

文章编号:1671-1637(2015)06-0118-09

## 基于单双目视觉融合的车辆检测和跟踪算法

蔡英凤<sup>1</sup>, 王海<sup>2</sup>, 陈小波<sup>1</sup>, 江浩斌<sup>2</sup>

(1. 江苏大学 汽车工程研究院, 江苏 镇江 212013; 2. 江苏大学 汽车与交通工程学院, 江苏 镇江 212013)

**摘要:**提出了一种基于单双目视觉融合的车辆检测与基于 Kalman 滤波的车辆跟踪算法,设计了一种基于二维深度置信网络的车辆检测器。在道路图像中利用单目视觉生成车辆可能存在的区域,构成双目视觉处理的车辆候选集合。在车辆可能存在的区域内利用双目视觉进行误检去除,并获得车辆的位置信息。在二维图像坐标系和三维世界坐标系内,利用 Kalman 滤波器对检测到的车辆进行跟踪。试验结果表明:算法的检测率为 99.0%,误检率为  $1.3 \times 10^{-4} \%$ ,检测时间为 57 ms,检测率高,误检率低,检测时间短;与单双目视觉弱融合算法、单目视觉算法和双目视觉算法相比,本文车辆检测与跟踪算法兼具双目视觉算法检测率高和单目视觉算法检测时间短的优点。

**关键词:**车辆检测;车辆跟踪;单双目视觉融合;二维深度置信网络;Kalman 滤波器

**中图分类号:**U491.116 **文献标志码:**A

## Vehicle detection and tracking algorithm based on monocular and binocular vision fusion

CAI Ying-feng<sup>1</sup>, WANG Hai<sup>2</sup>, CHEN Xiao-bo<sup>1</sup>, JIANG Hao-bin<sup>2</sup>

(1. Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013, Jiangsu, China;

2. School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China)

**Abstract:** The monocular and binocular vision fusion based vehicle detection and Kalman filter based vehicle tracking algorithm was proposed. The 2D deep belief network based vehicle detector was designed. In road images, the monocular vision was used to generate probably existing area of vehicle that composes vehicle candidate set processed by the binocular vision. The binocular vision was used to further eliminate error detection and obtain vehicle position information. The Kalman filter was used to track detected vehicles in 2D image coordinate system and 3D world coordinate system. Test result shows that the detection rate of the algorithm is 99.0%, the error detection rate is  $1.3 \times 10^{-4} \%$ , and the detection time is 57 ms. So the detection rate is high, the error detection rate is low, and the detection time is short. Compared to the monocular and binocular vision weak fusion algorithm, the monocular vision algorithm and the binocular vision algorithm, the proposed vehicle detection and tracking algorithm has both the advantage of binocular vision with high detection rate and the advantage of monocular vision with short detection time. 1 tab, 12 figs, 17 refs.

**Key words:** vehicle detection; vehicle tracking; monocular and binocular vision fusion; 2D deep belief network; Kalman filter

**Author resume:** CAI Ying-feng(1985-), female, lecturer, PhD, +86-511-88782390, caicaixiao0304@126.com.

收稿日期:2015-06-10

基金项目:国家自然科学基金项目(61203244,51305167,61403172);中国博士后科学基金项目(2014M561592,2015T80511);江苏省自然科学基金项目(BK20140555);交通运输部信息化技术研究项目(2013364836900)

作者简介:蔡英凤(1985-),女,江苏如皋人,江苏大学讲师,工学博士,从事基于视觉的智能车辆环境感知技术研究。

## 0 引言

先进驾驶人辅助系统作为提高道路与车辆安全性的重要手段,在世界范围内得到了高度重视和快速发展。该系统功能可以有效发挥的关键环节之一是对道路前方车辆的精确感知与跟踪。在车载环境下基于视觉的前方车辆检测与跟踪技术受到国内外广泛重视,视觉方式能够提供更多的信息量,且其核心技术很容易推广到基于其他传感器的系统中。然而,目前制约其发展的瓶颈就是视觉车辆感知的准确性不高<sup>[1-3]</sup>。目前的交通环境日趋复杂,天气与光照等环境因素对检测算法影响较大,同时,中国的混合交通流类型使得车辆常被遮挡,实际道路上行驶的车辆类别繁多,车辆视觉图像呈现较大类内差异性。上述特点增加了车辆检测和跟踪的难度。

基于单目视觉的道路车辆检测和跟踪,通常是在图像坐标系中应用训练好的车辆分类器,对遍历搜索得到的图像块进行检测,得出最可能的车辆位置,并完成与历史车辆区域的匹配<sup>[4-7]</sup>。这类方法可以在图像坐标系中获得目标的位置,运算时间短,但由于缺乏深度信息,检测效果受动态背景、光照和遮挡的影响较大。双目视觉的目标识别方法是另一种基于机器视觉的目标识别方法,在世界坐标系中通过对含有距离信息的点云进行分割和聚类实现包括车辆在内的障碍物检测,并通过滤波的方法完成与历史区域的匹配跟踪<sup>[8-10]</sup>。双目视觉技术可以获得被测目标在世界坐标系中的深度和位置信息,但由于需要搜寻大量匹配点,运算时间较长,无法满足先进辅助驾驶系统对车辆识别的高实时性需求。综上所述,单双目视觉技术具有原理上的互补性,将单目视觉提取的车辆区域作为双目视觉处理的兴趣区域,在兴趣区域内采用双目视觉获得的深度信息进行精准车辆定位,可提高目标识别的准确率。

现有研究大多集中在单独采用单目或双目视觉的手段进行车辆检测和跟踪,而融合这2种手段的研究报道较少。Toulminet等采用双目视觉对道路场景进行分割,重点对世界坐标系中的垂直边缘进行识别,以确定车辆可能存在的范围,利用单目视觉通过对称性、粗糙度等指标的约束进一步进行前方车辆的检测,在跟踪方面则采用相对简单的帧间相似度匹配加以实现<sup>[11]</sup>;Sivaraman等首先利用单目视觉以主动学习的方法在图像平面中检测出车辆,利用双目视觉获得被检测车辆的深度信息,最后采用Kalman滤波在图像坐标系和世界坐标系中同时

对被检测车辆进行跟踪<sup>[12]</sup>;Enzweiler等采用单双目视觉融合的手段进行了行人检测的研究<sup>[13]</sup>。

上述方法在单双目视觉融合的车辆检测和跟踪方面仍存在提升的空间。Toulminet等以双目视觉在全幅图像中进行车辆可能存在区域的搜寻,耗时长,而以对称性、粗糙度等简单的先验信息对车辆进行检测又存在检测率低的不足<sup>[11]</sup>;Sivaraman等仅以单目视觉进行车辆检测,未能利用双目视觉所提供的深度信息进一步对车辆检测结果进行验证,仅可定义为一种弱融合<sup>[12]</sup>。针对以上不足,本文提出了一种新的采用单双目视觉融合的车辆检测和跟踪算法,利用单目视觉计算耗时短的特点,采用基于二维深度置信网络的车辆检测器对图像中可能存在车辆的区域进行选择和确定,利用双目视觉技术准确获得车辆在图像坐标系和世界坐标系中的位置信息,由于实际道路上前车运动姿态变化的非线性程度较低而采用Kalman滤波器对检测到的车辆在图像坐标系和世界坐标系中进行跟踪,总体流程见图1。该方法避免了仅采用单目视觉技术存在的高误检率与仅采用双目视觉技术存在的深度图计算耗时长等缺点,具有一定的实际应用价值。

图1 算法流程

Fig. 1 Algorithm flow

## 1 基于单双目融合的车辆检测

本文所提单双目视觉融合的车辆检测算法分为2个步骤:第1步采用单目视觉方法,利用二维深度网络训练一个车辆检测分类器,在单幅图像上进行遍历搜索,生成存在可能车辆;第2步采用双目视觉方法,利用深度信息结合视差图对车辆可能进行再判断,实现车辆准确检测。

### 1.1 车辆可能存在区域生成

基于二维深度置信网络的车辆检测器见图2,它由可视层、隐层一、隐层二和标签层构成。可视层具有 $24 \times 24$ 个节点,其节点个数等于输入样本的维数。隐层一具有 $18 \times 18$ 个节点,隐层二具有 $12 \times 12$ 个节点。可视层与隐层之间通过一组权值进行

连接。为获得网络权值,首先采用包括车辆样本和非车辆样本的大量不含标签的图像样本进行无监督训练以获得初始权值,然后引入标签信息以反向传播的有监督训练方式优化网络权值。

图 2 车辆检测器

Fig. 2 Vehicle detector

将可视层和隐层一可以视为一个受限制玻尔兹曼机,可视层和隐层一节点间的状态能量  $E$  为

$$E = - \sum_{i=1}^{24} \sum_{j=1}^{24} \sum_{p=1}^{18} \sum_{q=1}^{18} v_{ij} a_{ij, pq} h_{pq} - \sum_{i=1}^{24} \sum_{j=1}^{24} b_{ij} v_{ij} - \sum_{p=1}^{18} \sum_{q=1}^{18} c_{pq} h_{pq} \quad (1)$$

式中:  $v_{ij}$  为可视层第  $i$  行第  $j$  列节点的状态;  $h_{pq}$  为隐层一第  $p$  行第  $q$  列节点的状态;  $a_{ij, pq}$  为可视层第  $i$  行第  $j$  列节点和隐层一第  $p$  行第  $q$  列节点的连接权值;  $b_{ij}$ 、 $c_{pq}$  分别为可视层和隐层一对应节点间的偏移量。

可得可视层状态和隐层一状态间的条件概率分布  $P_1$ 、 $P_2$  分别为

$$P_1 = \frac{1}{1 + \exp\left(- \sum_{i=1}^{24} \sum_{j=1}^{24} v_{ij} a_{ij, pq} - c_{pq}\right)} \quad (2)$$

$$P_2 = \frac{1}{1 + \exp\left(- \sum_{p=1}^{18} \sum_{q=1}^{18} h_{pq} a_{ij, pq} - b_{ij}\right)} \quad (3)$$

在此基础上,采用对比分歧算法<sup>[7]</sup>逐步更新连接权值和偏移量。

无监督训练可以使得网络间的权值获得较好的初始化,在此基础上将该二维深度置信网络视为普通的多层神经网络,采用反向传播的方法,引入标签层信息进一步微调网络权值,获得局部最优参数的目标函数  $D$  为

$$D = \min \left[ - \sum_l d_l \lg(e_l) \right] \quad (4)$$

式中:  $d_l$ 、 $e_l$  分别为第  $l$  个样本的真实标签和计算

标签。

经过上述训练可以获得一个能够在单目视觉图像上进行遍历搜索并判断出是否存在车辆的车辆检测器。对道路图像的检测结果与对应的车辆可能存在区域分别见图 3、4。具体获取方法是将所有车辆可能存在区域在原始图像上取其连通域的最小外接矩形,并将该矩形扩展 20%,以此作为可能的车辆兴趣区域。

图 3 检测结果

Fig. 3 Test result

图 4 可能存在区域

Fig. 4 Probably existing area

## 1.2 车辆存在验证

上述基于单目视觉的车辆检测,能够生成若干车辆假设,并依据获得的车辆假设建立车辆可能存在的感兴趣区域。然而,在这些车辆假设中,既可能

包含正确检出的车辆,也可能包含因为单目视觉的局限性造成的误检车辆。本节采用双目视觉技术在车辆可能存在的图像区域获取浓密视差图,生成对应的  $U$  视差图和  $V$  视差图, $U$ 、 $V$  为图像横、纵坐标,并在  $U$ - $V$  视差图中对单目视觉中所产生的车辆假设进行验证,剔除非车辆误检部分,以获得更加精确的车辆位置信息。

通过获取的  $U$ - $V$  视差图将原浓密视差图中的二维平面信息转换为一维统计信息,由此将障碍物的识别由平面检测转化为线段检测。与图 4 对应的浓密视差见图 5,其中较亮的区域对应的视差较大,而较暗的区域对应的视差较小。与图 5 对应的  $U$ - $V$  视差见图 6。

图 5 浓密视差

Fig. 5 Dense disparity

在实际道路情况下,路面区域是一个水平面,在  $V$  视差图中  $V$  方向上视差随着  $V$  坐标的增大(以左下角为原点)而减小,因此, $V$  视差图中路面区域对应的视差统计点在理想状态下是一条斜线段。同理,由车辆后部产生的垂直平面在  $V$  方向上视差保持不变,理论上是一条垂直于  $U$  轴的直线段。而斜线段和直线段的交点是车辆原始图像中车辆和路面接触点的纵坐标。与  $V$  视差图类似,车辆后部垂直平面在  $U$  视差图上是一条水平线段,且该水平线段包含了车辆后部垂直平面在原始图像中的横向范围。而路侧房屋、护栏等形成的与世界坐标系  $Z$  方向平行的垂直平面在  $U$  视差图上呈现斜线段。

通过以上分析可知, $V$  视差图确定了具有一定深度垂直面在图像坐标系下的纵坐标范围,而  $U$  视差图确定了该垂直面在图像坐标系下的横坐标范围。结合  $U$ - $V$  视差图采用 Hough 变换对视差图中的点进行直线拟合,就可以确定该垂直面在图像中的准确位置。直线拟合结果见图 7,最终车辆位置见图 8。

通过基于  $U$ - $V$  视差图的车辆验证,不仅可以剔

图 6  $U$ - $V$  视差

Fig. 6  $U$ - $V$  disparity

图 7 直线拟合结果

Fig. 7 Straight line fitting result

除误检车辆,还可以在视差图中获得深度信息并进一步获得前车相对于本车在世界坐标系中的位置坐标,为后续的车辆跟踪提供重要信息。

图 8 车辆位置  
Fig. 8 Vehicle position

## 2 基于 Kalman 滤波的车辆跟踪

在实际道路情况下,相对于本车,包括纵向运动和横向运动的前车运动状态变化的非线性程度一般较低,因此,采用 Kalman 滤波器在二维图像坐标系和三维世界坐标系中内同时对车辆进行跟踪。令车辆状态向量为

$$\mathbf{W}_k = (m_k, n_k, X_k, Y_k, Z_k, x_k, y_k, z_k) \quad (5)$$

式中: $m_k, n_k$  为被检测车辆  $k$  在图像坐标系下的中心点坐标; $X_k, Y_k, Z_k$  为被检测车辆  $k$  在世界坐标系下的中心点坐标; $x_k, y_k, z_k$  为被检测车辆  $k$  的中心点速度分量。

系统的状态方程和观测方程为

$$\begin{aligned} \mathbf{W}_{k+1} &= \mathbf{A}\mathbf{W}_k + \boldsymbol{\eta}_k \\ \mathbf{M}_k &= \mathbf{C}\mathbf{W}_k + \boldsymbol{\xi}_k \end{aligned} \quad (6)$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

式中: $\mathbf{M}_k$  为观测矩阵; $\boldsymbol{\eta}_k, \boldsymbol{\xi}_k$  分别为状态方程和观测方程的噪声; $\mathbf{A}, \mathbf{C}$  为系数矩阵; $\Delta t$  为步长。

在跟踪过程中采用 Kalman 滤波更新方程进行递归。对约 120 个连续图像序列中的黑色车辆进行跟踪,部分跟踪结果见图 9,被跟踪车辆的相对纵横向距离见图 10。可以看出,采用 Kalman 滤波器在对前方车辆进行跟踪的效果良好。

## 3 试验结果分析

在深度置信网络车辆检测器的训练中选取 Caltech 1999 数据库中的 126 个车辆后视图的图像以及项目组所采集的另外 600 个车辆后视图图像构成训练集正样本,同时选取 5 000 个不含有车辆的道路图像作为负样本。车辆检测器训练所用的某些典型正、负样本分别见图 11、12。

采用 Malaga 标准视觉图像库<sup>[14-15]</sup>中的 4 个图像序列进行试验。该序列中共包含 7 500 余个图像,其中含有车辆 6 391 veh。采用检测率和误检率作为评价指标

$$f_1 = \frac{F_1}{F_2} \quad (7)$$

$$f_2 = \frac{F_3}{F_1 + F_3} \quad (8)$$

式中: $f_1$  为检测率; $f_2$  为误检率; $F_1$  为正确检测车辆数; $F_2$  为总车辆数; $F_3$  为错误检测车辆数。

检测和跟踪性能比较结果见表 1,检测时间为检测一个图像耗费的时间。可以看出,本文算法采用单目视觉方法缩小了车辆可能存在区域的范围,因此,大大减少了双目视觉深度信息的计算时间,同时在双目视觉环节增加了通过深度信息对车辆进行进一步检测的步骤,故而检测率最高、误减率最低。本文算法处理一个图像耗时 57 ms,虽然大于单双目视觉弱融合算法的 46 ms 和单目视觉算法的 35 ms,但是远小于双目视觉算法的 175 ms。

## 4 结 语

本文基于单双目视觉方法在目标检测方面的互补性,提出一种单目视觉和双目视觉相融合的车辆精准定位算法,利用单目视觉处理速度快的特点,在

图 9 跟踪结果

Fig. 9 Tracking results

图 10 纵横向距离

Fig. 10 Vertical and horizontal distances

图 11 正样本

Fig. 11 Positive samples

图 12 负样本

Fig. 12 Negative samples



表 1 算法性能比较

Tab. 1 Performance comparison of algorithms

算法	来源	总车辆数/veh	正确检测车辆数/veh	检测率/%	误检率/%	检测时间/ms
单双目视觉弱融合	文献[12]	6 391	6 154	96.3	0.000 19	46
单目视觉	文献[16]	6 391	6 116	95.7	0.000 27	35
双目视觉	文献[17]	6 391	6 256	97.9	0.000 24	175
单双目视觉融合	本文	6 391	6 327	99.0	0.000 13	57

道路图像中生成车辆可能存在区域,利用双目视觉可以获得深度信息的优势,进一步在车辆可能存在区域内进行误检去除和准确定位,采用 Kalman 滤波方法在图像坐标系空间和世界坐标系空间对检测出的车辆进行跟踪。试验结果表明,本文所得车辆检测与跟踪方法兼具双目视觉检测率高和单目视觉检测时间短的优点,具有一定的实际应用价值。

#### 参考文献:

#### References:

- [1] TEOH S S, BRÄUNL T. Symmetry-based monocular vehicle detection system[J]. Machine Vision and Applications, 2012, 23(5): 831-842.
- [2] SIVARAMAN S, TRIVEDI M M. Integrated lane and vehicle detection, localization, and tracking: a synergistic approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 906-917.
- [3] 何 力,曲仕茹. 基于 PLS-VIP 特征降维的车辆检测[J]. 中国公路学报, 2014, 27(4): 98-105.  
HE Li, QU Shi-ru. Dimensionality reduction based on PLS-VIP for vehicle detection[J]. China Journal of Highway and Transport, 2014, 27(4): 98-105. (in Chinese)
- [4] CARAFFI C, VOJÍŘ T, TREFNÝ J, et al. A system for real-time detection and tracking of vehicles from a single car-mounted camera[C]//IEEE. 2012 15th International IEEE Conference on Intelligent Transportation Systems. New York: IEEE, 2012: 975-982.
- [5] ZHANG Zhao-xiang, TAN Tie-niu, HUANG Kai-qi, et al. Three-dimensional deformable-model-based localization and recognition of road vehicles[J]. IEEE Transactions on Image Processing, 2012, 21(1): 1-13.
- [6] MILANÉS V, LLORCA D F, VILLAGRÁ J, et al. Intelligent automatic overtaking system using vision for vehicle detection[J]. Expert Systems with Applications, 2012, 39(3): 3362-3373.
- [7] SIVARAMAN S, TRIVEDI M M. Active learning for on-road vehicle detection: a comparative study[J]. Machine Vision and Applications, 2014, 25(3): 599-611.
- [8] SEO D, PARK H, JO K, et al. Omnidirectional stereo vision based vehicle detection and distance measurement for driver assistance system[C]//IEEE. 39th Annual Conference of the IEEE Industrial Electronics Society. New York: IEEE, 2013: 5507-5511.
- [9] NGUYEN T N, MICHAELIS B, AL-HAMADI A, et al. Stereo-camera-based urban environment perception using occupancy grid and object tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(1): 154-165.
- [10] BARROIS B, HRISTOVA S, WÖHLER C, et al. 3D pose estimation of vehicles using a stereo camera[C]//IEEE. 2009 IEEE Intelligent Vehicles Symposium. New York: IEEE, 2009: 267-272.
- [11] TOULMINET G, BERTOZZI M, MOUSSET S, et al. Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis[J]. IEEE Transactions on Image Processing, 2006, 15(8): 2364-2375.
- [12] SIVARAMAN S, TRIVEDI M M. Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways[C]//IEEE. 14th International IEEE Conference on Intelligent Transportation Systems. New York: IEEE, 2011: 1249-1254.
- [13] ENZWEILER M, EIGENSTETTER A, SCHIELE B, et al. Multi-cue pedestrian classification with partial occlusion handling[C]//IEEE. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2010: 990-997.
- [14] KLAUS A, SORMANN M, KARNER K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure[C]//IEEE. 18th International Conference on Pattern Recognition. New York: IEEE, 2006: 15-18.
- [15] BLANCO J L, MORENO F A, GONZALEZ J. A collection of outdoor robotic datasets with centimeter-accuracy ground truth[J]. Autonomous Robots, 2009, 27(4): 327-351.
- [16] WANG Hai, CAI Ying-feng, CHEN Long. A vehicle detection algorithm based on deep belief network[J]. The Scientific World Journal, 2014, 2014: 1-7.
- [17] BAK A, BOUCHAFA S, AUBERT D. Detection of independently moving objects through stereo vision and ego-motion extraction[C]//IEEE. 2010 IEEE Intelligent Vehicles Symposium. New York: IEEE, 2010: 863-870.