proceedings of International Conference on Robotics: Science and Systems*. Philadelphia, PA, USA: IEEE, 2006: 362 – 367.
- [20] DAI W, CUHADAR A, LIU P X. Robot tracking using vision and laser sensors [C] // *Proceedings of IEEE Conference on Automation Science and Engineering*. Washington DC, USA: IEEE, 2008: 169 – 174.
- [21] GRITTI A P, TARABINI O, GUZZI J, et al. Kinect-based people detection and tracking from small-footprint ground robots [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. Chicago, USA: IEEE, 2014: 4096 – 4103.
- [22] LINDSTROM M, EKLUNDH J O. Detecting and tracking moving objects from a mobile platform using a laser rang scanner [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. Maui, USA: IEEE, 2001: 1364 – 1369.
- [23] GUO R, SUN F, YUAN J. ICP based on polar point matching with application to graph-SLAM [C] // *Proceedings of IEEE International Conference on Mechatronics Automation*. Changchun, China: IEEE, 2009: 1122 – 1127.
- [24] KAILATH T. The divergence and Bhattacharyya distance measures in signal selection [J]. *IEEE Transactions on Communication Technology*, 1967, 15(1): 52 – 60.
- [25] DOUCET A, GODSILL S, ANDRIEU C. On sequential Monte Carlo sampling methods for Bayesian filtering [J]. *Statistics and Computing*, 2000, 10(3): 197 – 208.
- [26] LUO R C, LAI C C. Enriched indoor map construction based on multi-sensor fusion approach for intelligent service robot [J]. *IEEE Transactions on Industrial Electronics*, 2012, 59(8): 3135 – 3145.

作者简介:

苑晶 (1980–), 男, 副教授, 目前研究方向为移动机器人导航与控制、目标跟踪, E-mail: nkyuanjing@gmail.com;

刘钢墩 (1990–), 男, 硕士研究生, 目前研究方向为移动机器人目标跟踪, E-mail: liugangdunai@163.com;

孙沁璇 (1990–), 女, 硕士研究生, 目前研究方向为SLAM、环境感知与识别, E-mail: sunqinxuan0122@163.com.