

面向服务机器人的多行人检测与跟踪^①

李鹏飞, 夏克付

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

摘要: 服务机器人在给人提供帮助, 带来生活便利的同时, 需要检测并跟踪行人. 然而, 环境复杂, 多个行人之间存在遮挡等问题, 给行人的检测与跟踪带来了挑战. 在行人检测方面, 本文提出了最近邻方法融合激光人腿检测和 Kinect 人体检测的结果, 有效改善了行人检测的精度和完整性. 针对多行人跟踪, 本文提出了一种改进的粒子滤波算法对行人的位置和速度进行了估计, 克服了传统粒子滤波算法计算量大, 重采样阶段粒子贫化的缺点. 最后, 在实际场景中采用改造的 turtlebot 机器人进行了测试, 并进行了计算机可视化, 实验结果证明本文提出的方法具有很好的准确性, 实时性和鲁棒性.

关键词: 服务机器人; 多行人检测与跟踪; 数据融合; Kinect; 改进的粒子滤波算法

Multi-People Detection and Tracking for Service Robots

LI Peng-Fei, XIA Ke-Fu

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Multi-people detection and tracking is a basic function for service robots when they serve people and interact with people. However, there exists a big challenge because of complex environment and occlusion of multiple people. To detect people, a new method based nearest neighbor algorithm combining leg detection of laser and body detection of Kinect is proposed and it improves the accuracy and completeness of detection. For multi-people tracking, an improved particle filter is proposed to predict the people's position and velocity, which can overcome the problem of heavy computational amount and particle degradation. In the experimental stage, a turtlebot robot with laser and Kinect is used to test the method in a real indoor environment. Results show the method can detect and track people in real time, and the method is robust and accurate.

Key words: service robots; multi-people detection and tracking; data fusion; Kinect; improved particle filter

1 引言

行人检测与跟踪是服务机器人的一项基本功能. 服务机器人需要动态地感知周围环境, 从中检测出行人, 并且进行跟踪, 预测出行人的位置和运动状态. 从而更好的与人进行交互, 例如导航过程中避开行人等. 传统的行人检测方法主要基于二维的 RGB 图像, 具体的方法有基于模板匹配的方法和基于统计特征的方法^[1-5]. 随着廉价 Kinect 传感器的问世, 大批学者开始研究基于 RGB-D 图像的行人检测, 并且将其运用在机器人上. 相比于传统的 RGB 图像, RGB-D 图像拥有

丰富的深度信息, 能够确定物体在三维空间中的位置, 更加具有实用价值. 面向服务机器人的行人检测与跟踪和传统的行人检测与跟踪有很大的不同, 主要体现在: (1) 机器人本身存在运动, 因此行人的背景是动态的; (2) 机器人获取的信息由二维的 RGB 图像变为三维的 RGB-D 图像; (3) 服务机器人长期活动的区域比较单一, 主要是室内. 这些不同点一方面给行人检测与跟踪提高了难度, 另一方面由于其特殊性也提供了新的解决思路.

很多学者研究过基于 RGB-D 图像的行人检测问

^① 基金项目:中央高校基本科研基金(WK0110000038);安徽省自然科学基金重点项目(KJ2016A050)

收稿时间:2016-01-27;收到修改稿时间:2016-03-14 [doi: 10.15888/j.cnki.csa.005385]

题^[6-13]. 文[6]基于深度信息提出了HOD(深度方向直方图)特征和Combo-HOD(HOG和HOD概率组合)特征,并用其检测行人.然而,该方法在出现行人部分遮挡时容易检测失败,并且检测需要使用GPU才能达到实时的效果.文[7]通过检测相邻行人间的头部特征改善了行人遮挡检测不准的问题,并通过点云预处理缩小了检测范围使得检测实时性大大提高.但采用Kinect对行人进行检测,当行人遮挡时仍然会存在检测不准的情况,并且对位置估计的精度偏低.

卡尔曼滤波^[14]和粒子滤波^[15]是解决跟踪问题的经典方法.卡尔曼滤波通过线性系统输入跟踪物体的位置等观测数据预测出跟踪物体的位置速度和方向,它的缺点是只适用于线性系统.针对非线性系统可以采用扩展的卡尔曼滤波^[16],它将非线性系统线性化,然后进行卡尔曼滤波估计.粒子滤波是一种更加有效的状态估计算法.它在状态空间中随机产生粒子表示概率密度,然后用粒子均值代替积分运算,从而获得系统的最优估计.不过粒子滤波在重采样阶段很容易丧失多样性造成粒子贫化,对多个目标的状态估计会出现错误.

为此,面向服务机器人的多行人检测与跟踪,本文提出了新的方法.首先,在行人检测阶段使用激光检测人腿,使用Kinect检测人体,然后利用最近邻算法进行数据融合,提高了检测的精确性和鲁棒性.其次,在行人跟踪阶段提出了SIR和MCMC相结合的分层粒子滤波算法,在保证实时性的同时有效减少了粒子贫化问题.最后,在改造后的turtlebot机器人上做了实验并进行了计算机可视化分析.

2 行人检测

行人在室内或静止或运动,可能会出现行人与物体或者行人与行人之间的遮挡.为了保持行人检测的准确性,并在有部分遮挡的情况下依然能检测到行人,本文采用了人腿检测与人体检测相结合的方法.通过融合两种方式的检测结果,使检测方法更加具有鲁棒性,图1显示了本文检测行人的框架.

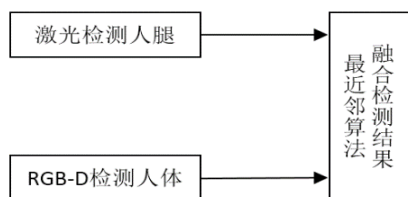


图1 行人检测框架图

2.1 基于激光簇的人腿检测

激光具有检测精度高,实时性好的特点,因此本文采用激光检测人腿.首先,需要对激光采集的点簇信息进行预处理,根据点簇的连续性将点聚集成激光线段,最终的结果是线段的集合S,即 $S=\{s_1, s_2, \dots, s_n\}$.其次,针对人腿的特殊性需要提取线段的特征用以描述人腿,本文定义了如下特征:

- 1) 线段中的点个数 Number.
- 2) 线段跨越的宽度 Width.
- 3) 线段跨越的深度 Depth.
- 4) 线段的周长 Contour.

有了线段的特征定义,就需要训练分类器,使用分类器判断检测到的激光线段是否为人腿,本文采用Adaboost算法作为分类器. Adaboost是一种迭代算法,它是针对同一个训练集训练不同的分类器(弱分类器),然后把这组弱分类器集合起来,构成一个更强的最终分类器(强分类器).本文采用Adaboost算法的主要流程如下:

1) 设置样本集 $(s_1, f_1), (s_2, f_2), \dots, (s_k, f_k)$, 其中 s_k 代表检测的样本, $f_k=+1$ 代表正样本, $f_k=-1$ 代表负样本, $k \leq N$.

2) 设置每个正样本的初始权重 $w_1(k) = \frac{1}{a}$, 其中 a 代表正样本的个数,每个负样本的初始权重 $w_1(k) = \frac{1}{b}$, 其中 b 代表负样本的个数.

3) 下面步骤循环 $t=1, \dots, T$

a) 归一权重 $w_t(k) = \frac{w_t(k)}{\sum_{i=1}^N w_t(i)}$

b) 以 w_t 的概率抽取样本训练弱分类器 c_j .

c) 计算每个分类器 c_j 的准确率 $r_j = \sum_{i=1}^N w_t(i) f_i c_j(e_i)$, 其中 $c_j \in \{+1, -1\}$.

d) 选择该次循环中使 $|r_j|$ 最大的分类器 c_j ,并设置 $(c_t, r_t) = (c_j, r_j)$.

e) 更新权重 $w_t(k) = w_{t-1}(k) \exp(a_t \cdot f_k \cdot c_{t-1}(e_k))$, 其中

$$a_{t-1} = \frac{1}{2} \log \left(\frac{1+r_{t-1}}{1-r_{t-1}} \right).$$

4) 最终的强分类器为 $C(e) = \text{sign}(F(e))$, 其中

$$F(e) = \sum_{t=1}^T a_t c_t(e) \text{ 且 } c_j(e) = \begin{cases} +1, & \text{如果 } p_j f_j(e) < p_j \theta_j \\ -1 & \end{cases}$$

θ_j 是判别的阈值.

2.2 基于 RGB-D 的行人检测

RGB-D 图像与 RGB 图像的区别在于 RGB-D 图像拥有深度信息, 可以检测出行人在空间的位置即三维坐标信息. RGB 图像可以用 (r, g, b) 来表示每个像素, RGB-D 图像可以用 (r, g, b, x, y, z) 来表示空间中的每个点. 由于拥有坐标信息, 因此对于行人的检测可以更加精确, 并可以判断行人在空间中的位置, 图 2 显示了本文采用 RGB-D 进行行人检测的流程.

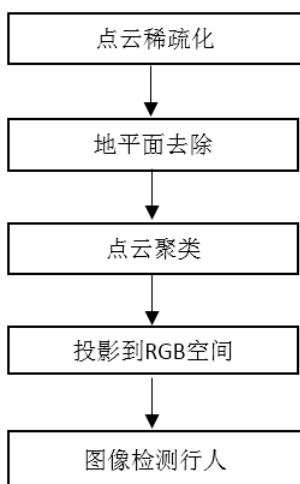


图 2 RGB-D 行人检测流程

为了充分利用 RGB-D 信息对人体进行检测, 必须对 RGB-D 图像进行预处理. 首先, 由于采集的点云数据量过大, 在进行检测时时间复杂度太高, 因此需要对点云进行稀疏化. 本文采用了 VoxelGrid 算法对点云进行下采样, 它通过创建三维体素格子, 计算每个体素格子中所有点的重心来近似一个体素的位置. 处理后的点云数据, 就是所有体素格子重心组成的点集. 下一步就是要去除地面对检测的影响, 因为地面包含的点云数据量大, 而且对检测人体有干扰作用. 本文采用 RANSAC 随机抽样一致性算法进行地面检测, 并予以剔除. RANSAC 算法通过建立地面的平面模型, 多次迭代抽样点云调整模型参数从而得出地面的模型方程式, 对于满足地面模型方程式的点将其删除. 通过预处理, 点云的数量减少了很多, 然后需要对点云进行聚类, 将描述一个物体的点云聚集到一起. 本文采用欧几里得距离作为聚类标准, 只要属于设定阈值范围内的相邻点即可划分为同一个聚类. 由于聚类体处于三维空间, 为了判别是否为行人, 需要将聚类体投影到 RGB 图像空间. 本文采用 HOG+SVM 的方法

对投影到 RGB 图像空间的候选集进行检测与判别. HOG(方向梯度直方图)特征是通过图像进行灰度化、归一化、计算像素梯度、划分 cell 和 block, 将梯度串联起来的局部特征, 它描述了待检测物体的轮廓信息. SVM(支持向量机)是一个性能良好的分类器, 它通过最大化间隔面使得分类的结构化风险最小, 在小样本、非线性和高维模式识别中表现优异. 本文即使用 HOG 特征训练 SVM, 并使用 SVM 判别人体.

2.3 检测结果融合

由激光检测出来的行人集合 P_1 和由 RGB-D 检测出来的行人集合 P_2 需要进行数据融合. 假设 $|P_1|=m$, $|P_2|=n$, 可以形成 $m*n$ 的距离矩阵 D , 距离矩阵中的每个元素 d_{ij} 表示集合 P_1 中的第 i 个元素与集合 P_2 中的第 j 个元素的距离. 采用最近邻的方法合并距离最近的两个检测体 P_1 和 P_2 , P_1 代表激光检测的行人状态, P_2 代表 RGB-D 检测的行人状态, 依次循环即可融合所有相关的行人状态. 在融合阶段, 采用不同的权重可以调整激光和 RGB-D 信息检测行人的置信度, 从而有所偏重.

3 行人跟踪

粒子滤波算法是一种基于贝叶斯估计的蒙特卡洛算法, 适用于非线性系统的状态估计, 在跟踪领域得到了广泛运用. 系统的状态方程和观测方程如下:

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}) \\ z_k &= h(x_k, v_k) \end{aligned}$$

其中 x_k 为系统 k 时刻的状态向量, f 为系统的状态转移函数. z_k 为系统 k 时刻的观测值, h 为系统的测量函数. u_k 和 v_k 分别为系统的过程噪声和测量噪声. 贝叶斯估计是利用先验概率分布计算后验概率分布, 系统的状态预测方程和状态更新方程如下:

$$\begin{aligned} p(x_k | z_{1:k-1}) &= \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1} \\ p(x_k | z_{1:k}) &= \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \end{aligned}$$

蒙特卡洛算法使用随机采样的方法将积分运算转换为样本点的求和运算, 构造了一个基于样本的后验概率密度函数. 使用由 N 个粒子构成的集合 $\{x_{0:k}^i, w_k^i\}_{i=1}^N$ 描述后验概率 $p(x_{0:k} | z_{1:k})$, 则 k 时刻的后验概率可以表示为

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^N w_k^i \delta(x_{0:k} - x_{0:k}^i)$$

基本的 SIS(序贯重要性采样)粒子滤波经过多次迭代后使得小部分粒子具有很大的权值,大部分粒子的权值很小,造成大量的计算集中在对估计作用很小的大部分小权值粒子上,形成了粒子退化问题. SIR(重要性重采样)算法是粒子滤波算法的一种,它的重采样过程通过复制权值大的粒子,抛弃权值小的粒子,提高了计算效率,克服了粒子退化问题,其流程如下:

1) 初始化粒子. 在状态空间中抽取 N 个粒子, 每个粒子的权重初始值设置为 $\frac{1}{N}$, 得到粒子集 $\left\{x_0^i, \frac{1}{N}\right\}_{i=1}^N$.

2) 循环 N 次, $i=1,2,\dots,N$

a) 根据状态转移函数产生新粒子 $x_t^i \sim p(x_t|x_{t-1}^i)$

b) 根据观测值更新权值, $w_t^i = w_{t-1}^i p(z_t|x_t^i)$

3) 权值归一化, $w_t^i = \frac{w_t^i}{\sum_{i=1}^N w_t^i}$.

4) 重采样. 在粒子集 $\{x_t^i, w_t^i\}_{i=1}^N$ 中采样 N 次生成新的粒子集 $\{x_t^{i'}\}_{i=1}^N$, 满足 $p(x_t^i = x_t^{i'}) = w_t^i$, 令新粒子集中每个粒子的权值为 $\frac{1}{N}$.

5) 状态估计. t 时刻的状态 $x_t = \sum_{i=1}^N w_t^i x_t^i$, 计算 $t+1$ 时刻的状态重新返回步骤 2.

然而 SIR 粒子滤波也引入了另一个问题, 那就是粒子的多样性减少, 造成粒子贫化, 最终使得所有粒子都集中在一个样本点上. MCMC(马尔科夫链蒙特卡洛)粒子滤波是一种能有效解决粒子贫化问题的算法. MCMC(马尔科夫链蒙特卡洛)构造了一条非周期、不可约的马尔科夫链 (x_0, x_1, \dots, x_n) , 并且它有不变分布 π , 当这一马尔科夫链迭代次数接近无穷时, 得到的 x_n 的分布可以无限接近 π . 利用这种方法我们可以找到一条以 P 为不变分布的马尔科夫链, 进而可以得到近似服从 P 的随机样本. MCMC 算法的一种经典实现形式是 Metropolis-Hastings 算法, 其流程如下:

1) 任意选取马尔科夫链的初始状态.

2) 从重要性分布中产生备选点.

3) 由均匀分布 $U(0,1)$ 中产生一个参考分布, 即 $v \sim U(0,1)$.

4) 判断是否接受移动. 如果 $v \leq \min\left\{1, \frac{p(z_k|x_k^i)}{p(z_k|x_k^j)}\right\}$

则移动 $x_t^j = x_t^i$, 否则不移动 $x_t^j = x_{t-1}^j$.

MCMC 粒子滤波是一种弱重采样算法, 其粒子收敛到后验概率的速度较慢, 克服了 SIR 粒子滤波算法粒子贫化的问题, 将粒子推向先验概率密度和后验概率密度都较大的区域. 然而 MCMC 算法中粒子采样次数过多, 影响跟踪系统的实时性. 为此, 本文采用 SIR 和 MCMC 相结合的分层粒子滤波算法. 在 SIR 粒子滤波算法的重采样阶段, $NP(0 < P < 1)$ 个粒子采用 SIR 算法根据上阶段粒子的权重采样生成. 然后针对剩余的 $N(1-P)$ 个粒子, 先在低分辨率空间采用 MCMC 粒子滤波算法生成部分粒子, 然后在高分辨率空间围绕生成的粒子使用 SIR 算法生成 $N(1-P)$ 个新粒子.

4 实验结果与分析

为了验证系统的有效性, 本文在改造后的 turtlebot 机器人上做了实验测试. 改造后的 turtlebot 机器人如图 3 所示, Kinect 传感器通过圆柱架高到 1.290m, Hokuyo 激光器安装在距离地面 0.215m 的底盘前方. 改造后的 turtlebot 机器人可以通过激光器检测人腿, 通过 Kinect 检测行人的躯体. 系统运行的环境是 Intel® Core™ i5-4200M CPU@2.50GHz×4 的笔记本电脑.



图 3 改造后的 turtlebot 机器人

首先对激光检测人腿模块进行了测试, 采用 URG-04LX-UG01 激光雷达, 它的测量距离范围是 0.02~5.6m, 扫描角度 240° , 扫描时间 100ms/scan. turtlebot 机器人和行人在实验场地中随意运动, 激光扫描出周围环境的 2D 信息. 通过前面叙述的检测算法, 环境的 2D 信息通过聚类被分割成不同的激光段, 提取激光段的特征, 采用 Adaboost 算法判别激光段是否是人腿. 在实验场地, 用激光采集了 567 帧包含行

人的正样本, 3218 帧不包含行人的负样本. 表 1 分别记录了有无行人情况下激光检测人腿模块的准确率.

表 1 激光检测人腿准确率表

实际情况	激光检测结果		总数
	检测为正样本	检测为负样本	
正样本	559 (98.24%)	8 (1.76%)	567
负样本	32 (0.99%)	3186(99.01%)	3218

图 4 是检测单个行人时的可视化, 其中红色的点和线段是激光对环境的扫描, 红绿蓝组成的坐标架是 turtlebot 机器人的位置, 被蓝色圆环圈住的两段红色激光是系统检测出来的人腿.

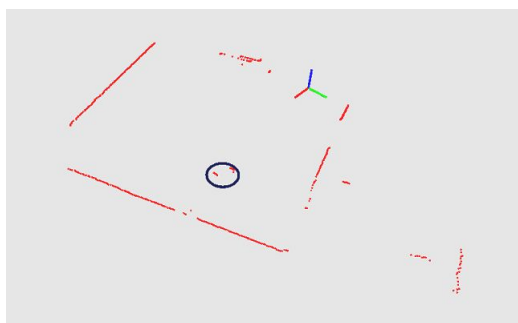


图 4 激光检测人腿效果图

其次对 RGB-D 检测人体模块进行测试, 采用微软 Kinect, 它能以每秒 30 帧的速率传送 640*480 像素的深度图和 640*480 像素的 RGB 图像, 传感深度范围是 1.2~3.5m, 主要在室内使用, 图 5 中(a)和(b)显示了使用 Kinect 采集的 RGB 图像和深度图像. 使用 RGB-D 图像对行人进行检测, 首先必须对点云数据下采样, 减少点云数量. 图 5(c)是原始的点云图像, (d)是使用 VoxelGrid 算法进行下采样后的点云图像. 针对点云数据, 接下来将经历地平面删除、点云聚类、点云投影、SVM+HOG 行人检测等阶段. 图 6 是在实验环境下检测到人体的效果图, 图像是 RGB-D 信息格式, 其中红色长方体将检测出的行人框出.

最后, 将激光检测出的行人与 RGB-D 信息检测的行人通过距离矩阵的形式采用最近邻算法数据融合得到完整精确的行人信息. 既可以检测行人躯体被遮挡的情况, 也可以检测行人腿部被遮挡的情况. 并且相比于图像信息, 激光的精读更高, 对行人状态的估计更加精确. 针对检测出的行人, 采用 SIR 粒子滤波和 MCMC 粒子滤波相结合的分层粒子滤波方法进行跟踪. SIR 粒子滤波保证了跟踪的实时性和准确性,

MCMC 粒子滤波保证了跟踪的多样性. 图 7 针对跟踪的行人进行计算机可视化, 场景中有机器人和三个行人, 显示了每个人运动的速度朝向和轨迹.

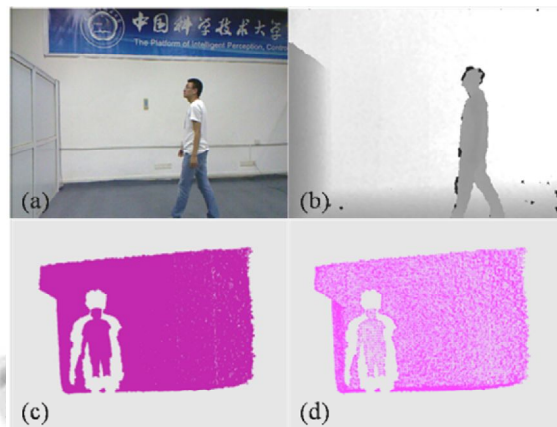


图 5 RGB-D 信息分解



图 6 RGB-D 检测人体效果图

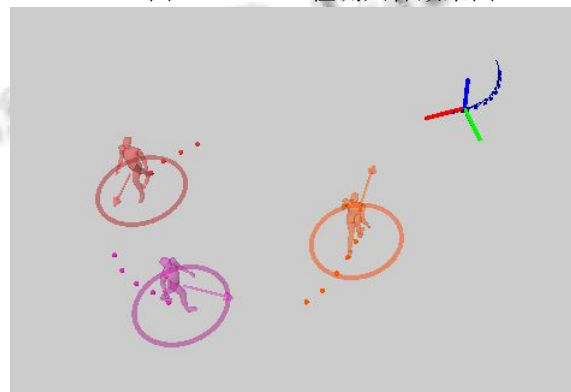


图 7 多行人跟踪效果图

本文为了验证系统的实时性, 在实验场地通过改变检测跟踪行人的个数进行了 4 组实验, 每组跟踪时长为 5 分钟, 最后得到了系统的平均运行时间如表 2 所示. 通过计算可知系统的平均运行时间为 68ms, 满足实时性要求. 分析可知, 系统的主要耗时集中在

HOG 特征的提取和粒子滤波的计算方面, 行人数量的增多会增加运行时间, 同时需要考虑环境复杂性问题造成的影响。

表 2 系统运行时间

行人个数	平均运行时间
1	63ms
2	70ms
3	67ms
4	73ms

5 总结与展望

本文设计了适用于服务机器人在室内环境检测跟踪多行人的实时系统, 提出了基于激光与 RGB-D 信息融合的检测方法. 通过自定义的人腿特征, 采用 Adaboost 算法对激光段进行人腿判别. 针对 RGB-D 点云数据, 通过稀疏下采样、去地平面、聚类、RGB 投影、HOG+SVM 方法对点云簇进行躯体判别. 行人跟踪方面, 分析了基本粒子滤波、SIR 粒子滤波的优势与不足, 提出了 SIR 和 MCMC 相结合的分层粒子滤波算法, 充分利用了 SIR 粒子滤波跟踪速率快, 跟踪精度高和 MCMC 粒子滤波在重采样阶段粒子丰富的特点. 在现实场景中, 本文通过改造的装载有 Hokuyo 激光和 Kinect 的 turtlebot 机器人进行了实验, 分别测试激光检测人腿, RGB-D 检测躯体, 信息融合检测行人以及跟踪的效果, 实验表明本文设计的系统实时性高, 精确度高, 同时具有鲁棒性. 下一步工作将融入多机器人系统, 在环境中同时运行多个机器人, 并将它们的检测结果予以融合.

参考文献

- Pedersoli M, Gonzalez J, Hu X, et al. Toward real-time pedestrian detection based on a deformable template model. *IEEE Trans. on Intelligent Transportation Systems*, 2014, 15(1): 355–364.
- Soudani A, Zagrouba E. Statistical prior based deformable models for people detection and tracking. *Neural Information Processing*. Springer International Publishing, 2015: 392–401.
- Liu H, Xu T, Wang X, et al. A novel multi-feature descriptor for human detection using cascaded classifiers in static images. *Journal of Signal Processing Systems*, 2015, 81(3): 377–388.
- Dollár P, Wojek C, Schiele B, et al. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 2012, 34(4): 743–761.
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision & Pattern Recognition*. 2005. 886–893.
- Spinello L, Arras KO. People detection in RGB-D data. 2011 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2011. 3838–3843.
- Munaro M, Basso F, Menegatti E. Tracking people within groups with RGB-D data. *IEEE/RSJ International Conference on Intelligent Robots & Systems*. 2012. 2101–2107.
- Munaro M, Menegatti E. Fast RGB-D people tracking for service robots. *Autonomous Robots*, 2014, 37(3): 227–242.
- Luber M, Spinello L, Arras KO. People tracking in RGB-D data with on-line boosted target models. 2011 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2011. 3844–3849.
- 王宁波. 基于 RGB-D 的行人检测[硕士学位论文]. 杭州: 浙江大学, 2013.
- 耿波亭. 基于 Kinect 的人体检测和跟踪算法研究[硕士学位论文]. 杭州: 杭州电子科技大学, 2012.
- Jafari OH, Mitzel D, Leibe B. Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras. 2014 *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014. 5636–5643.
- 许素萍, 苏松志. 深度图像下基于特征学习的人体检测方法研究[硕士学位论文]. 厦门: 厦门大学, 2014.
- Ligorio G, Sabatini AM. A novel Kalman filter for human motion tracking with an inertial-based dynamic inclinometer. *IEEE Trans. on Biomedical Engineering*, 2015, 62(8): 2033–2043.
- Arulampalam MS, Maskell S, Gordon N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. on Signal Processing*, 2002, 50(2): 174–188.
- Daum FE. Extended Kalman filters. *Encyclopedia of Systems and Control*, 2015: 411–413.