

基于 Stixel-world 及特征融合的双目 立体视觉行人检测*

白中浩,王鹏辉,李智强

(湖南大学 汽车车身先进设计制造国家重点实验室 长沙 410082)

摘要:针对单目视觉行人检测无法获得深度信息从而导致冗余信息较多、检测效率和准确度存在局限性的问题,首先,在图像的预处理阶段提出了一种利用双目立体视觉产生的视差信息优化分析来简化复杂场景的动态规划棒状像素场景(stixel-world)表达方式;然后,在行人目标检测阶段,对传统 HOG 特征中 block 尺度进行分析、降维,采用 Fisher 准则筛选得到了适用于道路环境下的多尺度 HOG(multi-HOG)特征,将 Multi-HOG 特征与 LUV 颜色通道特征进行融合,最后采用交叉核支持向量机(hik-SVM)分类器对行人目标分类。实验结果表明,采用改进过后的 Stixel-world 算法用于图像预处理极大地减少了计算时间。缩小了行人检测的候选区域,基于特征融合和 hik-SVM 的目标检测算法在保证检测准确度的前提下,具有较好的实时性和鲁棒性。

关键词: 行人检测;双目立体视觉;Stixel-world;特征融合;交叉核支持向量机

中图分类号: TH701 **文献标识码:** A **国家标准学科分类代码:** 510.40

Stereo visual pedestrian detection based on Stixel-world and feature fusion

Bai Zhonghao, Wang Penghui, Li Zhiqiang

(State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, China)

Abstract: The monocular vision pedestrian detection can not obtain the depth information, so detection efficiency and accuracy are limited. Firstly, in the image pre-processing, the disparity information optimization analysis is proposed to simplify the expression of dynamic programming stixel-world in complex scenes based on stereo vision. Then, at the stage of pedestrian target detection, this paper analyzes the influence of the block scale on the detection effect in the traditional HOG feature, and obtains the multi-HOG feature which is suitable for the road environment using the fisher criterion. The multi-HOG feature is integrated with the LUV color channel feature. Finally, the hik-SVM is used for pedestrian target classification. The experimental results show that the improved Stixel-world algorithm for image preprocessing greatly reduces the computation time and reduces the candidate region of pedestrian detection, target detection algorithm based on feature fusion and hik-SVM has good real-time and robustness under the premise of guaranteeing the detection accuracy.

Keywords: pedestrian detection; stereo vision; Stixel-world; feature fusion; hik-support vector machine (hik-SVM)

0 引 言

行人是城市交通的主要参与者和弱势群体,保障行人的安全也就成为了建设城市交通系统的首要目标。2016年全球道路交通事故中,行人死亡人数占比为22%^[1]。中国国家统计局的数据表明^[2],我国道路交通

事故现状特点为行人事故占比高(20%)、行人事故伤亡人数占比高(45%)。因此对交通环境中行人群体的保护至关重要。在车辆驾驶辅助系统中,行人检测作为保障行人、车辆安全的主动式安全辅助系统,已经成为学术界和产业界共同的研究热点,行人检测的目的在于建立智能、自主的驾驶辅助系统,具有保障行人及驾驶员安全的实用意义。

目前国内外针对传统可见光图像中的行人检测技术研究成果颇丰。早期的行人检测技术主要是基于形状模板和轮廓匹配,通过构造大量模板的方式在图像中逐级搜索^[3],进而识别出行人,但是这种方法检测速度较慢,且对于遮挡情况下的行人检测鲁棒性较差。为了提高行人检测的准确度,Dalal N 等人^[4]提出了一种描述图像局部梯度幅值和方向的梯度方向直方图特征(histograms of oriented gradients, HOG),HOG 特征能准确表达出行人目标在图像中明显的轮廓特征,因此是目前使用最为普遍的行人特征描述子;另外还有基于稀疏表示^[5]、深度学习^[6-8]以及其他特征^[9-10]的行人检测算法,都一定程度上达到了单一图像中行人检测的目的。

但是,基于单目视觉的行人检测无法提供深度信息,大量冗余背景也被计算在内,复杂的背景不仅会加大计算量,还会导致误检,如何准确区分行人和背景提高检测效果成为一大难题。近年来双目立体视觉系统在机器人领域、监控领域和视觉测量领域的高速发展引起了学者对其应用在汽车主动驾驶环境中的思考和探索^[11-14],双目立体视觉系统可以利用距离信息供驾驶辅助系统进行安全决策,还可以剔除对检测意义不大的背景区域得到行人的潜在位置和尺度,从而简化图像的理解和分析。

基于深度图的双目立体视觉算法信息量大、复杂程度高,交通环境中的行人检测相比于其他环境对实时性格外重视,因此目前阶段的双目立体视觉行人检测算法应用起来还有一定的局限性。本文提出的基于双目立体视觉的交通场景中的行人检测,利用视差信息生成基于棒状像素(stixels)的场景表达,可以有效减少行人检测候选区域的同时避免了像素级的深度图计算,在预处理阶段显著减少算法用时。将预处理阶段后的每一帧图像运用多维 HOG (multi-HOG) 特征 + LUV 颜色空间特征 + 交叉核支持向量机(hit-support vector machine, hit-SVM) 的特征融合目标检测策略,在改进传统 HOG 特征的基础上融合了颜色特征,显著提高算法实时性的同时保证了检测准确度。从而达到实时、准确、高鲁棒性的交通场景行人检测的目的。

1 基于 Stixel-world 场景表达的图像预处理

交通场景中的行人目标区别于其他环境下的行人目标,有其独有的特点,这些特点构成了本文基于双目立体视觉及 Stixel-world 图像预处理的基础。

1) 随着车载相机不断移动,背景复杂多变,在每一帧图像中遍历提取感兴趣区域的方式^[15]往往十分耗时,且存在大量冗余信息,准确度也不高,利用双目立体视觉提供的视差和深度信息提取感兴趣区域的方法则更为直接,检测效率和准确度也更高。

2) 行人目标基本存在于图像的上下居中位置,背景区域与行人目标往往有较为简单而明确的分界;行人目标基本与摄像机像平面平行,同一目标的视差值基本相同,方便判断。

3) 一般情况下的典型交通场景可以看作是由一个二维平面(地面)和垂直树立在上面的凸起物(行人、树木、车辆等)组成,计算像素级的深度图意义不大。

因此本文提出了采用棒状像素^[16]来简化场景表达的预处理手段,图 1 所示为不同形式的场景表达。

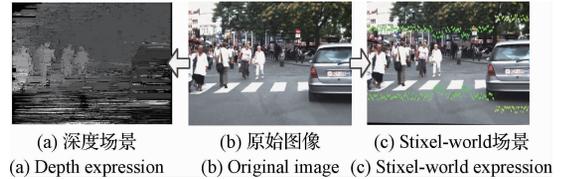


图 1 基于深度信息和基于 Stixel-world 的场景表达
Fig. 1 Scenes expression based on the depth information and based on Stixel-world

1.1 算法基本流程

由双目立体视觉图像构建 Stixel-world 的算法流程:首先将输入的原始双目图像进行立体匹配,构建视差空间;再根据视差信息完成地面估计、高度分割;最后将得到的 Stixels 显示出来即完成了整个 Stixel-world 的构建。

1.2 构建视差空间

双目立体视觉计算的前提是对双目图像进行立体匹配,立体匹配系统通过模拟人眼视觉感知原理,只有经过立体匹配将左右两幅独立的视觉图像联系起来才能进一步得到更有价值的信息。本文算法采用了半全局立体匹配算法(semi-global stereo matching, SGM)^[17]。这种方法基于局部最优的视差求取方法上,添加正则化约束并通过动态规划思想来提高计算结果平滑性,以达到求解出最优视差图的目的,视差计算结果以视差图形式显示如图 2 所示。

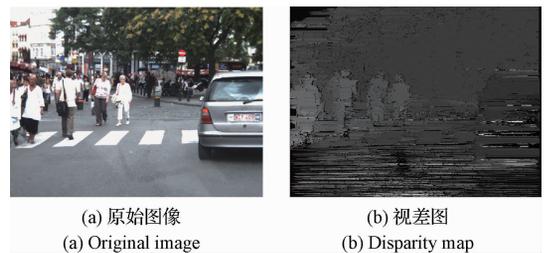


图 2 视差图构建结果
Fig. 2 Disparity map construction results

1.3 地面估计

在典型驾驶环境中,障碍物可以分为两类并抽象为

平面:行人、树木、路牌等垂直地面障碍物可以近似表示为垂直平面;地面则可以近似表示为水平面。构建的双目立体视觉系统三维几何模型如图3所示。

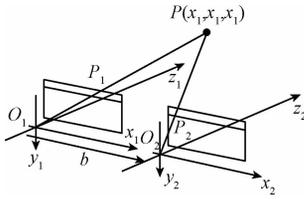


图3 双目立体视觉系统三维几何模型

Fig. 3 3-D Geometric Model of Stereo Vision System

地面估计算法的核心依据是处于同一距离的地面在水平方向具有相同的视差值,在图像 $I(u, v)$ 中, u 是图像的行坐标, v 是图像的列坐标。由双目立体视觉原理,可以求得视差 d 为:

$$d = u_l - u_r = f \frac{b}{z} \quad (1)$$

式中: u_l, u_r 表示目标点在左右图像中行坐标的位置, f 为摄像机焦距, b 为两个摄像机基线距离, z 为目标点到摄像机成像平面的距离。在本文中相对于光心的图像坐标系 (U, V) 中,左右图像行坐标 U 为:

$$U_l = u_l - u_0 = f \frac{X_w + b/2}{Y_w \sin\theta + Z_w \cos\theta} \quad (2)$$

$$U_r = u_r - u_0 = f \frac{X_w - b/2}{Y_w \sin\theta + Z_w \cos\theta} \quad (3)$$

推导得到水平方向视差 Δ 为:

$$\Delta = U_l - U_r = f \frac{b}{Y_w \sin\theta + Z_w \cos\theta} \quad (4)$$

当左右相机极线平行且与图像扫描线重合的时候 $v_r = v_l = v$, 因此左右图像列坐标 V 为:

$$V = v - v_0 = f \frac{Y_w \cos\theta - Z_w \sin\theta}{Y_w \sin\theta + Z_w \cos\theta} \quad (5)$$

在理想情况下,可以将路面抽象为水平面,在真实世界坐标系中平面的公式为:

$$Y_m = h \quad (6)$$

代入式(4)和(5)中,整理得到:

$$\frac{h}{b} \Delta = V \cos\theta + f \sin\theta \quad (7)$$

由式(7)可知,路面在 $V - \Delta$ 视差图中近似地投影为一条直线,图像坐标系 (U, V) 中, V 越大代表距离越远,对应的水平方向视差就越小,因此 $V - \Delta$ 视差图中地面应表现为一条从左上到右下的倾斜直线。利用 Hough 变换检测到直线并映射到原图像中即完成了地面估计部分,如图4中黑色网格所示,白色网格代表潜在的行人高度分布区域。

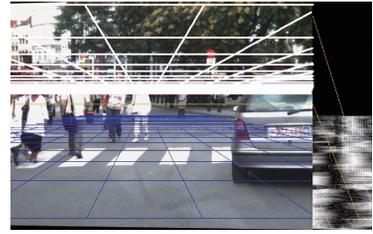


图4 地面估计结果

Fig. 4 Ground estimation results

1.4 高度分割

通过1.3节的方法提取出了交通环境中的地面,为构建 Stixel 提供了前提。在对每一个横坐标构建 Stixel 之后需要确定高度,从而确定 Stixel 上的像素点是否属于同一物体。最简单的情形是对每一个像素点求取视差值并判断,若大于某一阈值,则判断为背景,否则为物体。但是,该方法易受到错误视差的影响,并且对于不同距离的物体,阈值很难准确定义。本文采用设定隶属度函数的方式,对 Stixel 上的每个像素点进行划分,当隶属度函数值为正时表示物体,为负时表示背景。这样一来就将物体的高度分割出来,也就表达出了感兴趣区域的边缘信息。

定义隶属度函数如下:

$$M_{u,v}(d) = 2^{(1 - (\frac{d - \hat{d}_u}{\Delta D_u})^2)} - 1 \quad (8)$$

式中: d 表示某一像素点的视差值, \hat{d}_u 表示该像素点所在水平坐标 u 下的前景物体的视差值, ΔD_u 是人为给定的一个视差阈值参数。 ΔD_u 定义如下(本文中设定为2):

$$\Delta D_u = \hat{d}_u - f_d(z_u + \Delta Z_u) \quad f_d(z) = \frac{b \cdot f_x}{z} \quad (9)$$

式中: $f_d(z)$ 表示深度 z 对应的视差, 这样通过一个函数值来量化隶属度从而避免了透视效果带来的误差。对隶属度函数 $M_{u,v}(d)$ 通过式(10)计算得到代价函数, 代价函数 $C(u, v)$ 取极值时表示高度边缘点。

$$C(u, v) = \sum_{i=0}^{i=u-1} M_{u,v}(d(u, i)) - \sum_{i=v}^{i=v_j} M_{u,v}(d(u, i)) \quad (10)$$

再对代价图像进行动态规划计算,并以代价函数 $C(u, v)$ 作为动态规划的数据项:

$$c_{u,v_0,v_1} = C(u, v_0) + S(u, v_0, v_1) \quad (11)$$

$$S(u, v_0, v_1) = C_s |v_0 - v_1| \cdot \max\left(0, 1 - \frac{|z_u - z_{u+1}|}{N_z}\right) \quad (12)$$

式中: $S(u, v_0, v_1)$ 表示各个 Stixel 的顶点之间在垂直方向上的突变,表征的是分割边缘的平滑度。 C_s 是一个人为设定的比例因子, N_z 表示相邻 Stixel 之间深度阈值, 本文设为5 m, 高度分割结果如图5所示。



图 5 感兴趣区域分割结果

Fig. 5 Region of interest segmented results

由图 5 高度分割结果可知,采用 Stixel-world 的场景表达方式有效地划分出了行人和机动车在图像中的潜在位置,避免大量不必要区域的检测,可以极大提高后续行人检测的效率。在 TUD 数据库上的预处理时间每帧约为 10~20 ms(55 Hz 以上),考虑到预处理之后检测区域的缩小可达 50% 以上,因此不会对整个系统的实时性有太大影响。

2 基于特征融合的行人检测算法

基于统计学习的目标检测算法是目前行人检测领域应用最多的算法,这类算法的核心内容是特征的提取和分类。用于行人检测的特征中,Dalal N 等人^[4]在 2005 年提出的 HOG 特征是目前应用最为广泛的一种,后续很多针对行人检测的特征算子多是基于 HOG 特征的改进和算法融合^[18-20]。

2.1 构建 Multi-HOG 特征

在一幅图像中,像素的梯度主要存在于物体边缘的位置,因此局部目标的表象和形状能够被梯度的统计信息很好地描述。HOG 特征获取过程^[4]如图 6 所示,首先按固定大小将图片划分成小的 cells,统计每个 cell 的梯度直方图,其次将每几个 cells 组合成一个 block,串联得到一个 block 的 HOG 特征描述子,最后将图片内的所有 block 的 HOG 特征串联起来,得到的这个图片的 HOG 特征向量就可用于后续的学习和分类。

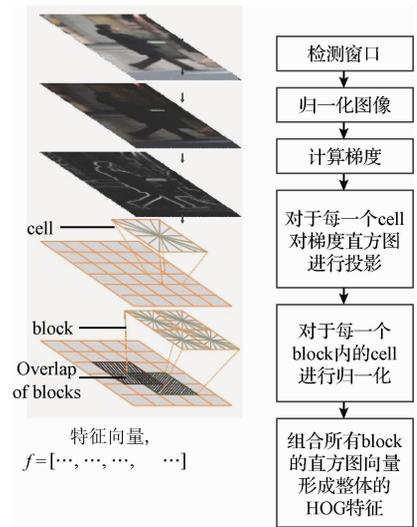


图 6 获取 HOG 特征基本流程

Fig. 6 Process of acquiring the HOG feature

虽然传统的 HOG 特征对光照变化和偏移旋转具有鲁棒性,但是存在冗余多、维数大(3 780 维)、计算速度慢的缺点^[4]。因此本文采用了多种尺度的 block 来优化传统的 HOG 特征从而实现降维,并采用 Fisher 准则来选取区分能力强的特征块,保留了传统 HOG 特征优点的同时,减少冗余特征的影响,达到优选特征块的目的。提取 Multi-HOG 特征包括如下两步。

1) 设计 block 尺寸

实验结果^[18]表明,不同尺度的 block 之间的重叠程度对 HOG 特征提取的好坏有直接影响,根据应用场景合理地组织 block 结构对行人检测的准确性和实时性有很大帮助,由于交通环境场景相比于监控录像等其他场景有其特殊性,存在着车速影响,block 尺度设置为 3 个梯度可以达到很好的标记效果,另外根据场景中绝大部分行人的宏观比例和局部特征,本文初步设置 block 尺度为 64×128 、 32×64 、 16×32 三个梯度供后续挑选。

由于本文采用 64×128 的滑动窗格来提取特征,不难计算得到,多尺度 HOG 特征在每个窗格中可以得到 21 个 block,而每个 block 分为 4 个 cell,每个 cell 有 9 个维度,那么按照 block 重叠程度为 0 的遍历方式计算,每个 64×128 的窗格中的 HOG 特征维度为 756 维,相比于传统 HOG 特征^[4]的 3 780 维检测效率有很大程度的提升,也更能突出图像中的重点,特征块尺度设置示意图如图 7 所示。

2) Fisher 准则挑选 block

Multi-HOG 特征构建了 21 个具有代表性的大小不同的 block,但是其区分能力强弱差异较大,甚至有的会提供冗余和负面的信息,本文采用 Fisher 准则对以上得到的 21 个 block 进行筛选,得到了最终 12 个最有价值的 block。

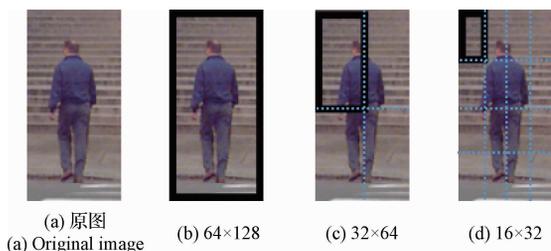


图7 设置特征块尺度
Fig. 7 Set the block scale

Fisher 准则通过寻找一个低维子空间,使得高维样本数据投影上来时类内距离最小,类间距离最大,算法上可利用类间距离和类内距离的比值最大化来表达这一目标。

类内离散度矩阵和总类内距离:

$$S_i = \sum_{X \in \omega_i} (X - m_i)(X - m_i)^T \quad i = 1, 2 \quad (13)$$

$$S_\omega = \sum_{i=1,2} S_i \quad (14)$$

类间距离 S_b 为:

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (15)$$

特征特征区分能力值的函数 $F(j)$ 为:

$$F(j) = S_b / S_\omega \quad (16)$$

根据区分能力大小对 block 进行取舍,结果如图8所示。

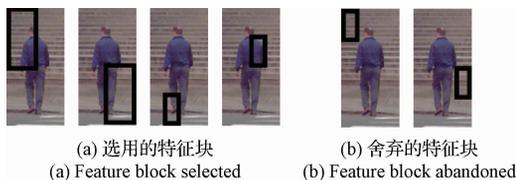


图8 部分选用的特征块与舍弃的特征块
Fig. 8 Part of the selected blocks and abandoned blocks

由图8筛选结果可以看出,分类能力强的 block 是有明显语义特征的,主要存在于行人头肩部、腿部等边缘处,符合 HOG 特征的检测思路。实验结果如图9所示,仅采用 12 个 block 形成的 432 维特征,就能达到采用 21 个 block、756 维特征检测效果的 99%。

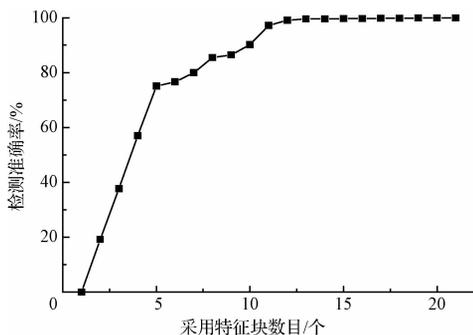


图9 特征块数目对检测准确率的影响
Fig. 9 The influence of the number of blocks on the accuracy

2.2 特征融合

通过单一特征进行的行人检测往往受到多变的环境和光照影响,而不同的底层特征具有不同的优势,通过特征融合的手段能实现检测算法的最优化。HOG 特征对光照变化和偏移旋转具有鲁棒性,在进行多尺度优化之后的 Multi-HOG 特征更是有了冗余特征少,计算速度快的特点,本文在对 HOG 特征进行多尺度优化的基础上融入了 LUV 颜色空间特征,考虑到颜色特征是图像检索与识别领域应用最为广泛的特征,而且与包含的场景和物体有直接联系,对图像本身的尺寸、方向、视角的依赖程度小,因此本文选择 Multi-HOG + LUV 颜色空间的特征串行组合策略来进行图像预处理之后的行人目标检测,如图10所示。

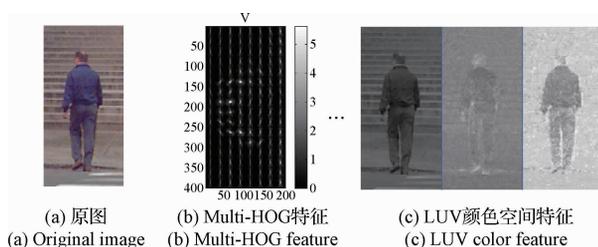


图10 特征通道和特征串接示意图
Fig. 10 Feature channel and feature fusion

2.3 分类器实现

现阶段基于机器学习的算法中,目标分类器主要是采用 SVM 和自适应增强学习算法 (AdaBoost) 两种,AdaBoost 在人脸检测方面^[21-22]的成功应用使它在行人检测方面被推广。AdaBoost 算法利用了弱分类器进行级联,具有很高的准确度,但是训练比较耗时;SVM 在解决非线性问题时优势明显,并且泛化能力强,不同核函数的 SVM 分类器适用于不同的场景,线性核 SVM 分类器训练和分类速度快,非线性核 SVM 分类器分类精度高,但是算法计算复杂度高,实时性较差。本文中的 hik-SVM 分类器采用的是直方图交集核函数^[23],采用折半查找法的线性搜索方法将时间复杂度从 $O(mn)$ 降为 $O(n \log m)$,具体实现流程如下。

本文采用的核函数为:

$$K_{\min}(x, z) = \sum_{i=1}^n \min(x_i, z_i) \quad (17)$$

将式(17)代入 SVM 分类函数表达式:

$$h(z) = \sum_{l=1}^m \alpha_l y_l K_{\min}(z, x_l) + b = \sum_{l=1}^m \alpha_l y_l \left(\sum_{i=1}^n \min(z_i, x_{l,i}) \right) + b \quad (18)$$

调换求和顺序,并将每一维度 n 中的所有样本的累加提取出来作为一个函数:

$$h_i(s) = \sum_{l=1}^m \alpha_l y_l \min(s, x_{l,i}) \quad (19)$$

每个样本要比较 n 次较小值,因此 m 个样本的计算复杂度为 $O(mn)$ 。

本文引入线性搜索与插值的概念,先将所有样本的第 i 维从小到大排列,再将 s 插值进去, $\min(s, x_{l,i})$ 变为两段的分段线性连续函数,前半段值为 $x_{l,i}$,后半段值为 s ,示意图如图 11 所示。

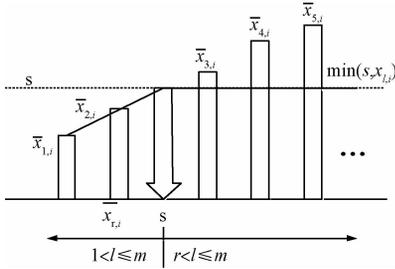


图 11 线性搜索及插值算法示意图

Fig. 11 Linear search and interpolation algorithm

$$h_i(s) = \sum_{1 \leq l \leq r} \alpha_l y_l x_{l,i} + s \sum_{r < l \leq m} \alpha_l y_l \quad (20)$$

r 为 $x_{l,i}$ 不大于 s 的最大整数下标,插值的关键在于如何确定 s 对应的 r ,本文采用数据分析中的折半查找法,经过排序和插值,算法的时间复杂度由 $O(mn)$ 降为 $O(n \log m)$,因此在行人检测样本数量 m 庞大的情况下,该方法能明显缩短训练时间,提升算法实时性。

3 实验结果

本文的实验在 Intel® Core™ i5, 双核 2.0 GHz, 4 G 内存的 PC 平台上完成。实验结果与传统 HOG 特征方法在 TUD、ETH 车载视频行人数据集和本文采集到的驾驶场景中的结果进行对比,分别从检测时间和检测准确度两个方面定量进行评价,定义检测准确度评价指标为误检率和漏检率。

3.1 车载视频行人数据集实验

本文采用目前被广泛使用于多行人检测与跟踪研究的基于双目立体视觉的 TUD 行人数据集和 ETH 行人数据集进行实验,其中 TUD 测试集采用一对车载摄像头进行拍摄,含有 508 对图像,图像对的时间间隔为 1 s,分辨率为 640×480 ;ETH 测试集采用一对车载的 AVT Marlins F033C 摄像头进行拍摄,帧率 13 ~ 14 fps,分辨率为 640×480 。

表 1、图 12、13 所示为本文算法和传统 HOG + SVM 算法在 TUD 行人数据集和 ETH 行人数据集上的检测结果对比。从表 1 可以看出,本文算法在 TUD 测试集上平均检测用时 87.66 ms,比传统的 HOG + SVM 算法的 379.42 ms 缩短了 76.9%;在 ETH 测试集上平均检测用

时 52.60 ms,比传统的 HOG + SVM 算法的 260.72 ms 缩短了 79.8%,基本达到了车载行人检测系统对实时性的要求。从图 12 可以看出,本文算法检测准确率较传统算法有很大提升,在误检率为 10% 的情况下漏检率分别从 70% 下降到 15%、从 52% 下降到 12%,在复杂的环境下总体行人目标误检率控制在 10% 以下,漏检率在 15% 以下。

表 1 算法执行用时对比

Table 1 Comparison of algorithm execution time (ms)

测试集	HOG + SVM	Stixel-world + 特征融合
TUD 测试集	379.42	87.66
ETH 测试集	260.72	52.60

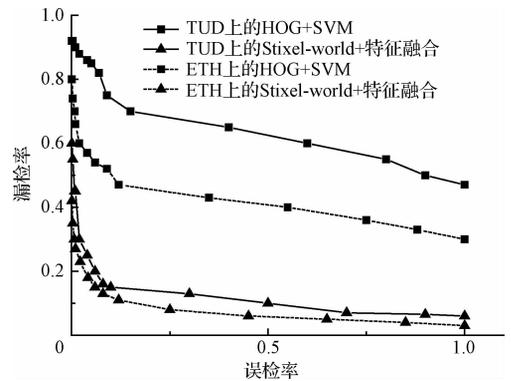


图 12 算法检测准确度对比

Fig. 12 Comparison of algorithm detection accuracy

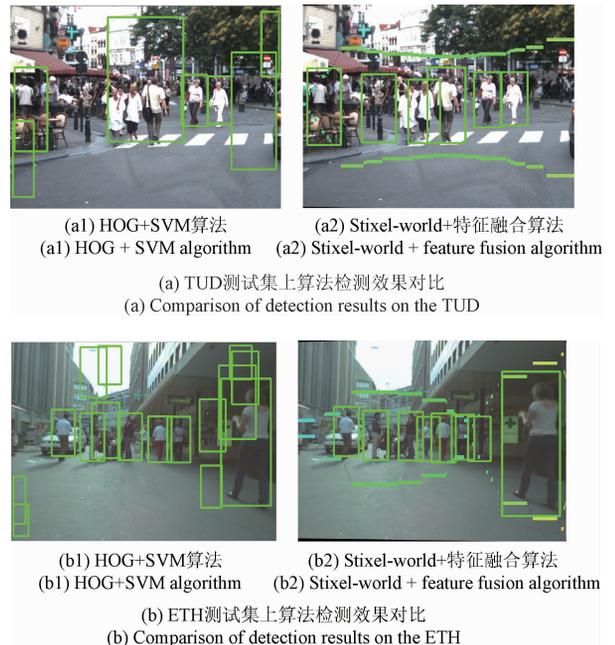


图 13 TUD 和 ETH 测试集上算法检测效果对比

Fig. 13 Comparison of detection results on the TUD and ETH

由图 12 和 13 结果对比可知,传统行人检测算法在 TUD 和 ETH 行人数据库中检测效果不佳,较多误检发生在行人潜在区域外的复杂背景中,目标框选结果也不够精确。这是因为在以往用于验证算法的 INRIA 等数据库中场景比较单一,行人较为规则而且冗余信息少。而 TUD 和 ETH 数据库是通过车载相机获得的,更加贴近真实交通场景,行人占了较小部分,背景也更为复杂,因此本文通过双目立体视觉信息划分出感兴趣区域的检测方法取得了较好的效果,也更适用于实际道路环境。

3.2 真实驾驶场景实验

相比于 TUD 和 ETH 复杂的行人数据,本文采集了背景较为简单、目标比较单一的交通场景作为补充,如图 14 所示。实验结果表明本文算法能够较为准确地划分出行人的潜在区域,部分场景缩小遍历检测区域达 73%。

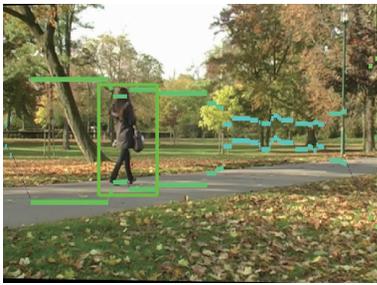


图 14 简单驾驶场景中的检测效果

Fig. 14 Results in a simple driving scene

4 结 论

本文利用双目立体视觉系统进行行人检测,根据得到的视差信息简化场景表达,在提取出感兴趣区域之后对行人目标进行了多特征融合检测。

1) 利用双目立体视觉可以得到视差和深度信息的优点,结合交通场景中行人的普遍特征,采用 Stixel-world 的场景表达方式,运用视差信息的同时避免了像素级的计算,降低了算法的复杂度。

2) 使用动态规划算法优化 Stixels 的合理性,有效提取出了感兴趣区域,与传统算法相比大幅减少了行人检测的候选区域,部分场景减小检测区域可达 73%。

3) 为保证检测准确性,优化筛选了 block 尺度形成 Multi-HOG 特征,并与 LUV 颜色通道特征相融合,实验结果表明本文算法在较为复杂的环境下总体行人目标误检率控制在 10% 以下,漏检率在 15% 以下,检测准确度较传统算法有了很大提升。

4) 结合了交叉核 SVM 分类器,在大量样本的情况下显著提升了系统实时性,在 TUD 和 ETH 数据库上的检测用时分别缩短了 76.9% 和 79.8%,基本满足车载行人检测系统对实时性的要求。

行人检测中的多姿态和复杂背景能够通过本文算法一定程度上得到解决,但是在出现严重遮挡情况下,本文算法还有局限性,行人目标存在的遮挡问题仍是未来行人检测领域需要解决的难题,也是下一步非常值得开展的工作。此外,随着计算机硬件的不断发展,更加高效的机器学习算法将会为行人检测领域增加更多可能性。

本文在提取图像中行人潜在区域的同时也将道路上的非机动车、机动车以及其他障碍物目标的潜在区域一并提取出来,为驾驶辅助乃至无人驾驶领域的各种障碍物识别提供了参考。

参考文献

- [1] World Health Organization. Global status report on road safety 2016[R]. 2017; 8-25.
- [2] 国家统计局. 中华人民共和国 2016 年国民经济和社会发展统计公报[R]. 2017;1-5.
National Bureau Of Statistics. Statistical communique of the people's republic of china on national economic and social development in 2016[R]. 2017;1-5.
- [3] DANIEL P H, KLANDERMAN G A, RUCKLIDGE W J. Comparing images using the Hausdorff distance[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9):850-863.
- [4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005:886-893.
- [5] 谈宇奇, 王雪, 林奎成. 基于视觉压缩感知的传感网络行人目标辨识方法[J]. 仪器仪表学报, 2014, 35(11): 2433-2439.
TAN Y Q, WANG X, LIN K CH. Visual compressive sensing-based pedestrian identification in sensor networks[J]. Chinese Journal of Scientific Instrument, 2014, 35(11): 2433-2439.
- [6] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. 自动化学报, 2016, 42(9): 1300-1312.
CHANG L, DENG X M, ZHOU M Q, et al. Convolutional neural networks in image understanding [J]. Acta Automatica Sinica, 2016, 42(9): 1300-1312.
- [7] 王恬, 李庆武, 刘艳, 等. 利用姿势估计实现人体异常行为识别[J]. 仪器仪表学报, 2016, 37(10): 2366-2372.
WANG T, LI Q W, LIU Y, et al. Abnormal human body behavior recognition using pose estimation [J]. Chinese Journal of Scientific Instrumentation, 2016, 37(10): 2366-2372.
- [8] AURISANO A, RADOVIC A, ROCCO D, et al. A convolutional neural network neutrino event classifier [J]. Journal of Instrumentation, 2016, 11(9): P09001- P09024

- [9] 李庆武, 仇春春, 俞楷, 等. 基于多尺度聚合通道特征的实时行人检测[J]. 电子测量与仪器学报, 2015, 29(11):1691-1697.
LI Q W, QIU CH CH, YU K, et al. Real time pedestrian detection based on multi-scale aggregation channel features[J]. Journal of Electronic Measurement and Instrument, 2015, 29(11):1691-1697.
- [10] 李钰, 苏育挺. 基于 Android 平台的行人检测系统[J]. 电子测量技术, 2012(12):99-101.
LI J, SU Y T. Pedestrian detection system based on Android platform [J]. Electronic Measurement Technology, 2012(12):99-101.
- [11] 金志刚, 赵明昕, 张瑞, 等. 基于双目视觉的聚合积分通道行人检测优化算法[J]. 天津大学学报: 自然科学与工程技术版, 2016, 49(12):1225-1230.
JIN ZH G, ZHAO M X, ZHANG R, et al. Optimized aggregated channel features pedestrian detection algorithm based on binocular vision [J]. Journal of Tianjin University : Science and Technology, 2016, 49(12):1225-1230.
- [12] ZHANG D. Stereo vision-based pedestrian detection using dense disparity map-based detection and segmentation [C]. Proceedings of SPIE-The International Society for Optical Engineering, 2015, 9443: 94430D-94430D-5.
- [13] AL-MUTIB K, EMADUDDIN M, ALSULAIMAN M, et al. Motion periodicity-based pedestrian detection and particle filter-based pedestrian tracking using stereo vision camera[J]. Mechatronics & Machine Vision in Practice, 2012, 5(6):32-37.
- [14] ZHANG D. Stereo vision-based pedestrian detection using dense disparity map-based detection and segmentation[C]. Proceedings of SPIE-The International Society for Optical Engineering, 2015, 9443: 94430D-94430D-5.
- [15] PIOTR D, CHRISTIAN W, BERNT S, et al. Pedestrian detection: An evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(4):743-761.
- [16] BENENSON R, TIMOFTE? R, GOOL L V. Stixels estimation without depth map computation [C]. IEEE International Conference on Computer Vision Workshops, 2012, 28(5):2010-2017.
- [17] HIRSCHMÜLLER H. Stereo processing by semi-global matching and mutual information[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008, 30(2):328-341.
- [18] 张楚金, 王耀南, 卢笑, 等. 基于假设验证和改进 HOG 特征的前车检测算法[J]. 电子测量与仪器学报, 2015, 29(2):165-171.
ZHANG CH J, WANG Y N, LU X, et al. Front-vehicle detection algorithm based on hypothesis and verification of improved HOG feature [J]. Journal of Electronic Measurement and Instrument, 2015, 29(2):165-171.
- [19] 刘威, 段成伟, 遇冰, 等. 基于后验 HOG 特征的多姿态行人检测[J]. 电子学报, 2015, 43(2):217-224.
LIU W, DUAN CH W, YU B, et al. Multi-Pose pedestrian detection based on Posterior-HOG feature[J]. Acta Electronica Sinica, 2015, 43(2):217-224.
- [20] ZENG CH B, ME H D. Robust head shoulder detection by PCA based multilevel HOG LBP detector for people counting[C]. Proceedings of the 2010 20th International Conference on Pattern Recognition. 2010:2069-2072.
- [21] 白中浩, 焦英豪, 白芳华. 基于主动形状模型及模糊推理的驾驶疲劳检测[J]. 仪器仪表学报, 2015, 36(4):768-775.
BAI ZH H, JIAO Y H, BAI F H. Driving fatigue detection based on active shape model and fuzzy inference[J]. Chinese Journal of Scientific Instrument, 2015, 36(4):768-775.
- [22] 毕雪芹, 惠婷. 基于肤色分割与 AdaBoost 算法的人脸检测[J]. 国外电子测量技术, 2015, 34(12):82-86.
BI X Q, HUI T. Face detection based on skin color segmentation and AdaBoost algorithm [J]. Foreign Electronic Measurement Technology, 2015, 34(12):82-86.
- [23] MAJI S, BERG A C, MALIK J. Efficient classification for additive kernel SVMs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(1):66-77.

作者简介



白中浩(通讯作者), 分别在 2000 年、2002 年和 2006 年于湖南大学获学士学位、硕士学位和博士学位, 现为湖南大学教授, 主要研究方向为汽车安全、汽车电子等。

E-mail: baizhonghao@hnu.edu.cn

Bai Zhonghao (Corresponding author) received his B. Sc., M. Sc. and Ph. D. degrees all from Hunan University in 2000, 2002 and 2006, respectively. Now he is a professor in Hunan University. His main research interests include automotive safety and automotive electronics.



王鹏辉, 2015 年于湖南大学获学士学位, 现为湖南大学硕士研究生, 主要研究方向为汽车电子、图像处理等。

E-mail: wangph_94@126.com

Wang Penghui received his B. Sc. degree from Hunan University in 2015. Now he is a M. Sc. candidate in Hunan University. His main research interests include automotive electronics and image processing.