交通路口监控视频跨视域多目标跟踪的可视化

刘彩虹 张 磊 黄 华
(北京理工大学计算机学院 北京 100081)

摘 要 跨视域大场景的多目标跟踪与展示是智能监控的基本需求之一.该文设计了一种基于视域拼接的跨视域多目标跟踪的可视化算法,借助于视频场景中几何信息实现视域拼接,从而实现将交通路口不同视角监控视频中的跟踪目标在统一的视场下展示.算法主要包含四个步骤:视域背景拼接、目标检测、跨视域多目标跟踪以及可视化显示.其中,视域背景拼接步骤利用交通场景背景图像几何信息辅助的半交互方式确定特征点对,计算不同视角到参考视域平面的单应变换矩阵,并利用SPHP算法保形后对所有配准图像线性融合以完成背景拼接;目标检测步骤利用ViBe背景建模算法分离目标,并进行阴影消除以提高检测准确性;跨视域多目标跟踪则结合各个视角到拼接视域平面的映射关系获得目标的定位信息,采用Kalman滤波和最小均方的轨迹匹配实现跨视域多目标的一致性跟踪;最后可视化显示步骤则在拼接的视域背景上对跟踪目标进行动态可视化展示.实验结果表明,该算法能够在统一视场下展现多个视域的监控场景信息,更方便于交通路口的监控.

关键词 跨视角;多目标跟踪;单应变换;ViBe;Kalman滤波;轨迹匹配
中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2018.00221

Visualization of Cross-View Multi-Object Tracking for Surveillance Videos in Crossroad

LIU Cai-Hong ZHANG Lei HUANG Hua
(*School of Computer Science, Beijing Institute of Technology, Beijing* 100081)

Abstract Cross-view Multi-object tracking and displaying for large scene is a basic requirement of surveillance video intelligent processing system. However, because of the limited field of view of a single camera, it's impossible to meet the exact requirement for long distance object tracking. In this paper, we propose an algorithm for visualization of cross-view multi-object tracking. We try to combine object information from multiple views and achieve the goal of tracking object in a wider field of view that is obtained by stitching the multiple views with overlapping areas. The proposed algorithm mainly contains four steps: background stitching step, object detection step, cross-view multi-object tracking step and visualization step. In the background stitching step, at first, we need to know the mapping from each view to the reference view plane. But there are too few feature points to find in the crossroad and we can't find the corresponding points with traditional methods that based on the feature point detection and matching. So we present a half-interactive method to determine the corresponding points between each view and reference view. The half-interactive method is based on geometrical information in background image, such as the vanishing point, horizontal line. It's easy to operate and reliable. And then we can use the obtained point pairs to calculate the homography matrix, by which we can project the camera image plane onto reference plane. Then the SPHP algorithm is adopted. It is on

收稿日期:2016-12-04;在线出版日期:2017-03-30.本课题得到国家杰出青年科学基金(61425013)、国家自然科学基金面上项目(61472035)资助.**刘彩虹**,女,1992年生,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为计算机图形学.E-mail:1090859045@qq.com.
张 磊,男,1981年生,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为计算机图形学.**黄 华**(通信作者),男,1975年生,博士,教授,中国计算机学会(CCF)会员,主要研究领域为图像与视频处理.E-mail:huahuang@bit.edu.cn.

account of that the severe projective distortion is introduced into the images after the perspective projection with the homography matrix and we use the SPHP algorithm to keep the perspective so that we can get an as wider as possible field of view. At last, we stitch the calibrated background images with linear fusion. Thus we get the mapping relation from each view to the stitching view. In the object detection step, we detect the objects with ViBe background modeling algorithm. For object detection, background subtraction is commonly used. One of most popular algorithm for background subtraction method is Gaussian mixture model. But in this paper, the ViBe algorithm is a more appropriate choice because it's less noisy and more convenient to solve the problem of ghost and the case when the object stops for a long time compared to the Gaussian mixture model. Although the ViBe algorithm is suitable, it's a pixel-level algorithm, and it is sensitive to shadows. So in order to get more correct object detection information, we have to remove the shadows to improve accuracy of object detection. Then, in the cross-view multi-object tracking step, we can use the obtained mapping relation to find the location of every object in reference view. And in each view, we achieve single-view multi-object tracking by Kalman filter. Once we get the trajectory in each single view, to determine the same object in different views by trajectory correspondence which uses the least mean square to matching the trajectory from different views. Thus finally in the visualization step, we implement the cross-view multi-object tracking and visualization algorithm. The experimental results demonstrate that our algorithm is efficient to the persistent object tracking and visualization.

Keywords cross-view; multi-object tracking; homography; ViBe; Kalman filter; trajectory correspondence

《计算机学报》

1 引 言

多目标跟踪及可视化是计算机视觉和计算机图形学领域的重要技术，是目标识别，交通分析，智能监控，异常行为检测等众多后续研究的基础，具有广阔的应用前景. 对于交通路口单个监控摄相机而言，其视域有限，不能满足对目标进行大范围、长距离持续跟踪的需求. 而对每个视角摄像机拍摄的视频进行独立跟踪和展示，又不便于跟踪目标的监测和分析. 跨视域多目标跟踪是研究多个具有重叠区域的视角的目标信息在同一视域下时空一致的跟踪，涉及目标检测、跟踪及同一视域下数据的关联和融合[1]. 跨视域大场景的多目标跟踪与展示是智能监控的基本需求之一.

虽然目标跟踪的研究很广泛，但跨视域多目标持续跟踪，由于除跟踪外还涉及视角间的数据关联(即重构)，导致相关研究工作比较少. Liu 等人[2]采用 MMPF 粒子滤波结合时空分析实现十字路口跨视角目标跟踪，其中目标的持续跟踪是利用视角间的转换实现的. Leal-Taixe 等人[3]对多个视角的已标定相机和目标数据，利用时域关联和空间约束，将跟踪问题转化成图的全局优化问题从而实现跨视域多目标跟踪；Hofmann 等人[4]利用 Leal-Taixe 等人构造流图的思想，并对其进行了简化，仅全局构造一个流图，实现 3D 空间的目标跟踪，这两者工作主要集中在对数据关联部分的研究，同样缺乏统一视域下不同视角视频目标跟踪的可视化呈现. Thingom 等人[5]则是以视频摘要的方式将不同视角监控结果进行集中展示和可视化，但其是在时域上对不同视角视频片段做摘要，并没有实际实现多个视角视频在同一视场下的同时展示.

本文设计实现了一种基于视域拼接的跨视域多目标跟踪的可视化算法，将交通路口不同视角监控视频中的跟踪目标在统一的视场下进行展示. 算法利用交通场景背景图像中的几何信息辅助进行交互的方式以确定特征点对，计算不同视角到参考视域平面的单应变换矩阵，实现大视差视域背景图像的拼接，进而实现目标跟踪信息的可视化. 为缓和单纯单应变换所造成的形变失真和视角信息丢失问题，本文采用 SPHP[6]算法进行保形. 另外，鉴于背景减除的检测算法不需要任何先验知识便能从视频流中

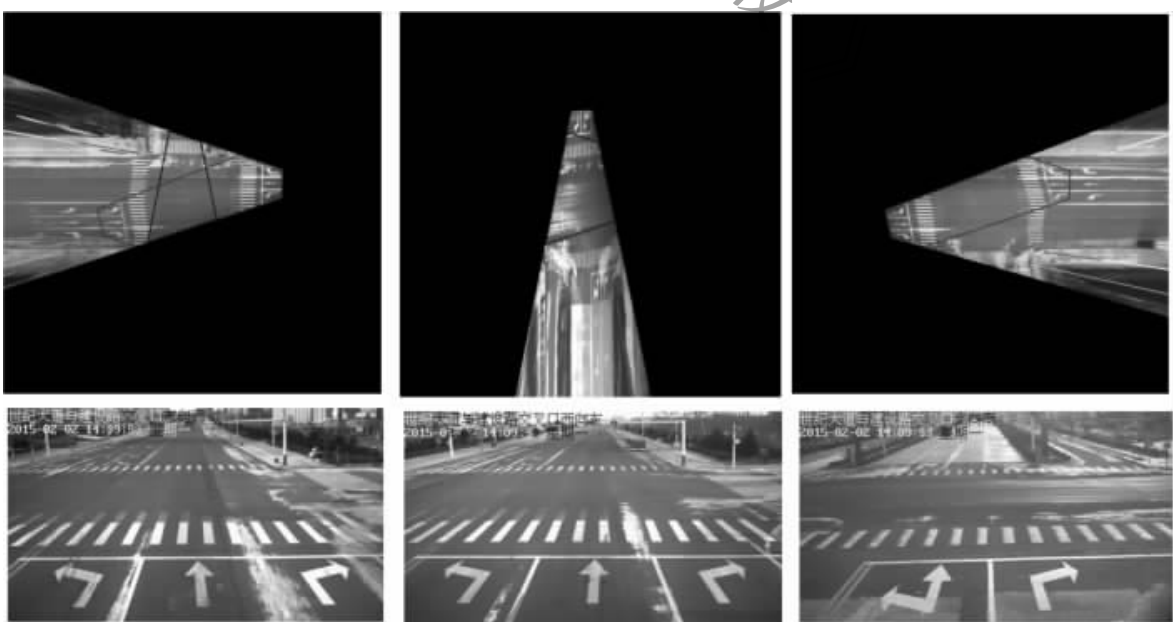
检测运动目标以及 ViBe[7] 算法的能够及时初始化，抗噪性能较强、计算量小，易于实现等特点，文中采用 ViBe 进行前景目标的检测. 同时对于已知背景图像的情况，ViBe 的背景减除算法也便于处理鬼影现象以及静止目标. 接下来的目标跟踪则是结合各个视角到拼接视域平面的映射关系和目标检测结果计算目标在拼接视域的定位信息，采用 Kalman[8] 滤波和最小均方的轨迹匹配实现可靠跟踪. 实验结果表明，本算法能够在统一视场下同时展现多个视域的监控场景信息，更有利于交通路口的监控.

2　相关工作

与本文相关的工作包括视角间配准和融合、运动目标检测、多目标跟踪等.

跨视域配准和融合. 为完成多个视角数据的融合，需先进行多个视角的配准. 视角间配准可通过相机标定信息建立各相机视角到世界坐标系的映射来实现，该方法适用范围广，例如文献[2-4]等都是在已标定相机的基础上进行研究工作. 传统的相机标定需要尺寸已知的标定物来确定三维空间点到二维图像点间的对应关系. 目前应用比较广泛的是 Zhang 等人[9] 的利用至少两幅不同视角平面图案进行标定的方法. 这种标定方法对于监控视频则需在监控相机所在地才能实现，需额外操作过程. 另一种方式是通过视角间对应点对的方式实现视角间配准. 视角间对应点对可通过手动或自动方式确定，自动的对应点对获取方法主要包括轨迹匹配、视域(FOV)线匹配、特征点匹配. 前两者对单个视角目标检测和跟踪的性能依赖性很大，特征点匹配的方法则要求每个视角地面场景具有足够的纹理信息[1]. 然而交通路口监控视频场景纹理单一，特征信息较少，而且视角间视差过大，重叠区域也比较小，图 1 是监控路口视角间重叠区域情况的一个实例，可见重叠区域仅包含道路中间的纹理信息较少的部分，而且各视角相似性也比较大，特征点的选取和匹配都比较困难. 图 2 是对其中两个视角进行特征点提取和匹配的结果，可见特征点难以准确匹配. 对于这种情况一种简单的方法是完全通过交互的方式人为指定所有对应点对，但在没有任何参照信息情况下指定对应点，操作困难且精度不高. 姜明新等人[10] 采用人为放置标志物的方式计算多层单应性进行信息映射，通过在拍摄环境中放置 4 根垂直于地面的带底座杆子，在不同高度的 3 层分别设置标志物来计算每层单应矩阵，虽精度较高但操作过程复杂.

视角间融合则是利用配准得到的单应矩阵，将各视角所检测前景映射到参考平面，以完成各目标的定位，确定同一目标. 所使用单应矩阵可以是单层单应矩阵或多层单应矩阵[10].



(a) 视角1　　(b) 视角2　　(c) 视角3

图 1　第 1 行图像为第 2 行单应变换后图像(深色区域为视角 1 和视角 2 重叠区域，浅色为视角 1 和视角 3 重叠区域)

图 2 交通路口不同视角背景图像进行特征点匹配

运动目标检测. 目标检测算法是通过对视频进行处理和分析,得到所需感兴趣的前景目标. 前景检测算法有很多,大致可分为两大类:背景减除法、训练分类法. 背景减除法是从静态背景中找到区别于静态背景的运动前景像素. 这种方法不区分目标的类型,不管是人还是车辆,只要是运动的感兴趣目标,均能够检测得出,但相对而言,其精度所受影响因素也较多,例如光照的变化、阴影等,一般常用于监控视频等背景基本不发生变化的静态相机所拍摄视频. 背景减除的算法有很多,文献[11-13]中都介绍了各种背景减除算法. 其中最简单一种是帧差法,其直接采用视频中的一帧为背景图像,采用当前帧减去背景的方式提取前景,对于捕捉目标的边缘比较有效,但易出现鬼影和空洞现象. 近年来使用比较广泛的是高斯混合模型[14]或者基于高斯混合模型的改进模型[15-17],这类方法是在时间域上建立像素的颜色分布模型,能够对光照渐变等复杂场景进行建模. 其计算复杂度较高,为此,Szwoch 等人[13]对其进行了并行实现,以提高其效率. 另外其对运动物体长时间静止的情况无法处理,且易出现误检情况和空洞现象,噪点较多. 另外一类全新视角的性能比较好的背景减除算法是 Barnich 等人[7]提出 ViBe 非参数化前景检测算法,其为每个像素建立一个样本集,利用像素变化的随机性,随机替换该像素及其邻域像素样本,据作者文中所述,ViBe 算法优于高斯混合模型、SACON[18-19]等复杂背景模型,具有易于实现,抗噪能力比较好、计算准确率高,计算量小,能够及时初始化的特点. 此后 Hofmann 等人[20]结合 SACON 和 ViBe 两种算法的优势,提出了性能更优的 PBAS 算法. 训练分类法是指对特定目标进行训练,建立模型的方法. 如车辆目标的检测[2,21-22],行人目标的检测[23-24]以及近来广泛应用的利用梯度直方图和 SVM 训练建立模型进行目标匹配的 DPM 目标检测算法[25],这类算法进行目标检测,一般结果会更精确和鲁棒,但这种方法通常只能检测特定类别的目标,甚至特定视角的目标,具有一定的局限性,而且还需样本数据集进行训练.

多目标跟踪. 近年来多目标跟踪的研究集中在对已检测目标进行跟踪(即 tracking by detection principle),即已知各视角间的空间对应关系和目标检测结果,建立各视角各检测目标的数据关联. 多目标跟踪算法分为单个视频多目标跟踪即单目标跟踪以及多个视角视频的目标跟踪即跨视域多目标跟踪.

单目标跟踪的方法有很多,例如利用目标区域的色彩、纹理等信息计算目标和候选区块相似度进行匹配的跟踪算法,比较常用的有 CamShift 算法,这种算法在发生很大遮挡或者目标发生较大形变时,

跟踪效果会变差；基于预测的跟踪方法，如利用可对目标运动有效进行线性最优估计的 Kalman 滤波器进行跟踪；另外还有维护多个假设，直到有足够多的观测信息来作出决策的跟踪方法，如基于马尔科夫链蒙特卡洛的方法[26-27]，比较早的基于粒子滤波的方法[28-29]等，这样的方法，由于假设空间的组合，复杂度呈指数级增长；基于匈牙利算法或者贪心策略实现数据关联和跟踪的方法，如 Breitenstein 等人[30]利用粒子滤波结合贪心的关联策略进行跟踪. 这些都是在线的方法，还有一些比较成功的离线跟踪方法，例如将通过预处理步骤得到可靠的长度短的轨迹再进行连接的多层跟踪策略[31]等.

跨视域多目标跟踪算法则主要分为两类：跟踪优先(track-first)算法和融合优先(fuse-first)算法[1]. 跟踪优先算法对每个视角进行跟踪，然后再在参考平面内进行投影和目标数据信息的连接. 例如文献[32]中通过 Kalman 滤波器在 3D 世界坐标系和 2D 图像坐标系分别进行每个目标的跟踪实现. 这种算法需进行多次跟踪，复杂度高. 融合优先算法则是先将检测所得目标信息投影到参考平面，再进行跟踪. 最近的跨视域多跟踪数据关联算法多是通过整个序列或者大的视域窗口的轨迹优化来实现[3-4].

3　算　法

本文算法流程如图 3 所示，主要包括四步：交通路口多个视角视域背景图像的拼接、目标检测、跨视角视频目标跟踪以及其可视化显示.

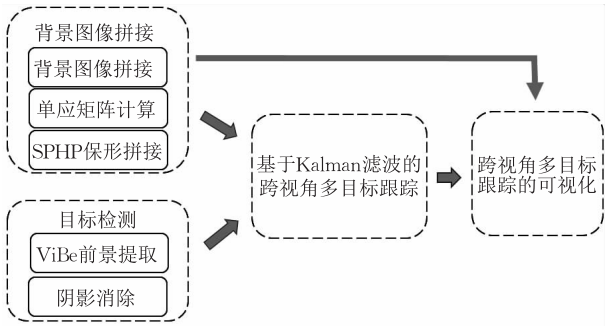


图 3　算法流程图

3.1　视域背景拼接

交通路口各视角图像一般视差很大，甚至达到 90°，且监控相机拍摄多为路面信息，纹理信息过少，很难通过特征点的方式完成各视角间的准确配准. 本文借助视频场景中主要几何信息以确定各视角图像到参考平面的单应变换来进行图像配准，并利用保形的图像拼接算法，建立统一的视域背景.

由于各视角的相机曝光差异等，可能导致各视角拍摄所得视频亮度差异过大而使得拼接图像视觉效果不好. 因此拼接前需进行亮度调整，使得各视角背景图像亮度尽可能相似，以提高拼接结果的视觉质量. 本文采用 Reinhard 等人[33]的颜色转换算法进行亮度调整，该算法利用简单的统计分析将原图的颜色特性转移到目标图像. 颜色调整结果如图 4 所示. 以下过程均基于颜色调整后图像.

(a) 南　　(b) 北

图 4　以东向图像颜色为基准，对南北方向图像进行亮度调整的结果(其中第 1 行为基准图像，第 2 行为原始图像，第 3 行为颜色调整后图像)

3.1.1 单应变换的计算

计算机视觉中,单应变换定义为一个二维平面到另一个二维平面的投影映射.通常由给定特征点对确定两个平面的单应变换矩阵:$\boldsymbol{p}(x,y,1)^{\mathrm{T}}\mapsto\boldsymbol{p}'(x',y',\omega)^{\mathrm{T}}$,$\boldsymbol{H}$ 为 3×3 矩阵,即

$$\boldsymbol{p}'\sim\boldsymbol{H}\boldsymbol{p},$$

其中,$\boldsymbol{H}=\begin{bmatrix}h_1&h_2&h_3\\h_4&h_5&h_6\\h_7&h_8&1\end{bmatrix}=\begin{bmatrix}\boldsymbol{h}^{1\mathrm{T}}\\\boldsymbol{h}^{2\mathrm{T}}\\\boldsymbol{h}^{3\mathrm{T}}\end{bmatrix}$,$\boldsymbol{h}^{j\mathrm{T}}$ 表示 $\boldsymbol{H}$ 矩阵的第 j 行.方程可叉积表示为

$$\boldsymbol{p}'\times\boldsymbol{H}\boldsymbol{p}=\boldsymbol{0},$$

去除线性相关项后方程为

$$\begin{bmatrix}\boldsymbol{0}^{\mathrm{T}}&-\omega\boldsymbol{p}^{\mathrm{T}}&y'\boldsymbol{p}^{\mathrm{T}}\\\omega\boldsymbol{p}^{\mathrm{T}}&\boldsymbol{0}^{\mathrm{T}}&-x'\boldsymbol{p}^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^1\\\boldsymbol{h}^2\\\boldsymbol{h}^3\end{bmatrix}=0\qquad(1)$$

一般取 $\omega=1$,可知平面间一个特征点对至多可得两个线性无关方程组,故至少需四对特征点方可确定一个单应矩阵 $\boldsymbol{H}$.

为获取摄相机成像平面到参考平面(本文选取地平面)的特征点对,本文依据已知的背景图像或者地面实况数据确定交通路口的简化模型(如图 5 所示).基于该模型,本文采用半交互方式进行单应变换的计算,即通过从路面背景图像获取信息并辅助交互来确定两平面间的对应关系,包括消隐点约束、水平线约束及用户交互的点约束三部分.由于各视角计算方式相同,下面以如图 6(a)所示东向西视角的图像为例进行说明.

消隐点约束.本文利用真实空间平行直线相交于消隐点[34]这一特性对图像中线段进行聚类来检测消隐点.其中线段采用 von Gioi 等人[35]的 LSD 方法进行检测.检测的线段按消隐点进行分类后结果如图 6(b),其中,相同颜色的线段代表同一类线段.对分类得到的每组线段进行拟合,便可得到各消隐点.交通路口图像检测的线段往往多为人行道线,人行道线段相互平行,与相机拍摄方向一致的线段会很快收敛于一点,利用这些特性便可过滤出图像垂直方向的消隐点(如图 6(c)所示).

将计算得到的消隐点向量表示为 $\boldsymbol{vp}=(v_x,v_y,1)^{\mathrm{T}}$,消隐点对应在地平面上点应为相应无穷远点,即$(v_x,v_y,0)^{\mathrm{T}}$,代入式(1)可化简为

$$\begin{bmatrix}\boldsymbol{0}^{\mathrm{T}}&\boldsymbol{0}^{\mathrm{T}}&(\boldsymbol{vp})^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^1\\\boldsymbol{h}^2\\\boldsymbol{h}^3\end{bmatrix}=0\qquad(2)$$

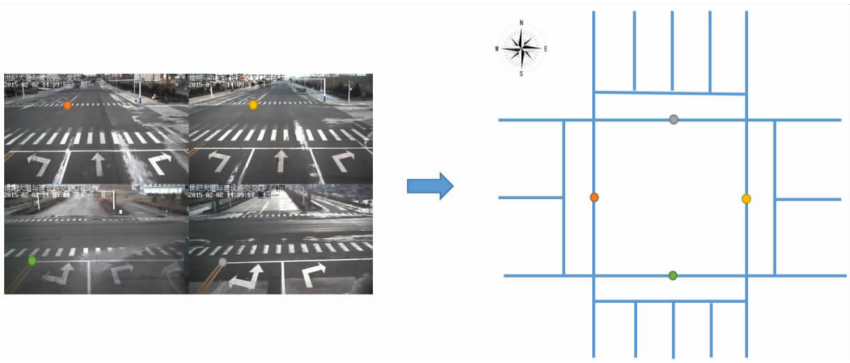
水平线约束.水平线约束与消隐点约束相同,也是利用图像中线段集分类得到的各组线段进行信息提取,以找到与图像水平方向的线段集.这类线段通常具有方向相同、消隐点在无穷远处的特性.同样利用这些性质对每组线段拟合出一个方向,便可从中找出所需水平方向(如图 6(d)所示).

将计算得到的水平方向向量表示为 $\boldsymbol{hd}=(h_x,h_y,0)^{\mathrm{T}}$,其变换后也必为水平方向,即 $\boldsymbol{p}=\boldsymbol{hd}$,$\omega=0$,$x'\oplus y'=1$,代入式(1)为

$$\begin{bmatrix}\boldsymbol{0}^{\mathrm{T}}&\boldsymbol{0}^{\mathrm{T}}&(\boldsymbol{hd})^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{h}^1\\\boldsymbol{h}^2\\\boldsymbol{h}^3\end{bmatrix}=0\qquad(3)$$

用户交互的点约束.利用上述两个约束不足以计算单应变换,本文利用图像上人行道信息辅助交互的方式另外确定 3 对特征点,相对于完全用户交互的方式,可靠性和精确性都更高.

利用上述算法提取水平方向线段端点进行 K-means 聚类(如图 6(e)),对各类线段点进行拟合,提取人行道线(如图 6(f)).人行道线位置在简化参考平面上已知,依据所得人行道线(图 7 中 line1 和 line2),只需交互确定线上特定的两个点(图 7 中点 x_1 和点 x_2),即可得到所需三对点.



(a) 各视角背景 (b) 简化模型

图 5 依据各视角背景建立的简化路口地面模型

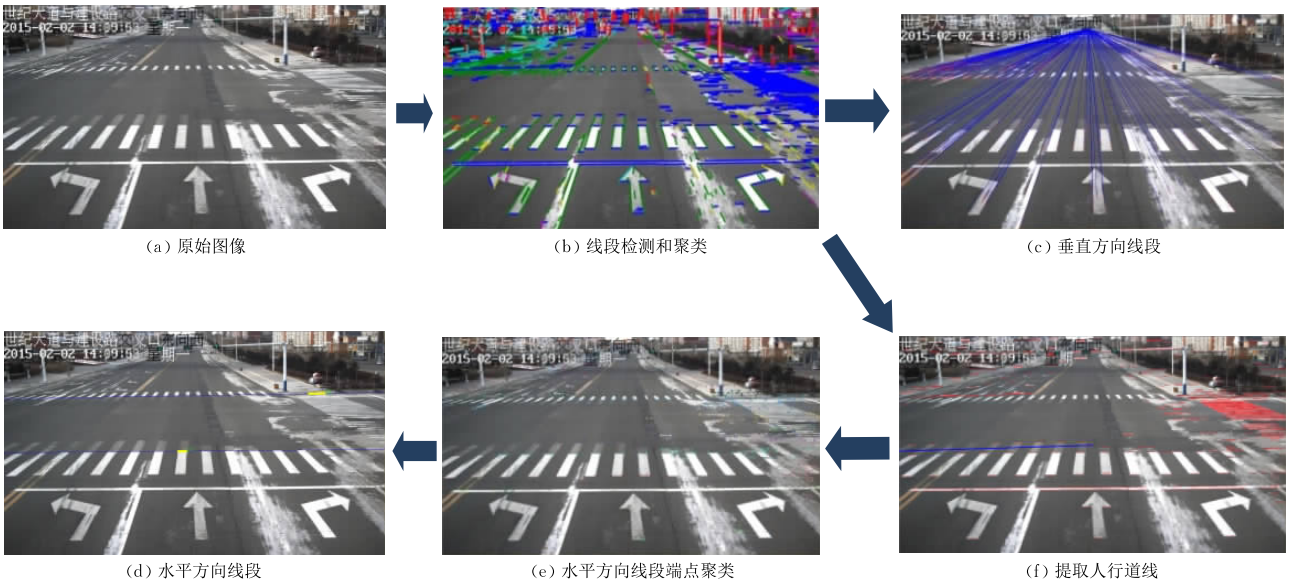
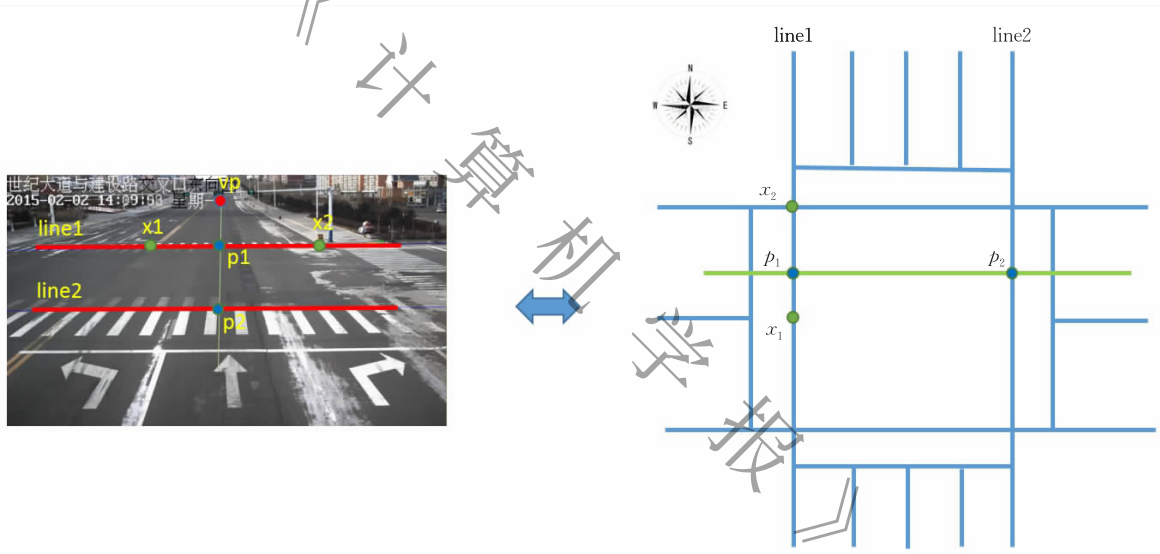


图 6　背景图像几何信息提取



(a) 图像平面　　(b) 参考视角

图 7　半交互确定特征点对

从消隐点作 line1 和 line2 的垂线分别交于 p_1，p_2. 由于 p_1、x_1、x_2 共线，设 $\frac{x_1 p_1}{x_1 x_2} = k$，则 p_1 可由 x_1、x_2 表示为：

$$p_1 = kx_2 + (1-k)x_1 \tag{4}$$

由于人行道线在水平方向，且水平方向消隐点在无穷远处，因此线段长度满足线性关系，即参考平面上 p_1 点与 x_1，x_2 应具有相同的比例关系，由此可得参考平面上 p_1 坐标. 而 p_1，p_2 垂直与 line2，可得参考平面上 p_2 坐标. 选择 p_2 及 line1 上任意 2 点，即可得 3 对对应点.

结合两个约束，以及交互得到的 3 对对应点，便可确定图像视角到地平面的单应变换矩阵 $\boldsymbol{H}$.

3.1.2　交互鲁棒性讨论

依据前文所述，由于交通路口监控视频场景纹理信息单一、视角间视差较大、重叠区域较小，特征点匹配方法不可行. 对这种情况，可以考虑的一种配准方式就是完全交互. 依据单应性条件，需交互确定四对点方能实现图像视角到参考平面的单应矩阵，在没有任何参考信息的情况下，点的确定会比较困难，且精度误差会比较大.

而本文所采用的场景信息辅助的交互方式，如图 7 中所示，在以画出的人行道线为参考的条件下，可以方便地找到特定的点(角点、中心点等)，从而使得确定的对应点具有较高的精度. 在整个交互过程中，用户只需在已标记直线上选取特定的两点即可，

《计算机学报》

所需确定点的数目也较少,简单易操作,精确度高于完全交互.

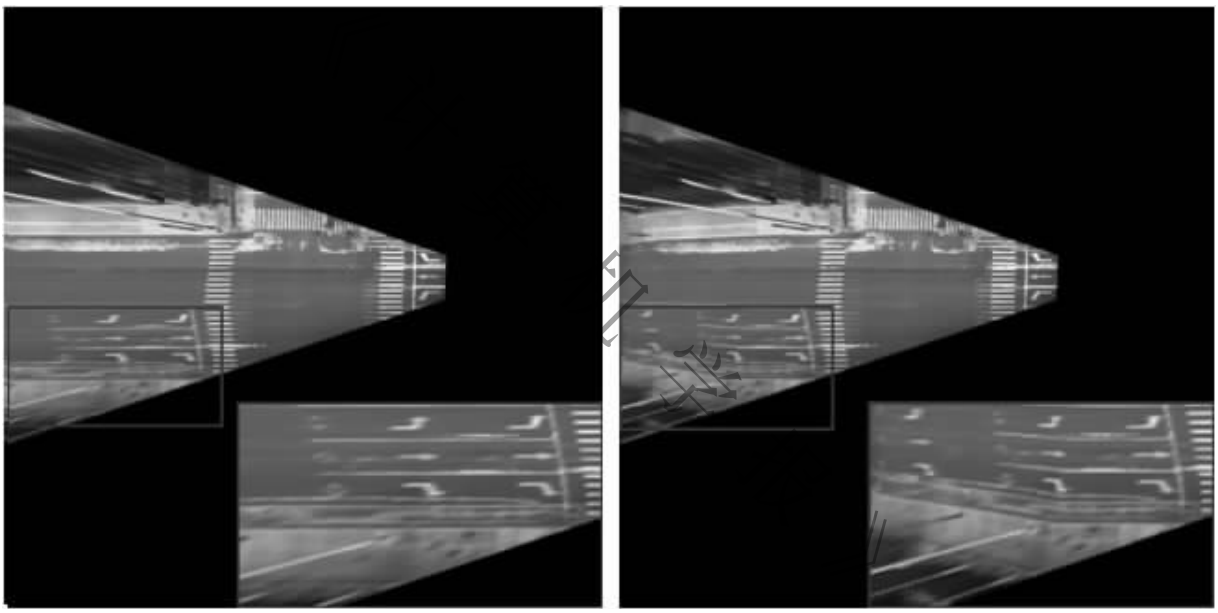
3.1.3 SPHP 保形

上述已得各方向图像成像平面到地平面的单应变换,若仅依靠单应变换进行图像变换,图像中远离相机的部分将会被剧烈拉伸变形,形状失真严重,各视角信息损失也较大(如图 8(a)).本文采用 SPHP 的思想修改原单应变换以保持拼接后图像形状和视角信息.

SPHP 算法是在不损失拼接精度的情况下,基于相似变换能够更好保持图像形状的原理提出的一种在目标/参考图像的重叠/非重叠区域从单应变换平滑转换为相似变换的拼接算法. Chang 等人[6]通过约束所有视角图像尽可能经历相似变换来计算单应变换到过渡变换以及过渡变换到相似变换的分界线 u_1 和 u_2,但本文所拼接背景图像视角差距过大,所以不适用.我们对每个视角图像单独进行 SPHP 变换.如图 9(b)中直线所示,为使重叠区域依然保持对齐,分界线 u_1 设置为各视角与其他视角重叠区域的最外侧,而 u_2 确定为拼接图像最外侧,使得变换后图像具有尽可能平滑的视觉效果.单个视角进行 SPHP 保形变换后结果如图 8(b)所示.

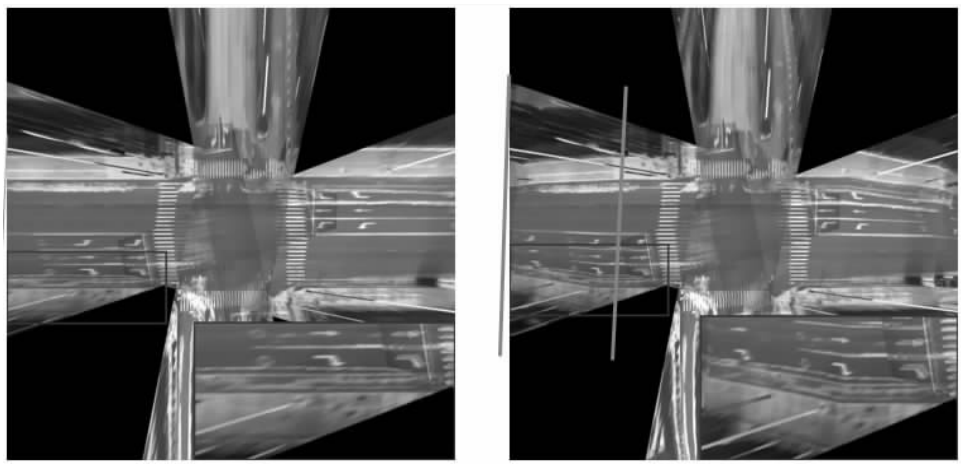
假设各视角依据上述方法得到的映射关系为 $\{\omega_i\}$, $i=1,2,\cdots,N_v$,其中 N_v 为视角数目.对视角 i 的背景图像 $\hat{I}_i$ 及其变换后图像后 I_i,则有 $I_i=\omega_i\cdot\hat{I}$.

对各视角变换后图像,进行线性融合即可得到拼接图像(如图 9(b)所示).



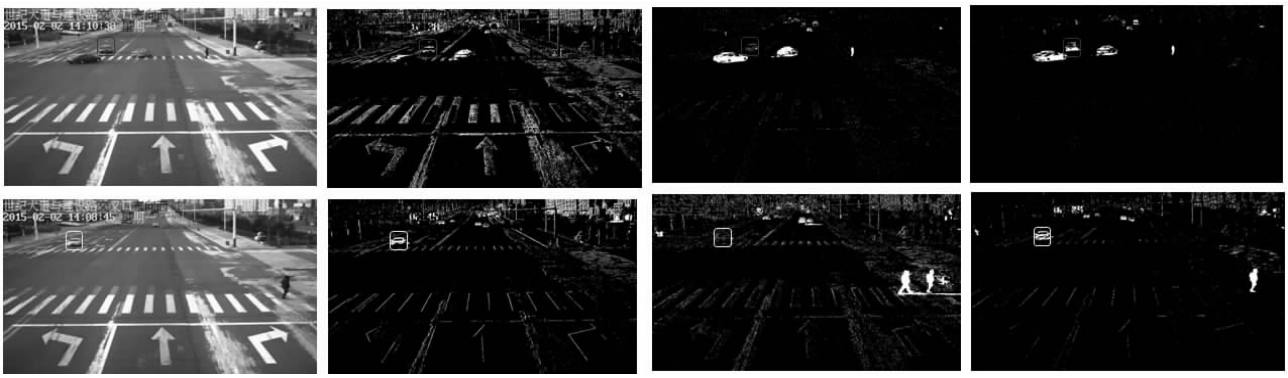
(a) 单应变换 (b) SPHP保形

图 8 进行单应变换后图像以及 SPHP 保形变换后图像(其中灰色框为局部放大效果)



(a) 仅单应变换 (b) SPHP保形

图 9 两种拼接方法的比较((a)为仅采用单应变换进行配准拼接;(b)对单应变换结果采用 SPHP 保形后进行拼接)



(a) 原始图像　(b) 帧差法　(c) 高斯混合模型　(d) ViBe

图 10　帧差法、高斯混合模型以及 ViBe 算法前景检测结果比较

3.2　目标检测

虽然背景图像已知，但是若直接采用简单帧差法进行前景提取，会对光照变化敏感，且噪声点过多，不利于检测，本文目标检测采用 Barnich 等人[7]提出的 ViBe 背景建模算法提取前景，能够逐帧调整背景模型，从而适应背景的变化，具有较强的抗噪能力.

ViBe 前景检测算法的基本思想是通过为每个像素存储一个样本集来建立背景模型，然后对当前帧的每个像素，将其与对应样本集进行比较来判断其属于前景还是背景，其中每个像素的样本集为其过去的像素值和其邻域像素值. 具体判别规则是若背景模型样本集中个数与当前像素接近的样本个数超过指定阈值，则判别为背景，否则为前景.

公式化表示为：记 $v(pt)$ 为图像像素点 pt 处的像素值；$M(pt)=\{v_1,v_2,\cdots,v_N\}$ 为 pt 处的样本集，N 为样本集大小；$SR(v(pt))$ 为以像素值 $v(pt)$ 为中心的半径为 R 的颜色区域. 则当 $M(pt)$ 满足 $SR(v(pt))\bigcap M(pt)>\sharp\min$ 时判断为背景点，$\sharp\min$ 为给定阈值，即背景点满足样本集中与当前帧 pt 点像素值的距离在 R 内的样本数大于 $\sharp\min$.

初始化过程中，背景减除的很多流行的算法都需利用相当多的视频帧数据来估计背景像素的时域分布，这样对突然的亮度变化就会很敏感. ViBe 算法则是仅利用单帧图像进行背景模型初始化，由于视频单帧不存在时域信息，其依据相邻像素的像素值具有相近时空分布特性的假设，对每个像素从单帧图像的空间邻域中随机选取像素填充入其背景模型. 这种算法对于本文已知背景图像的情况很适用.

ViBe 算法的更新过程是利用随机的思想，模拟像素变化的不确定性，采用随机子采样方式，随机选择像素样本且随机选择邻域像素进行更新. 设采样因子为 φ，即若当前点判断为背景点，有 $1/\varphi$ 的概率更新该点模型的样本值，且有 $1/\varphi$ 的概率更新其邻域像素的模型样本集，而样本集中选择要更新的样本，同样是采用随机选取的策略，样本集中每个样本有相同的概率被更新.

程序中取 $N=20$，$R=20$，$\sharp\min=2$，$\varphi=16$.

由于本文采用背景图像进行初始化，故除噪声外，由初始化过程前景被错误检测为背景而造成的鬼影现象很少. 另外，算法实现过程中对于模型更新过程本文直接采用保守的更新策略，使得前景点始终不会被用来填充背景模型，这样静止目标便不会融入背景.

相对于近年来广泛使用的高斯混合模型[15]，该算法对鬼影现象更鲁棒，且噪声少，计算量小，更高效. 但由于其依然是像素级的颜色背景模型，对阴影敏感，为提高目标提取的精确性，本文采用了 KaewTraKulPong 等人[36]的阴影检测方法进行阴影消除(如图 10(d)).

前景检测获得前景二值图后，便可对其进行区块分割获得每个目标，其中目标采用矩形框表示，目标位置为矩形框下边界的中点.

3.3　跨视角目标跟踪

本文跨视角目标跟踪采用简单的跟踪优先(track-first)的方法[1]，类似 Black 等人[32]提出的跟踪算法. 我们在每个视角利用 Kalman 滤波进行目标跟踪的同时在拼接的统一视域下同步进行 Kalman 跟踪并进行各视角数据的关联和融合.

Kalman 滤波中不管是各个视角还是拼接视域的目标状态均是目标在对应图像坐标系中的位置. 由于单个视角中相邻两帧中目标形态变化不会很大，因此在各个视角，我们利用其预测的位置信息结合其颜色直方图进行目标匹配. 当出现遮挡时，

Kalman 滤波器可预测目标的位置，当其再次可见时可继续跟踪，另外，当前视角遮挡的目标在其他视角可能是可见的，利用拼接视域目标的同步信息，可以矫正单个视角的跟踪结果，从而提高了跟踪对目标遮挡情况的鲁棒性，图 11 和图 12 分别是单个视角中出现车辆遮挡和人车遮挡的情况，图中可见，在遮挡前后目标颜色不变，可见算法对遮挡情况具有一定的鲁棒性.

图 11 车辆遮挡情况跟踪结果

图 12 人车遮挡情况跟踪结果

对于 Kalman 滤波器，本文设定各个相机视角下状态模型为

$$\boldsymbol{S}_L=[x\quad y\quad w\quad h\quad \mathrm{d}x\quad \mathrm{d}y]^{\mathrm{T}},$$

其中，$[x\quad y]$为目标在各视角图像坐标系中的位置，实验中设置为其矩形框下边界的中点；w 和 h 分别为对应目标矩形边界框的宽度和高度；$[\mathrm{d}x\quad \mathrm{d}y]$则为目标在图像坐标系中的速度.

转移矩阵为

$$\boldsymbol{A}_L=\begin{bmatrix}1&0&0&0&T&0\\0&1&0&0&0&T\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix},$$

测量矩阵为

$$\boldsymbol{H}_L=\begin{bmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\end{bmatrix}.$$

设定拼接视域下状态模型为

$$\boldsymbol{S}_G=[X\quad Y\quad \mathrm{d}X\quad \mathrm{d}Y]^{\mathrm{T}},$$

其中，$[X\quad Y]$为目标在拼接视域坐标系中的位置；$[\mathrm{d}X\quad \mathrm{d}Y]$则为目标在拼接视域图像坐标系中的速度.

转移矩阵为

$$\boldsymbol{A}_G=\begin{bmatrix}1&0&T&0\\0&1&0&T\\0&0&1&0\\0&0&0&1\end{bmatrix},$$

测量矩阵为

$$\boldsymbol{H}_G=\begin{bmatrix}1&0&0&0\\0&1&0&0\end{bmatrix}.$$

数据关联涉及各视角同一目标的判断，融合即是确定目标在拼接视域下的位置. 拼接视域平面与每个视角目标的一致性判断是利用背景拼接所得映射关系，即$\{\omega_i\}$，$i=1,2,\cdots,N_v$，采用轨迹匹配的方式进行决策，具体为：

设 Kalman 滤波得第 i 个视角第 j 个目标跟踪所得目标轨迹表示为 $\boldsymbol{T}_{ij}=\{P_{ij}^t\}$，$t=f,f+1,\cdots,c$. 其中 P_{ij}^t 表示目标第 t 帧的轨迹点，f 表示目标出现的第一帧，c 表示当前帧. 拼接视域平面用 s 表示，即拼接视域平面第 k 个目标轨迹表示为 $\boldsymbol{T}_{sk}=\{P_{sk}^t\}$. 与视角目标匹配的拼接视域目标满足轨迹的最小均方距离足够小(阈值 d_ξ)，即满足如下公式：

$$dist(\boldsymbol{T}_{ij}-\boldsymbol{T}_{sk})=\frac{1}{n_t}\sum_t\sqrt{(\omega_i P_{ij}^t-P_{sk}^t)^2}<d_\xi.$$

若存在多个，则取其中距离最近的目标，即

$$k'=\arg\min_k dist(\boldsymbol{T}_{ij}-\boldsymbol{T}_{sk}),$$

其中，n_t 两目标均存在的帧数，k'即为所求匹配目标.

《计算机学报》

这样使得匹配目标的运动轨迹尽可能一致，相对于仅采用当前位置进行匹配的方法，提高了目标匹配的准确性.

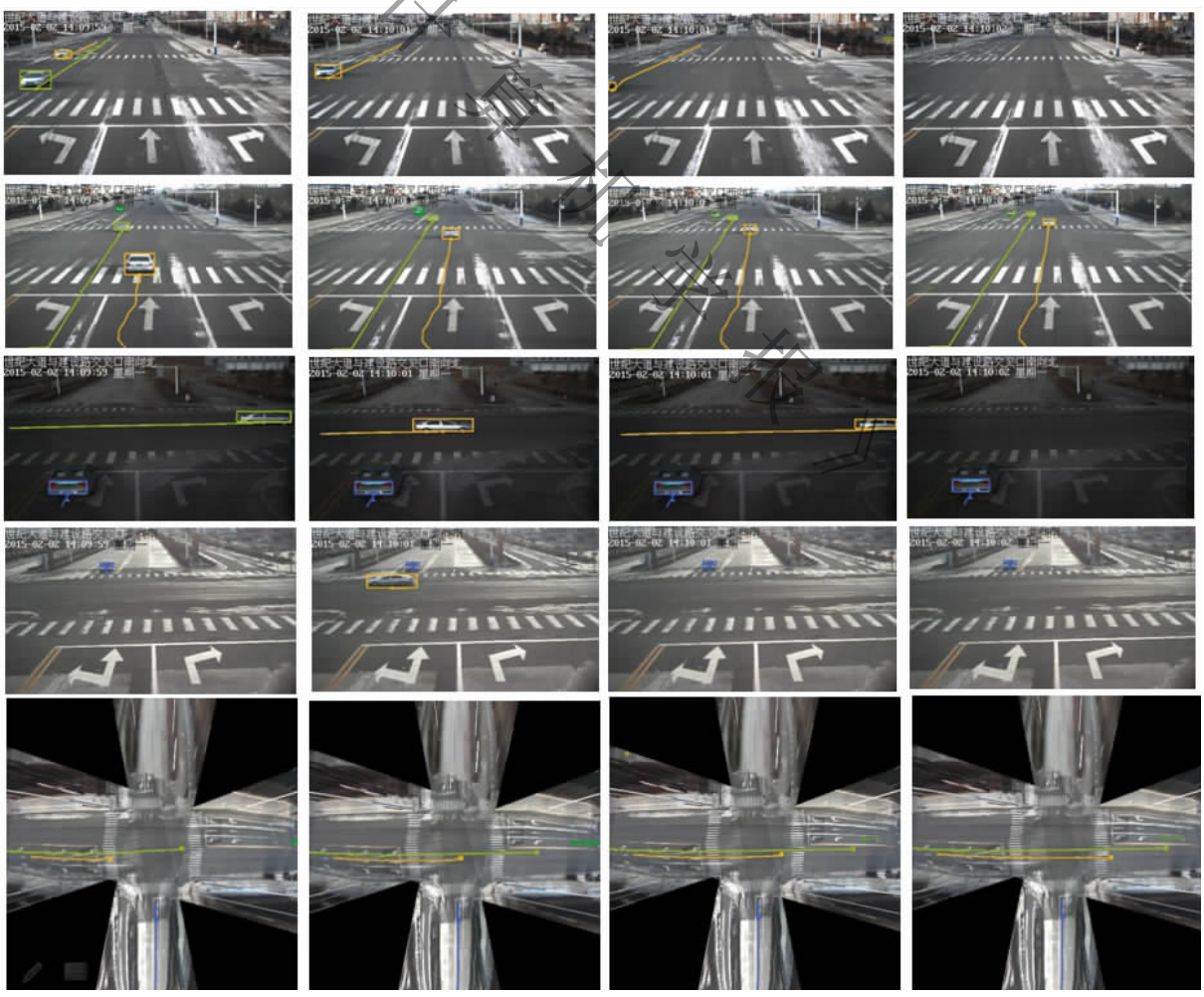
融合过程目标在拼接视域当前帧的位置的计算则是简单通过其对应的各视角目标位置的均值得到.

这种方法简单易于实现，但依然存在着局限. 由于各视角目标的匹配仅采用目标映射到拼接视域下的位置进行判断，虽然进行了保形变换，各视角映射到拼接视域时，远离相机的部分形变依然会比较严重，即离摄像机位置越远的位置不确定性越大，因此对目标检测的精确度的依赖就会越大. 远离相机处目标检测位置的微小差异可能引起在拼接视域下位置的很大差距，从而会使匹配出错. 而远离相机部分目标本身也不明显，检测过程难免出现检测的目标不完整，或者漏检的情况，所以越远出现误匹配的可能性便会较大. 例如，如图 13 中最后一行从右向左的目标，即图中第 2 行图片中远处的绿色目标，由于漏检，后续帧再次出现时已判断为新的目标. 可见特别是在远离相机部分，跨视域多目标跟踪的结果对目标检测的精度敏感度较高.

3.4　跨视角多目标跟踪的可视化

跟踪目标的可视化是在各视角视频中以及视域背景拼接所得到的拼接图像上，一致性展示各视角目标及其跟踪轨迹. 即需在拼接视域和各视角同步进行可视化显示.

拼接视域上目标可视化包括轨迹以及用方块表示的目标当前帧所在位置，其中每个目标采用一种颜色标志. 而各视角目标可视化包括轨迹和目标检测所得矩形边界框，为实现各目标在各视角的一致性标识，即使得在不同的视角中同一目标采用同一种颜色表示，各视角目标采用其对应拼接视域上目标的颜色进行绘制(如图 13 所示).



(a) 第746帧　(b) 第763帧　(c) 第770帧　(d) 第775帧

图 13　视频中若干帧目标跟踪及可视化结果(其中第 1 行至第 4 行是输入 4 个视角各帧目标跟踪轨迹，最后一行则是在拼接视域上目标的一致性跟踪轨迹)

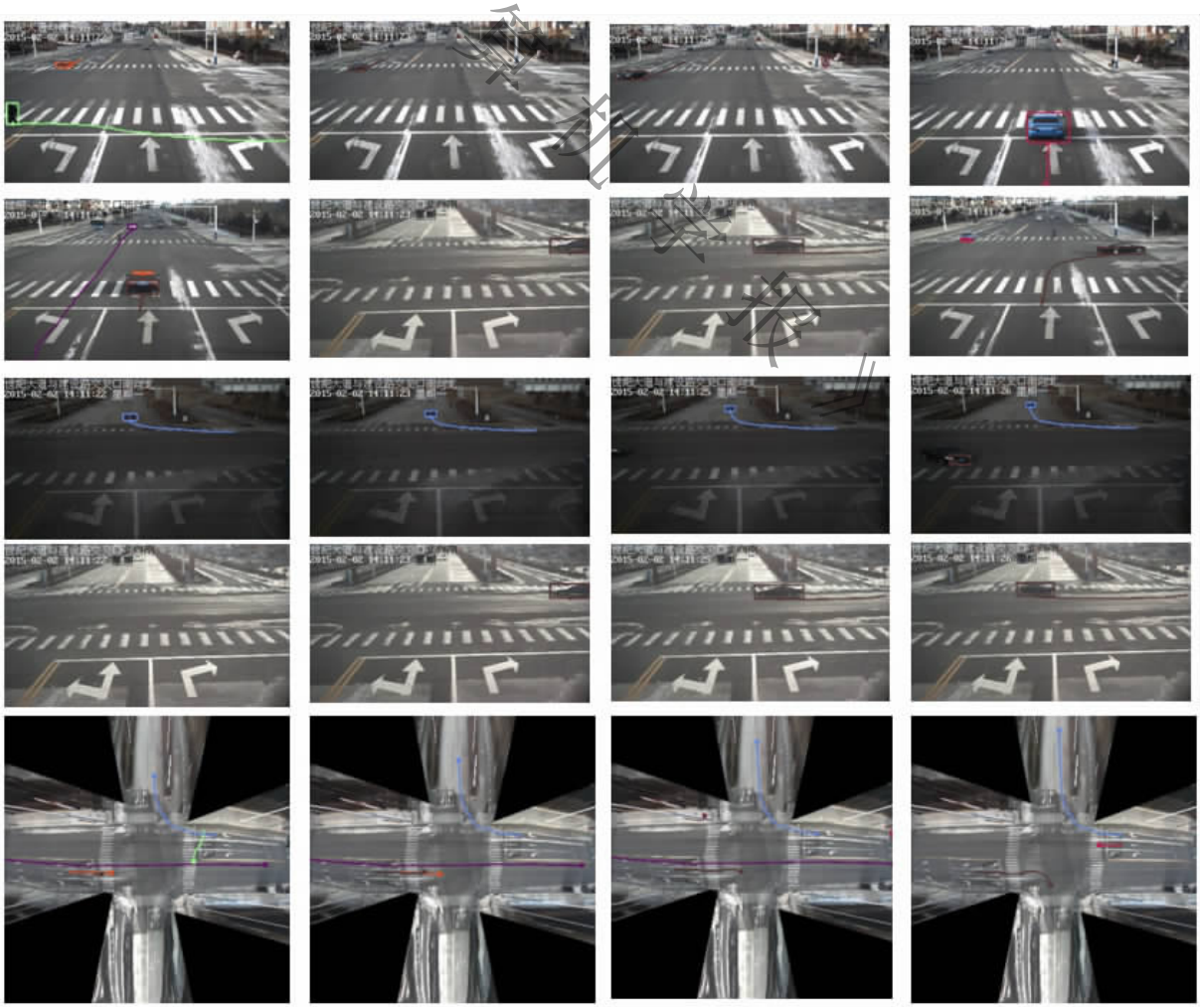
4 实验结果及分析

本文实验采用 Windows 8 下 Visual Studio 2013 结合 OpenCV 和 OpenGL 的软件环境，计算机配置为 Intel(R) Core(TM) i5-3470 CPU@3.20 GHz. 各视角初始背景图像可利用手工选取视频帧或者通过背景建模等方式得到. 下面为验证算法的有效性，从视域背景拼接、目标检测、跨视角多目标跟踪及其可视化三个部分展示和分析实验结果.

图 9 为本文视域背景拼接算法进行拼接的结果图像，可以看出本文的算法可以达到比较精确地拼接效果. 虽然由于相机镜头畸变，仍然存在着失真(图 9 中人行道)，但对于路口整体状况影响不大. 另外，图 9 中对仅采用单应变换的拼接结果以及单应变换结合 SPHP 进行拼接的结果进行了比较. 从图 9 中可见，采用 SPHP 变换进行拼接减少了单纯单应变换下的透视失真，能够更好地保持各自视角的信息，从而获得了更大范围的视域信息，便于后续跨视角目标跟踪及可视化.

本文目标检测中采用 ViBe 进行前景检测. 图 10 中对采用不同的检测算法进行前景检测的结果进行了比较，其中 ViBe 算法和高斯混合模型均为进行阴影去除后的结果，图中带颜色的框表示静止的目标. 可见高斯混合模型存在明显的丢失静止目标现象以及鬼影现象，而且噪点也较多，不利于目标的提取. 由于背景图像已知，帧差法虽然能够正确检测到静止目标，但是由于其背景是单帧静态图像，背景光照的变化使得噪点很多，另外，颜色与背景相近的目标也会丢失，如图 10 中的行人目标. 而由于本文对 ViBe 算法直接采用单帧背景图像进行初始化，可以解决鬼影现象以及静止运动目标的情况，相对高斯混合模型其具有更高的检测准确率(图 10(d)).

图 13 和图 14 为在 4 个大小为 1280×720 的不同视角视频进行跨视角多目标跟踪的可视化结果，其中相同颜色代表同一目标. 其中图 13 为目标进行



(a) 第1580帧 (b) 第1590帧 (c) 第1605帧 (d) 第1617帧

图 14 目标转弯情况的跟踪结果

直线运动的跟踪情况，从图中可见实现了两种颜色目标的持续跟踪，而图 14 即为目标转弯的跟踪结果，记从上之下前 4 行分别为视角 1～4，则可见其中灰色目标在 1580 帧中仅出现在视角 2 中，在 1590 帧时其移出视角 2，进入视角 4，并在 1605 帧及 1617 帧时依然能正确连续跟踪，可见实现了该目标的连续跟踪. 实验结果采用 CLEAR[37] 指标评估跟踪性能，其包括 MOTP 和 MOTA 两部分，其中 MOTP 评估目标位置的精确度，MOTA 评估目标跟踪的准确度. 跟踪中设定距离阈值 $d_\xi=100$，位置精度评估利用 d_ξ 归一化. 对前 1000 帧跟踪结果进行统计后计算得 MOTP 值约为 83.2%，MOTA 值约为 83.9%. 可以看出，本文算法能够有效进行跨视角目标的连续跟踪和可视化.

在计算时间方面，视域拼接部分 1280×720 的 4 个图像映射关系获取需约 8.7 s，其中单应性计算 1.4 s，4 个图像进行变换和融合少于 5 s. 跟踪和可视化部分，为提高检测精度，得到比较好的可视化效果，在目标检测部分耗时较多. 实验中为提高程序的实时性，对每帧图像的 ViBe 算法执行中采用了多个线程进行分块计算的方式来提高计算效率. 平均每帧目标检测部分耗时 0.326 s，跟踪和可视化耗时 0.033 s，程序中包含目标检测的整个可视化跟踪过程平均性能为 2.3 fps.

5　结　论

本文针对单个摄像机视角视域有限，不能实现大范围、长时间目标跟踪的问题，设计实现了一种基于视域拼接的跨视角目标跟踪的可视化算法. 其利用图像中的几何信息辅助交互进行单应矩阵的计算，同时利用 SPHP 算法以保持各视角信息，实现视域背景的拼接；然后采用 ViBe 前景检测算法结合阴影检测进行目标检测；最后利用 Kalman 滤波同时在各个视角和拼接视域同步进行多目标跟踪来提高跟踪性能，从而实现了在背景图像上的跨视角多目标跟踪的可视化显示. 实验结果表明，本文算法能够有效实现跨视角多目标跟踪的可视化显示，为智能监控提供了有力的工具. 但是，本文算法中依然存在着局限和不足，由于视域背景拼接中需采用交通路口场景中的信息进行视域拼接，因此本文算法仅对交通路口监控视频有效；另外关于同一目标的判断和融合，本文仅采用目标在拼接视域下位置判断同一目标，拼接视域下目标也是仅采用方框表示，未来工作中可以利用更复杂的方式进行目标的匹配和判断，以获得更精确的匹配效果，并进一步实现大视差目标的融合，从而在拼接视域下得到更真实更直观的可视化效果.

参　考　文　献

[1] Taj M, Cavallaro A. Multi-view multi-object detection and tracking//Cipolla R, Battiato S, Farinella G M Eds. Computer Vision. Berlin Heidelberg, Germany: Springer, 2010: 263-280

[2] Liu L, Xing J, Ai H. Multi-view vehicle detection and tracking in crossroads//Proceedings of the 1st Asian Conference on Pattern Recognition. Beijing, China, 2011: 608-612

[3] Leal-Taixe L, Pons-Moll G, Rosenhahn B. Branch-and-price global optimization for multi-view multi-target tracking//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 1987-1994

[4] Hofmann M, Wolf D, Rigoll G. Hypergraphs for joint multi-view reconstruction and multi-object tracking//Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. Portland, USA, 2013: 3650-3657

[5] Thingom C, Yeon G. Multi-view video summarization. IEEE Transactions on Multimedia, 2010, 12(7): 717-729

[6] Chang C H, Sato Y, Chuang Y Y. Shape-preserving half-projective warps for image stitching//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 3254-3261

[7] Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724

[8] Kalman R E. A new approach to linear filtering and prediction problem. Journal of Basic Engineering, 1960, 82(1): 35-45

[9] Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(11): 1330-1334

[10] Jiang Ming-Xin, Wang Hong-Yu, Liu Xiao-Kai. A multi-target tracking algorithm based on multiple cameras. Acta Automatica Sinica, 2012, 38(4): 531-539(in Chinese)
(姜明新，王洪玉，刘晓凯. 基于多相机的多目标跟踪算法. 自动化学报，2012，38(4)：531-539)

[11] Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Computer Vision & Image Understanding, 2014, 122(5): 4-21

[12] Brutzer S, Hoferlin B, Heidemann G. Evaluation of background subtraction techniques for video surveillance//Proceedings of the IEEE Conference on Computer Vision and Pattern

《计算机学报》

Recognition. Colorado Springs, USA, 2011: 1937-1944

[13] Szwoch G, Ellwart D, Czyżewski A. Parallel implementation of background subtraction algorithms for real-time video processing on a supercomputer platform. Journal of Real-Time Image Processing, 2016, 11(1): 111-125.

[14] Friedman N, Russell S. Image segmentation in video sequences: A Probabilistic Approach//Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence. Providence, USA, 1997: 175-181

[15] Zivkovic Z. Improved adaptive gaussian mixture model for background subtraction//Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK, 2004: 28-31.

[16] Zivkovic Z, Heijden F V D. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters, 2006, 27(7): 773-780

[17] Yao J, Odobez J M. Multi-layer background subtraction based on color and texture//Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition. Minneapolis, USA, 2007: 1-8

[18] Wang H, Suter D. Background subtraction based on a robust consensus method//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China, 2006: 223-226

[19] Wang H, Suter D. Sacon: A consensus based model for background subtraction. Technology Report: MECSE-15-2005, Monash University, Australia, 2005

[20] Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, USA, 2012: 38-43

[21] Kuo C H, Nevatia R. Robust multi-view car detection using unsupervised sub-categorization//Proceedings of the IEEE Workshop on Applications of Computer Vision. Snowbird, USA, 2009: 1-8

[22] Wen X, Shao L, Fang W, et al. Efficient feature selection and classification for vehicle detection. IEEE Transactions on Circuits & Systems for Video Technology, 2015, 25(3): 508-517

[23] Wu B, Nevatia R. Detection and Tracking of Multiple, Partially occluded humans by bayesian combination of edgelet based part detectors. International Journal of Computer Vision, 2007, 75(2): 247-266

[24] Dalal N, Triggs B. Histograms of oriented gradients for muman detection//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005: 886-893

[25] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 32(9): 1627-45

[26] Oh S, Russell S, Sastry S. Markov chain monte carlo data association for multi-target tracking. IEEE Transactions on Automatic Control, 2009, 54(3): 481-497

[27] Benfold B, Reid I. Stable multi-target tracking in real-time surveillance video//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 3457-3464

[28] Okuma K, Taleghani A, Freitas N D, et al. A boosted particle filter: multitarget detection and trackings//Proceedings of the European Conference on Computer Vision. Prague, Czech Republic, 2004: 28-39

[29] Vermaak J, Doucet A, Prez P. Maintaining multi-modality through mixture tracking//Proceedings of the IEEE International Conference on Computer Vision. Nice, France, 2003, 2: 1110-1116

[30] Breitenstein M D, Reichlin F, Leibe B, et al. Online multi-person tracking-by-detection from a single, Uncalibrated Camera. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(9): 1820-1833

[31] Henriques J F, Caseiro R, Batista J. Globally optimal solution to multi-object tracking with merged measurements//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 2470-2477

[32] Black J, Ellis T, Rosin P. Multi view image surveillance and tracking//Proceedings of the Workshop on Motion and Video Computing. Orlando, USA, 2002: 169-174

[33] Reinhard E, Ashikhmin M, Gooch B, et al. Color transfer between images. IEEE Computer Graphics & Applications, 2001, 21(5): 34-41

[34] Chen Ning-Fan, Cai Li-Dong. An vanishing points detection algorithm based on clustering. Chinese of Stereology and Image Analysis, 2006, 11(1): 49-52(in Chinese)
(陈宁凡, 蔡利栋. 一种基于聚类的消隐点自动测量方法. 中国体视学与图像分析, 2006, 11(1): 49-52)

[35] von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: A fast line segment detector with a false detection control. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(4): 722-732

[36] KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for realtime tracking with shadow detection//Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance System. London, UK, 2002: 135-144

[37] Bernardin K, Stiefelhagen R. Evaluating multiple object tracking performance: The CLEAR MOT metrics. EURASIP Journal on Image & Video Processing, 2008, 2008(1): 1-10

《计算机学报》

LIU Cai-Hong, born in 1992, M. S. candidate. Her research interests focus on computer graphics and visualization.

ZHANG Lei, born in 1981, Ph. D., associate professor. His research interests focus on computer graphics.

HUANG Hua, born in 1975, Ph. D., professor. His research interests focus on image and video processing.

Background

Because of the limited field of view of a single camera, it's impossible to meet the exact requirement for long distance object tracking. To solve this problem, we propose an algorithm for visualization of multi-view multi-object tracking based on view stitching for surveillance videos in crossroad. And we can achieve the goal that visually display the objects tracked in a common wider field of view which is stitched by the views of several camera. This is the first time we do something for this research subject. And to our knowledge, there is no such research before.

This research is related to quite a few research fields, mainly includes image stitching, object detection and multi-view multi-object tracking. Then we will introduce these three fields one by one.

At first, for image stitching, it's a popular research topic in recent years. The algorithm is much mature today and the result stitched images is more and more natural with fewer and fewer artifacts. But the proposed algorithms are mainly based on feature point detection and matching, it's still not applicable in the case like our research that the parallax between images is too big, even to 90 degrees, the overlapping area is limited and feature information is too little.

Second, for object detection, the method most commonly used is still background modeling and foreground detection, such as the Gaussian mixture model and ViBe background modeling algorithm and so on. The Gaussian mixture model is still the most popular algorithm for background subtraction. But there are still other methods that get specific object model by training a lot of sample data. Comparing to the background subtraction, these methods can get more accurate detection result, but there are still limitation exists. First, these methods usually need a lot of data to train. Second, they usually only can train models for specific kind of object. They are not universal for every kind of object.

The last is the multi-view multi-object tracking. It mainly contains two kinds of methods: track-first method and fuse-first method. Track-first methods perform tracking in each camera and then project and link the resulting information on other views. And the fuse-first methods project detection information from each view to a common view and apply tracking. In recently years, the research has focused on the tracking-by-detection principle. In other words, they assumed the detection information has obtained and tracking becomes a data association problem. And recently they usually use the detection information to construct a graph, and transform the problem to be an optimization problem to solve.

《计算机学报》