

基于 Mean Shift 的多视频车辆跟踪研究

朱浩楠 许明敏 沈 瑛

(浙江工业大学计算机科学与技术学院 杭州 310023)

摘要 为了提高多视频下目标车辆跟踪的准确率,提出了一种基于 Mean Shift 结合视觉词袋的车辆跟踪方法。该方法采用 Mean Shift 提供的轮廓和颜色信息进行初匹配,并进行跟踪;针对车辆在不同视频下车视角、环境不同的情况,提出了尺度不变的识别方法,即利用视觉词袋特征作为车辆特征进行再次匹配。该方法能够利用高速路网中摄像机拍摄的视频确定目标车辆的具体位置。实验表明,基于 Mean Shift 的多视频车辆跟踪方法能够有效提高车辆跟踪的准确率。

关键词 多视频, 车辆跟踪, Mean Shift

中图分类号 TP391.4 **文献标识码** A

Research on Multi Video Vehicle Tracking Based on Mean Shift

ZHU Hao-nan XU Ming-min SHEN Ying

(School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract In order to improve the accuracy of target tracking in multi video, a vehicle tracking method based on Mean Shift combined with visual words was proposed. The method uses Mean Shift to provide the contour and the color information to carry on the initial match and the track. A scale invariant identification method was proposed for the situation of vehicle viewing angle and environment in different video, which regards the visual word bag feature as the vehicle features for match again. The method can be used to determine the specific location of the target vehicle by using the video camera in the high-speed network. Experimental results show that the Shift Mean based multi video vehicle tracking method can improve the accuracy of vehicle tracking.

Keywords Multi video, Vehicle tracking, Mean Shift

1 引言

随着安防科技的发展和网络带宽的增加,基于计算机视觉的智能交通系统得到了越来越普遍的应用。车辆跟踪、违章抓拍、车牌识别都是智能交通方面的热门研究领域。其中,多视频车辆跟踪是实现车辆识别、计数和交通流量统计的必要前提,也是智能交通的核心。各国研究人员对其进行了深入研究,取得了一定的研究成果。

多视频车辆跟踪是计算机视觉领域最活跃的研究课题之一^[1-2]。早在 20 世纪 80 年代,人们就开始研究多视频目标跟踪技术,如 Coifman 等^[3]提出了基于视频图像处理系统(Video Image Processing System)的交通监控系统,可以实现单视频车辆的类型识别和流量统计。随后 Haag 等^[4]在单视频基础上研究了多视频车辆跟踪问题。多视频车辆跟踪是在某个视频中确定目标车辆后在其他视频中搜索识别并跟踪的行为。在多视频跟踪中,车辆本身存在重叠、消失等情况,为此需要提供有效的车辆特征提取算法实时锁定车辆,同时减小车辆跟踪过程中图像改变引起的误差。在多视频跟踪过程中,难点集中在如何使计算机像人眼一样“认识”同一车辆目标。常规方法为建立车辆目标的特征向量集,采用多特征融合方法^[5-6]在车辆跟踪过程中不断提取车辆的特征,与目标车

辆特征进行对比来动态锁定目标。如国内研究者 Tai 等^[7]研究了用于车辆视频中事故判定的跟踪系统,其能够自动检测运动车辆并在多个视频中跟踪,但是使用的特征简单,识别和跟踪效率较低。Pai 等^[8]针对十字路口摄像头,采用车辆轮廓融合颜色特征跟踪路口车辆,但是没有考虑车辆速度和道路监控的分布,无法筛选有价值的视频。

为了实现复杂环境下的车辆特征匹配,本文提出了一种基于 Mean Shift 的多车辆跟踪算法,利用 SIFT 算子实时提取视频中车辆的特征,并与目标车辆比对。Mean Shift 结合 SIFT 算子具有尺度不变性,在多视频的情况下保证了特征不变性,在与原有特征的比较中可以更好地维持车辆特征的不变性,并能够用于实时跟踪。实验验证表明,针对多视频跟踪,采用优化 Mean Shift 方法结合视觉词袋模型,能够有效地在不同视频中识别目标车辆。

2 基于 Mean Shift 的多视频车辆跟踪算法

2.1 Mean Shift 目标初次匹配

在车辆跟踪领域基于单一特征或者多特征融合的方法是较为常规的,如利用车辆轮廓、颜色特征,但是其缺点也是显而易见的,寻找到一个对环境扰动、光线变化较不敏感的算法是车辆跟踪的关键。

本文受国家自然科学基金项目(61672464)资助。

朱浩楠(1991—),男,硕士生,主要研究方向为虚拟现实、图像处理等;许明敏(1990—),男,硕士生,主要研究方向为虚拟现实、图像处理等;沈 瑛(1976—),女,副教授,主要研究方向为信息安全、图形图像处理,E-mail:shenyings@zjut.edu.cn(通信作者)。

Mean Shift 是目标分割领域较为常用的算法,特点是计算量小,能够高效探索目标区域^[6];鲁棒性强,能配合一些尺度不变特征如 SIFT 特征在复杂环境下准确识别车辆目标。

但是该算法有一定的缺陷:

1) 车辆是刚性目标,但是在不同视频下车辆呈现不同角度,使车辆大小、车型轮廓发生变化,但在 Mean Shift 方法中跟踪核窗口的大小不变,极易造成目标跟丢的情况。

2) Mean Shift 的跟踪基于颜色直方图,由于不同视频处于不同场景中受到的光照不同,因此不同视频中相同车辆会有不同颜色特征。如光照强的视野中对象区域较亮,而光照不足的视野中对象的颜色较暗,因此即使提取同一目标车辆,在不同天气环境下得到的视觉直方图误差较大。

为此,本文提出了改进的基于 Mean Shift 的多视频车辆跟踪方法。算法的核心步骤为:调整视觉直方图、Mean Shift 特征初匹配、视觉词袋特征再匹配,具体流程如图 1 所示。

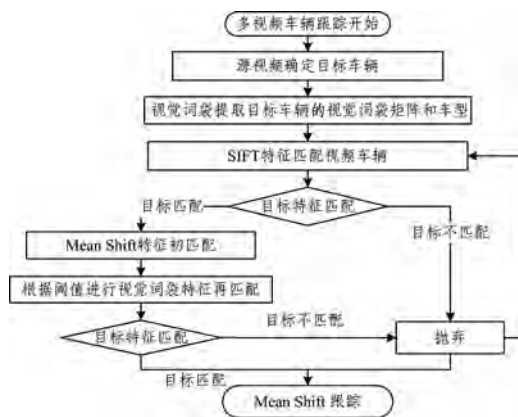
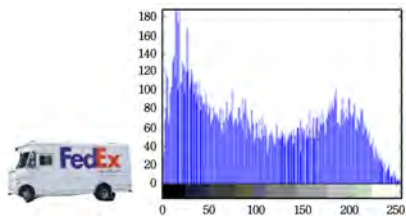
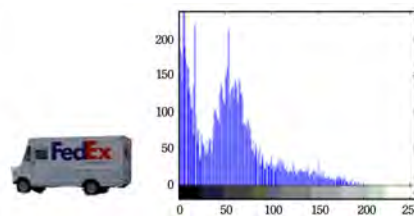


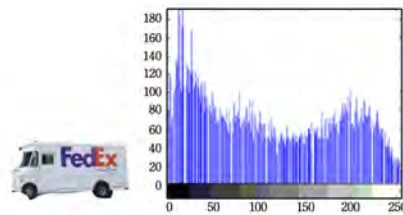
图 1 多视频车辆跟踪的流程图



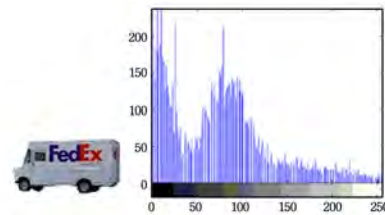
(a) 调整前 A 摄像机检测到的目标



(b) 调整前 B 摄像机检测到的目标



(c) 调整后 A 摄像机检测到的目标



(d) 调整后 B 摄像机检测到的目标

图 3 直方图调整前后对比

2.1.2 改进的 Mean Shift 跟踪初匹配

Mean Shift(均值漂移)车辆跟踪算法的本质是分别计算候选区域和目标区域像素的特征值概率得到候选模型和目标模型的概述,随后利用相似函数度量初始帧中嫌疑对象和目标之间的相似性,选取相似函数最大值,即相似函数的候选模型并得到目标对象 Mean Shift 向量,该向量正是目标由初始位置向正确位置移动的向量。Mean Shift 算法能够快速收敛,通过不断迭代向量,最终收敛到目标位置,实现识别和跟踪。Mean Shift 算法的描述如下。

1) 目标模型描述

通过手动标记获得视频帧中的目标区域。假设其中有 n 个相似,用 $\{z_i\}_{i=1 \dots n}$ 表示其位置,对选中的区域的灰度颜色空间均匀划分,得到由 m 个相等的区间构成的灰度直方图。目标模型的概率密度 $q_u (u=1, \dots, m)$ 可表示为:

$$q_u = c \sum_{i=1}^n K(\|z_i^* - u\|^2) \delta[b(z_i) - u] \tag{1}$$

$$c = 1 / \sum_{i=1}^n K(\|z_i^* - u\|^2) \tag{2}$$

$$z_i^* = \left(\frac{(x_i - x_0)^2 + (y_i - y_0)^2}{x_0^2 + y_0^2} \right)^{0.5} \tag{3}$$

为了减少不必要的视频查找,参考文献[8]中的方法,对当前监控中出现的目标车辆,按照车速匹配原则仅选取可能出现目标的视频。如图 2 中有多个摄像机对高速公路进行监控, O_1, O_2, O_3, O_4 代表车辆,以 O_3 为例,摄像机 C_2 下有目标 O_3 ,本文仅考虑该车辆在下一段时间或者几帧后出现在 C_1 或者 C_3 的情况,如果车辆处于摄像机视频盲区,目标车辆不能被检测到。根据 3.2 节中的方法判断目标车辆车速后假定目标会在 C_2 摄像机下出现。

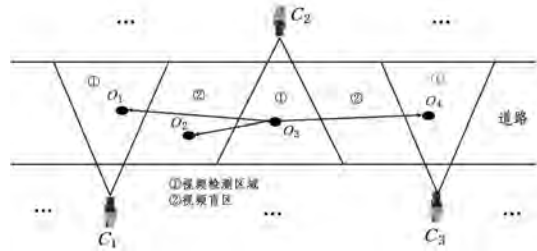


图 2 摄像机标定示意图

2.1.1 调整颜色直方图

在 Mean Shift 方法中需要计算目标的颜色直方图分布,但是由于两视频存在光照不同的情况,导致同一目标直方图差距较大,加大了目标识别的难度。为此,首先对颜色直方图进行调整,使不同视频下同一目标颜色尽可能接近。

直方图调整将使原本目标直方图像素调整到 $[0, 255]$,以提高图像对比度的方法。图 3 给出了同一车辆在不同场景下进行直方图像素调整后的对比效果。可以明显看出,调整前图(a)、图(b)的视觉直方图亮暗程度差别较大,在调整直方图后车辆的亮暗程度的差别明显缩小了。

其中, z_i^* 表示以目标中心为原点的归一化像素位置, (x_0, y_0) 为目标中心坐标; K 是核函数, 一般选用 Epanechnikov 核函数; $b(z_i)$ 表示 z_i 处像素属于哪个直方图区间; u 为直方图的颜色索引。 $\delta[b(z_i) - u]$ 函数的作用是判断目标区域中像素处的灰度值是否属于直方图中第 u 个单元, 属于则为 1, 否则为 0; c 是归一化系数。

2) 候选模型描述

在第 t 帧时, 根据第 $t+1$ 帧的目标中心位置 f_0 , 以 f_0 为搜索窗口的中心, 得到候选目标的中心位置坐标 f , 计算当前帧的候选目标区域直方图。该区域的像素用 $\{z_i\}$, $(i=1, \dots, n)$ 表示, 则候选模型的概率密度为: $p_u(f) = c \sum_{i=1}^n k(\| \frac{f-z_i}{h} \|^2) \delta[b(z_i) - u]$ 。 h 为核函数窗口大小; $\delta[b(z_i) - u]$ 决定权重分布, 以确定参数目标模型描述。

3) 相似性度量

相似性函数用于描述目标模型与候选目标之间的相似程度。本文采用 Bhattacharyya 系数作为相似性函数, 其定义为: $\rho(p, q) = \sum_{u=1}^m \sqrt{p_u(f)q_u}$ 。相似函数越大则两个模型越相似。将前一帧中目标的中心位置作为搜索窗口的中心, 寻找使得相似函数最大的候选区域, 即在本帧中目标的位置。

4) Mean Shift 迭代过程

均值漂移的迭代过程就是目标位置搜索的过程, 为使相似函数最大, 对上式进行泰勒展开, 得到 Bhattacharyya 系数的近似表达:

$$\rho(p, q) \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u(f)q_u} + \frac{c}{2} \sum_{i=1}^n w_i K(\| \frac{f-z_i}{h} \|^2) \quad (4)$$

$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u}{p_u(f)}} \delta[b(z_i) - u] \quad (5)$$

式(4)和式(5)中只有第二项随 f 变化, 其极大化过程可以通过候选区域中心向真实区域中心的 Mean Shift 迭代方程完成:

$$f_{k+1} = f_k + \frac{\sum_{i=1}^n w_i (f_k - z_i) g(\| \frac{f_k - z_i}{h} \|^2)}{\sum_{i=1}^n w_i g(\| \frac{f_k - z_i}{h} \|^2)} \quad (6)$$

其中, $g(x) = -K'(x)$, Mean Shift 方法就是从 f_k 起向两个模型相比颜色变化最大的方向不断移动, 直到最后两次移动距离小于阈值, 即找到当前帧的目标位置, 并以此作为下一帧的真实搜索窗口中心, 如此重复。

5) 运动目标跟踪的实现过程

运用传统的 Mean Shift 算法进行目标跟踪。首先对目标跟踪初始化, 通过目标检测方法得到需要跟踪的初始目标的外接矩形框, Mean Shift 算法中包括颜色和轮廓特征。本文中目标车辆特征已经通过手工框定获取, Mean Shift 根据目标车辆颜色和轮廓计算核函数加权下的搜索窗口的直方图分布, 用同样的方法计算第 N 帧对应窗口的直方图分布。随后调整视觉直方图, 减少两者的亮度差, 以提高识别的精度。以两个目标模板分布的相似性最大为原则, 使搜索窗口沿密度增加最大的方向移动, 得到目标的真实位置。

2.2 融合视觉词袋特征再次匹配

基于单纯改进的 Mean Shift 方法是“盲目”的, 即使优化了 Mean Shift 提取的直方图, 该算法仍然没有尺度不变的特征。但 Mean Shift 的轮廓信息为视觉词袋提供了待测车辆的

阈值(长宽比例)^[9], 即只有达到阈值时车辆才被视觉词袋分析, 否则直接抛弃。在初次匹配中, 算法只考虑了车辆的颜色和轮廓特征, 即便获取了车辆的这些特征, 也很难从众多车辆中筛选出目标车辆。基于视觉词袋的方法很好地解决了这个问题^[10-11], 视觉词袋对车辆来说是非常稳定的特征。不同型号、颜色的车辆有不同的视觉单词统计直方图, 在单视频车辆分类中可以轻松地地区分车辆, 在多视频跟踪中配合 Mean Shift 特征可实现有效跟踪。依靠核心的 SIFT 算子可获得稳定的特征, SIFT 特征的优势如下:

1) SIFT 特征在旋转、图像尺度变化、光照变化的情况下依然能够保持鲁棒性, 这一特性非常符合车辆跟踪的需要。

2) 车辆跟踪决定了图像处理的特征向量非常庞大, SIFT 特征丰富, 能在数量庞大的特征集中被准确找到。

3) SIFT 特征获取简单, 运算效率高, 能够与其他方法的特征向量融合。

视觉词袋算法通过提取车辆的视觉特征, 统计视觉单词直方图并利用机器学习分类得到。具体的流程如图 4 所示。

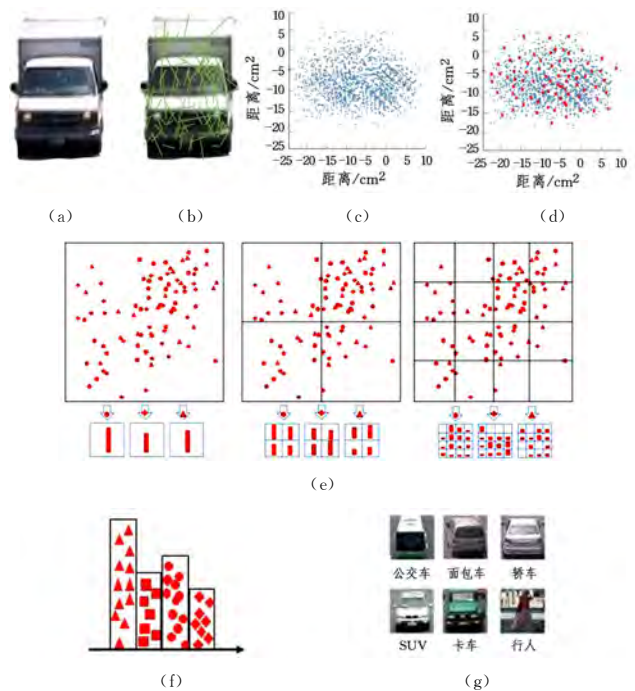


图 4 融合视觉词袋的分类方法

融合视觉词袋的分类方法的步骤如下:

1) 在训练阶段, 提取同类车辆的 SIFT 特征点。本次实验中共将车辆分为 6 类, 如图 4(g) 所示, 最终结果将在 6 类中获取。

2) 然后采用 E2LSH 聚类方法将特征点聚类为特征集合, 获取车辆的特征点并聚类得到视觉单词。

3) 通过空间金字塔矩阵对特征集合进行优化, 一般对特征集合做两次空间分割达到最优效果, 如图 4(e) 所示。形成某类车辆特定的视觉直方图, 如图 4(f) 所示。

4) 将视觉单词直方图矩阵送入 SVM 分类器进行学习。获得每类车辆的训练结果, 并将其作为分类的训练集。

5) 提取疑似车辆的 SIFT 特征, 利用空间金字塔分割视觉单词。

6) 生成视觉单词直方图后送入 SVM 分类器分类, 获得最终的车辆结果。

该方法的流程如图 5 所示。

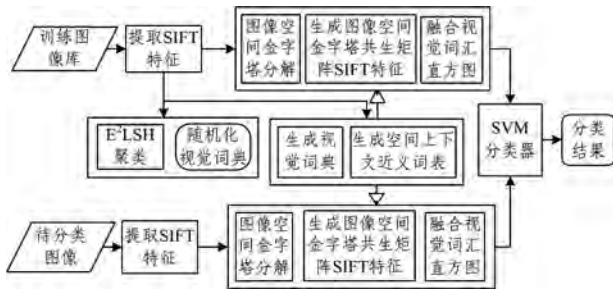


图 5 方法流程图

通过决策函数 $f(x) = \max[\omega_i \cdot x + b_i]$, 判别的结果为 i 类。这样在目标分类后就可以直观地获取车辆的类别。随后将 Mean Shift 首次匹配获得的嫌疑对象通过视觉词袋模型进行分类, 分类结果为 j 类。具体的判别过程如下:

1) 设目标车辆特征为 $T(C, L, BoW)$, 其中 C 为车辆的颜色直方图, L 为车辆的轮廓矩阵, BoW 为视觉词袋获得的类别。首先获取疑似车辆图像, 利用 Mean Shift 的视觉直方图和轮廓匹配, 因为已经对车辆颜色直方图进行改进, 初次匹配能够达到实用效果。

2) 如果车辆特征基本吻合, 则对疑似车辆进行视觉直方图匹配。通过 SIFT 特征和视觉金字塔分类得到视觉直方图并利用训练好的 SVM 分类器分类, 如果分类结果 $j = i$, 将疑似车辆设定为目标车辆, 利用 Mean Shift 进行跟踪; 如果车辆特征(颜色和轮廓)与目标车辆相差巨大, 并且 $j \neq i$, 则认定疑似车辆不是目标车辆, 不进行车辆跟踪。

2.2.1 SIFT 特征匹配

SIFT 提取的特征是视觉单词的基础, 在不同视频中车辆的特征是变化的, 只有遍历车辆并获得车辆“不变”的特征才能准确识别并跟踪车辆。由于 SIFT 局部特征具有平移、缩放、旋转不变性, 同时对光照变化、仿射及投影也有很好的鲁棒性, 因此非常符合车辆特征的匹配。具体匹配的方法如下: 利用向量的欧氏距离对特征点利用相似度度量准则判断特征点是否匹配, 利用目标车辆获取的特征点在另一视频中的疑似车辆中找到与之匹配的最近邻点和次近邻点。利用最近欧氏距离比值法(Nearest Neighbor Distance Ratio, NNDR)删除阈值大于设定阈值的特征对。NNDR 的大小决定了车辆匹配的精确度。NNDR 的计算式如下:

$$NNDR = \frac{d_1}{d_2} = \frac{|D_A - D_B|}{|D_A - D_C|} = \frac{\sqrt{\sum_{k=1}^{128} (a_k - b_k)^2}}{\sqrt{\sum_{k=1}^{128} (a_k - c_k)^2}} \quad (7)$$

其中, d_1 和 d_2 表示欧氏最短距离和次最短距离, D_A 是跟踪车辆的特征向量, D_B 和 D_C 是离 D_A 距离最近的两个目标特征向量。在不同场景下对 SIFT 特征进行了匹配, 结果如图 6 所示。



图 6 SIFT 实现特征匹配

2.2.2 误匹配剔除

如图 6 所示, 在特征匹配后存在很多误匹配点, 因此需要删除误匹配的特征点。本文采用 RANSAC 算法对误匹配点进行删除, RANSAC 算法是最常用的误匹配消除算法。由 Fishler 等^[12]提出的一种估计模型的迭代算法, 可以对粗糙的数据集进行筛选, 得到高置信度的数据集, 同时获得高精度的模型参数, 达到剔除误差较大匹配的特征点的目的。其算法的原理如下:

- 1) 假设模型 M 的参数初始化最少需要 n 个数据点, 从样本 P 中随机选取 n 个数据计算模型 M ;
- 2) 依据模型 M , 将 P 中其余点与模型 M 进行比较, 若误差小于给定距离阈值则归为内点, 使用内点集重新计算模型 M ;
- 3) 重复步骤 1) 和 2) K 次后, 最多内点对应的模型 M 即为所求参数模型, 对应内点提纯后即稳定值。

RANSAC 算法的实现过程中包含 3 个重要参数: 1) 距离阈值; 2) 估计次数 K ; 3) 一致集规模阈值。

图 7 给出了 SIFT 特征匹配后利用 RANSAC 剔除误匹配点后的效果。



图 7 RANSAC 剔除误差点后的场景效果

3 实验结果

本实验数据采用 CDW 提供的视频数据 (dataset2014/dataset/baseline/highway 和 dataset2014/dataset/PTZ/twoPositionPTZCam)。在 Windows7 Visual Studio2010 平台下对改进的 Mean Shift 结合视觉词袋方法和其他方法进行对比验证。

目标跟踪一共分为 3 部分, 分别为车辆大小变化和形变后的跟踪、多车辆下的目标车辆跟踪、车辆部分遮挡情况下的跟踪。

1) 车辆大小变化后的跟踪

视频中的车辆具有近大远小的特点, 但是 Mean Shift 是基于颜色和轮廓特征识别的, 大小不作为特征参与分析, 因此大小对车辆的跟踪没有影响。

2) 摄像机角度变化后的跟踪

通过分析结果发现, 由于视觉词袋 SIFT 特征具有尺度不变性特性, 车辆在图 8 中两图不同大小下能够准确跟踪, 在角度发生变化后能够有效识别目标车辆。



图 8 多视频下不同角度的车辆跟踪

通过图 8 分析发现, 利用车辆的轮廓和颜色特征能够准确识别目标车辆。如果视频中出现与目标车辆相似度极高的

车辆,通过学习机制算法能够寻找相邻帧中最近的相似车辆,基本不会产生车辆误识别的问题。

3) 车辆部分遮挡情况下的跟踪

如图 9 所示,遮挡情况下目标车辆分为无遮挡跟踪(见图 9(a))、部分遮挡跟踪(见图 9(b))和完全遮挡跟踪(见图 9(c))。在无遮挡下目标能够被准确跟踪,当车辆部分被遮挡时,基于 SIFT 特征和颜色特征能够识别目标车辆;如果车辆完全被遮挡并且短时间内再次出现,跟踪框能够智能识别并做短暂停留;如果目标长期消失,跟踪框将会漂移,目标跟踪失败。



(a)无遮挡跟踪 (b)部分遮挡跟踪 (c)完全遮挡跟踪

图 9 多视频目标车辆部分遮挡跟踪

为了进一步说明本文算法的效果,设计了实验流程和一种改进的 VRI(vehicle reidentification)方法[13]以进行对比。该方法使用滑动窗口代替全屏检测,大大缩短了车辆识别的时间。它通过车身特征(如颜色、轮廓)结合机器学习分类得到目标车辆。

另外,将本文的改进算法与原始的 Mean-Shift 和改进了颜色直方图的 Mean-Shift 算法进行比较,从时间和准确率两个角度衡量所提算法的效果。为此,笔者模拟道路摄像机位置,在留下镇 G2501 高速上录制了 6 段视频,在 3 个观测点各录制了 2 段 10 min 的视频(480 * 800pixel,每秒 20 帧),由于是绕城高速,车速较慢。3 个观测点录制同一批车辆通过 3 个观测点的视频。其中一批 3 段视频作为训练数据,另外一批 3 段视频作为测试数据。录制视频的位置如图 10 所示。



图 10 3 个观测点的位置

观测点 1 是目标车辆的入口,文中假设通过观测点 1 发现目标车辆,并且目标车辆存在一定几率在接下来某个时间点出现在观测点 2 和观测点 3 的视频中。

在算法准确率的实验中,我们设计了两个实验进行分析,分别是模拟两辆较特殊的车辆通过 2 个观测点时,4 个算法的识别和跟踪效果。为此选择了观测点 1 的一辆集装箱货车和卡车,实验 1 的步骤如下:

1) 选择一种算法,将观测点 1 的视频导入系统,手动框选目标车辆。系统根据不同算法计算得到不同的目标车辆特征。

2) 运行待查车辆,即观测点 2 或者观测点 3 的视频。系统自动筛选匹配车辆。

3) 如果目标车辆特征有疑似车辆吻合,则系统框选目标车辆,将车辆信息显示在系统中。

本次实验选取的目标车辆如图 11 所示。3 段测试视频经过改进的 ViBe 方法结合高斯模型[7]建模后的效果如图 12 所示。



(a)集装箱货车 (b)卡车

图 11 实验选取的目标车辆



(a)观测点 1 (b)观测点 2 (c)观测点 3

图 12 观测点背景图

车辆初次匹配后得到了阈值,理论上匹配车辆的数目和时间随着阈值的升高而降低,本次实验采用观测点 1 至观测点 3 的视频进行粗略匹配。取观测点 1 和观测点 2 中大型车(卡车、大巴)和小型车(SUV、面包车、轿车)各 1000 辆。将观测点 3 测试的大型车和小型车各 10 辆作为测试车辆,反复实验得出了如图 13 所示的结果。

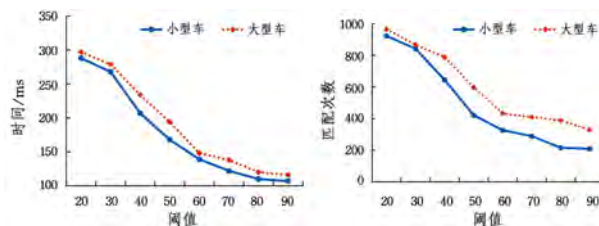
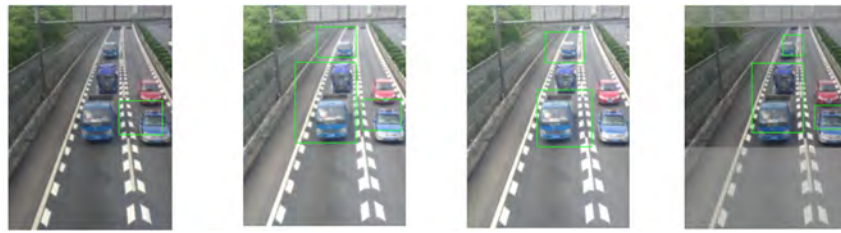


图 13 时间、匹配次数和阈值的关系

根据图 13 可以发现,合理地选取阈值能够有效提高车辆的检索效率。下面系统根据轮廓信息自动选取阈值,车辆跟踪结果如表 1 所列。目标车辆在观测点 2 中出现过 4 次,即存在 4 辆集装箱货车,分别是第 4 分钟 2 次,第 7、第 9 分钟各 1 次。图 14 为第 4 分钟时两辆集装箱货车出现的帧,4 种算法一定程度上能够筛选疑似车辆,改进的 Mean Shift 方法在 Mean Shift 方法基础上改良了颜色直方图,使视频颜色与目标背景相似。改进的 Mean Shift Bow 方法在改进的 Mean Shift 基础上加入视觉词袋模型,有效避免了轮廓、颜色相似情况下车辆重叠导致的不能识别。改进的 VRI 方法采用滑动窗口,只识别非蒙版范围内的车辆。虽然识别时间短,但是很容易造成误跟踪。第 4 分钟时两辆目标车辆的跟踪效果如图 14 所示。

表 1 集装箱货车跟踪结果

时间 (0~10 min)	方法(识别个数/误识别个数)				实际个数
	Mean-Shift	改进的 Mean-Shift	改进的 Mean-Shift&BOW	改进的 VRI	
1	0/1	0/0	0/0	0/0	0
2	0/0	0/0	0/0	0/0	0
3	0/0	0/0	0/0	0/0	0
4	1/0	1/0	2/0	1/1	2
5	0/0	0/0	0/0	0/0	0
6	0/1	0/0	0/0	0/0	0
7	0/0	1/1	1/0	1/0	1
8	0/0	0/0	0/0	0/0	0
9	0/0	0/0	0/0	1/0	1
10	0/0	0/0	0/0	0/0	0
总计	1/2	2/1	3/0	3/1	4



(a)Mean-Shift (b)改进的 Mean-Shift (c)改进的 Mean-Shift&BOW (d)改进的 VRI

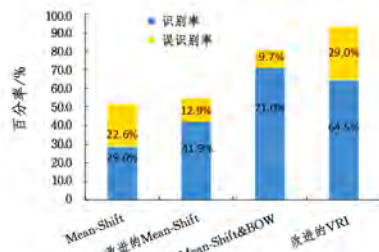
图 14 4 种算法跟踪集装箱货车的效果图

表 1 中的列代表视频时间,以分钟为单位。最后得到 4 类方法的识别率和误识别率。通过对比发现,相同条件下本文提出的基于视觉词袋改进的 Mean Shift 算法准确识别了 75%的车辆,与改进的 VRI 算法精度相同,但误识别率为 0%,低于改进的 VRI 的误识别率 25%。

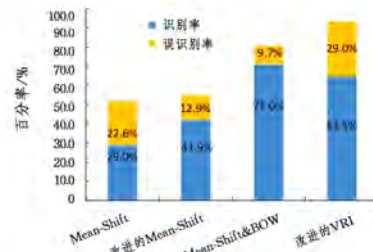
表 2 中的列代表视频时间,以分钟为单位。最后得到 4 类方法的识别率和误识别率,如图 15 所示。通过对比发现,相同条件下本文提出的基于视觉词袋改进的 Mean Shift 算法准确识别了 75%的车辆,和改进的 VRI 算法精度相同,但误识别率为 0%,低于改进的 VRI 的误识别率 25%。

表 2 卡车跟踪结果

时间 (0~10 min)	方法(识别个数/误识别个数)				实际个数
	Mean-Shift	改进的 Mean-Shift	改进的 Mean-Shift&BOW	改进的 VRI	
1	1/0	1/0	2/0	2/1	2
2	0/1	2/1	2/0	2/1	3
3	1/1	1/1	1/0	1/0	1
4	1/2	2/1	4/0	3/1	5
5	2/1	1/0	1/0	1/1	2
6	1/0	2/0	3/2	4/2	6
7	1/2	1/0	1/0	1/1	1
8	0/1	1/1	3/1	1/0	3
9	2/0	1/0	3/0	3/1	4
10	0/0	1/0	2/0	2/1	4
总计	9/7	13/4	22/3	20/9	31



(a)集装箱货车的识别率和误识别率



(b)卡车的识别率和误识别率

图 15 本文算法和其他算法对不同颜色卡车识别率的比较

实验 2 选择的目标车辆为普通卡车,测试 4 种算法对更普通车辆的识别和跟踪能力。实验步骤如实验 1 所示,但是手动框选的是观测点 1 的一辆卡车。待测视频为观测点 3 的视频。观察表 2 可知,卡车在视频中每一分钟都有出现。4

种算法的识别数量和误识别量也不相同。图 16 显示选取了一帧同时有 3 辆卡车、1 辆轿车的图像,基于视觉词袋的改进的 Mean Shift 算法的识别效果是最好的,同时拥有最低的误识别率。具体跟踪效果如图 16 所示。

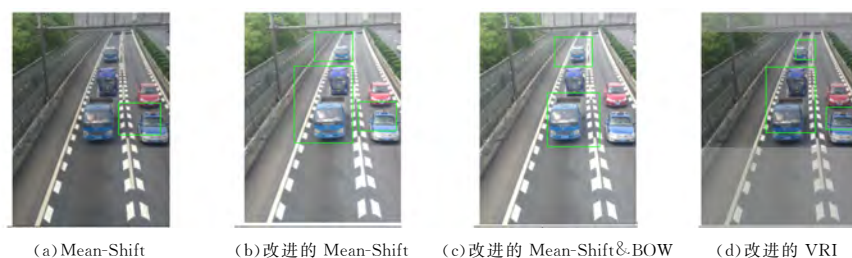


图 16 卡车的识别率和误识别率

为了测试 4 种算法的运行时间,设计了算法的时间耗费实验。4 类算法读入相同的视频,一共 600 帧(共 30 s,20 帧/s),计算每 10 帧消耗的时间。图 17 为计算得到的对比图,纵向观察发现,视觉词袋结合 Mean Shift 算法的时间耗费高于其他算法。Mean Shift 由于只提取车辆轮廓和颜色信息,因此跟踪最快。在 Mean Shift 基础上,改进了颜色直方图的算法略微比 Mean Shift 算法慢。而改进的 VRI 使用了车辆局部特征对车辆进行学习分类,一定程度上降低了车辆的识别速

度。视觉词袋结合 Mean Shift 的算法为了提高精度一定程度上牺牲了时间,算法中的空间金字塔视觉单词优化耗费了大量时间。但是通过图 17 发现 4 种算法的跟踪速度在同一数量级,对于实时性不高的车辆跟踪来说是可以接受的。另外从横向比较可以知道,第 190 帧到 250 帧,第 470 帧到第 530 帧出现了折线图的高峰,这是由于此时帧中前景图像复杂即车辆较多,需要比对每辆车导致耗费大量时间;在第 330 帧到 400 帧,视频中车辆较少,计算速度较快。

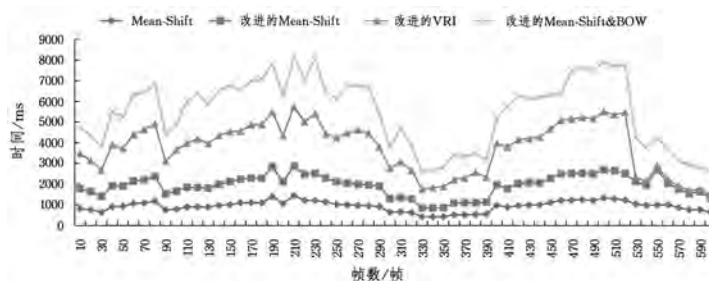


图 17 60 帧内 4 种算法耗费的时间

结束语 本文描述了目前多视频车辆跟踪领域的研究现状,针对存在的难点提出了解决方法。提出了采用 SIFT 特征匹配,视觉词袋结合 Mean Shift 的方法来解决多视频下的车辆跟踪这一难点。最后实验表明,本文提出的方法在一定程度上能够实现多摄像头下的车辆跟踪。

参考文献

[1] FLEISCHER K, NAGEL H H. Machine-vision-based detection and tracking of stationary infrastructural objects beside inner-city roads[J]. IEEE Proceedings Intelligent Transportation Systems, 2001, 1(1): 525-530.

[2] 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002, 25(3): 225-237.

[3] COIFMAN B, BEYMER D, MC LAUCHLAN P, et al. A real-time computer vision system for vehicle tracking and traffic surveillance[J]. Transportation research. Part C, Emerging technologies, 1998, 6(4): 271-288.

[4] NAGEL H H. Tracking of complex driving manoeuvres in traffic image sequences[J]. Image and Vision Computing, 1998, 16(8): 517-527.

[5] WANG Y Z, LIANG Y, ZHAO C H, et al. Kernel-based tracking based on adaptive fusion of multiple cues[J]. Acta Automatica Sinica, 2008, 34(4): 393-399.

[6] GU X, WANG H T, WANG L F, et al. Fusing multiple features for object tracking based on uncertainty measurement[J]. Acta Automatica Sinica, 2011, 37(5): 550-558.

[7] TAI J, TSENG S, LIN C, et al. Real-time image tracking for

automatic traffic monitoring and enforcement applications[J]. Image and Vision Computing, 2004, 22(6): 485-501.

[8] PAI C J, TYAN H R, LIANG Y M, et al. Pedestrian detection and tracking at crossroads [J]. Pattern Recognition, 2003, 37(5): 1025-1034.

[9] FUKUNAGA K, HOSTETLER L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. Information Theory, 21(1): 111-123.

[10] HE X H, ZHAO L. Moving object real-time detection and tracking based on improved Gaussian mixture model [J]. Jisuanji Yingyong Yanjiu, 2010, 27(12): 4768-4771.

[11] 陈伟宏, 肖德贵. 一种费重叠多摄像头的实时监控系統[J]. 计算机工程与应用, 2006, 42(18): 218-220.

[12] 磨少清, 刘正光. 多车道路面的大车车型分类特征及提取方法[J]. 计算机应用, 2009, 29(2): 586-589.

[13] 叶永伟, 任设东, 陆俊杰, 等. 基于 SVM 的汽车涂装线设备故障诊断[J]. 浙江工业大学学报, 2015, 43(6): 670-675.

[14] SONOYAMA S, TAMAKI T, Hirakawat, et al. Trade-off between speed and performance for colorectal endoscopic NBI image classification[J]. Medical Imaging, 2015, 12(3): 15-18.

[15] FISHLER M A, BOLLES R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and auto mated cartography[J]. Comm. of the ACM, 1981, 24(6): 726-740.

[16] WANG J, INDRA-PAYOONG N, SUMALEE A, et al. Vehicle Reidentification With Self-Adaptive Time Windows for Real-Time Travel Time Estimation[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 540-552.