

基于多边形顶点单应映射的多摄像机前景融合实时运动目标检测*

任 劼¹, 许 铭², 赵慧民³

- (1. 西安工程大学电子信息学院, 陕西 西安 710048;
2. 西交利物浦大学电气与电子工程系, 江苏 苏州 215123;
3. 广东技术师范学院电子与信息学院, 广东 广州 510665)

摘 要: 采用多摄像机多平面单应前景映射融合的方法可以减小目标遮挡所造成的影响并提高目标检测的准确性和可靠性。采用传统的前景图单应变换是一个图像级的运算, 需要将前景图中的全部像素逐一投影到参考视角中。为了消除透视效应产生空洞区域, 还需根据分辨率较高的参考视角反映射的方法确定前景投影图。由于计算量较大, 处理无法应用到实时目标检测中。该文提出了一种前景轮廓多边形顶点映射的方法, 首先对在单摄像机中检测出的前景进行多边形逼近, 只对前景多边形的顶点进行单应映射, 最后在参考视角中根据映射顶点进行前景重构, 提高前景映射的运算速度以满足多摄像机前景融合实时目标检测。实验中, 多边形顶点法与传统法、轮廓法、矩形框法进行了比较。实验表明, 重构的前景映射图不仅可以有效地逼近图像级单应变换的前景映射图, 而且在运动目标大小不同的情况下, 多摄像机前景融合目标检测的运算速度可以分别提高约12和69倍。

关键词: 运动目标检测; 单应; 多摄像机

中图分类号: TP391 **文献标志码:** A **文章编号:** 0529-6579(2016)04-0030-09

Real-time object detection with foreground fusion from multiple cameras using homography mapping of polygon vertices

REN Jie¹, XU Ming², ZHAO Huimin³

- (1. College of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China;
2. Department of Electrical and Electronic Engineering, Xi'an Jiaotong Liverpool University, Suzhou 215123, China;
3. School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: Multi-camera and multi-plane foreground fusion approach can relieve the effects of occlusion and improve the accuracy and robustness of moving object detection. The traditional homography mapping is an image-level transformation which projects each pixel in the binary foreground image into a reference view. To avoid perspective openings or holes which are generated during the mapping from the camera view to the top view, the number of the pixels in the homography mapping is decided by the resolution of the top view, which is usually higher than that of the camera view. The slow speed has dissuaded the foreground homography mapping from real-time applications. A foreground polygon approximation method

* 收稿日期: 2016-02-19

基金项目: 国家自然科学基金资助项目(60975082); 陕西省教育厅科研计划项目(15JK1310); 广东省自然科学基金资助项目(2015A030313672, 2016A030311013); 广东省教育厅省级重大资助项目(2014KZDXM060); 广东省普通高校国际合作重大项目(2015KGJHZ021)

作者简介: 任劼(1984年生), 女; **研究方向:** 机器视觉视频监控; E-mail: renjie_xpu@126.com

is proposed. After the foreground regions are identified in a camera view, each foreground region is approximated by a polygon and only the vertices of the polygon are projected to the reference view through homography mapping. Then the projected foreground region, which is rebuilt in the reference view, is utilized in real-time moving object detection with multiple cameras. To evaluate the performance, the proposed polygon approximation method has been compared with the contour based method and the bounding-box based method. The experimental results have shown that the proposed algorithm can produce competitive results in comparison with those using foreground bitmap mapping. Considering the differences of moving objects' size, the processing speed is about 12 and 69 times faster than the bitmap mapping method.

Key words: object detection ; homography ; multi-camera

近年来, 加速转型的中国社会所面临的越来越多的不稳定因素, 以及社会治安防范的日益复杂, 视频监控需求呈快速而稳定的增长。传统的实时监控方式需要工作人员在布满监视器的监控室中工作并对特殊事件做出预警, 效果很大程度上依赖于当前监测者的主观意识、反应速度和疲劳程度。智能视频监控作为人工智能和机器视觉的一个研究方向, 引起了广泛的关注。智能视频监控系统的目标就是可以自动并实时完成对目标的检测、跟踪、分类和识别, 对可能发生的异常情况发出预警信号, 以帮助监控人员快速做出应对, 对存储的各种视频进行快速恢复、寻找查询和分析^[1-2]。

根据所采用的摄像机的数目, 智能监控系统可分为单摄像机和多摄像机两类。如果采用单摄像机, 典型的视频监控系统需要承担的工作分为运动物体检测、目标跟踪、目标分类和特殊事件检测。视频序列中的运动物体可通过前景检测的办法逐帧进行检测并利用时空关系进行跟踪。被跟踪目标的特征信息在被提取后, 利用设定的特殊事件触发条件进行判断, 当特殊事件发生时将会产生预警信号。

当运动物体在屏幕中是单一无遮挡时, 单—摄像机可以轻松完成上述各种工作。在现实应用中, 来自其他运动物体的动态遮挡和来自树木、建筑物等障碍物的静态遮挡会严重影响智能监控的结果。考虑到当一个运动目标在一个摄像机中被其它物体遮挡时, 在其它不同位置的摄像机中观测时可能不会产生遮挡现象, 利用多摄像机协同监测可以有效地缓解遮挡带来的影响。由于多摄像机可被安置在不同的地点以检测不同的区域, 采用多摄像机还可以扩大监测区域。基于多摄像机视频监控的优越性, 对多摄像的智能视频监控系统的研究有着重要的现实和理论意义。

1 文献综述

根据多摄像机信息融合的程度, 国内外研究多摄像机目标检测和跟踪的文章可被划分为低度融合、中度融合和高度融合三类。

1.1 低度融合

采用低度信息融合方法的视频监控系统主要是采用多摄像机切换的办法来解决单一摄像机监测时视野受限的缺陷^[3]。具体办法是首先利用某一摄像机进行目标检测与跟踪, 当监测到当前视角在目标的运动过程中不再具有良好的检测效果或目标离开当前摄像机的视角范围时, 需要切换到另一台摄像机进行检测与跟踪。在这一类方法中, 目标的检测和跟踪是由不同的摄像机完成的, 在同一时间只有一个摄像机负责监控, 摄像机之间几乎没有信息融合。所以多摄像机切换的方法仅仅是增加了视角覆盖范围而无法缓解由于遮挡所带来的影响。

1.2 中度融合

当同一运动物体在不同摄像机中的对应关系被确定后, 在不同摄像机中检测到的运动轨迹或运动物体的观测信息就可以被融合到参考视角以提高检测和跟踪的可靠性。研究者们常采用中心投影^[4-5]、极线法和单应变换法等几何约束的方法来建立不同摄像机之间的对应关系^[6-10]。

采用运动轨迹融合的方法中, Khan 和 Shah^[11]在建立了不同摄像机之间的对应关系后, 将各个摄像机视角中所检测到的目标的运动轨迹映射到一个虚拟的顶视图平面进行融合, 以提高目标检测和跟踪的准确性。这类方法的缺陷是在一个视角的跟踪结果不会被用来提高其他视角中跟踪的可靠性。

改进的算法中目标跟踪不再由单一摄像机执行, 而是先将各个摄像机中所检测到的特征信息投射至参考视角并融合, 然后在参考视角内利用融合后的信息进行全局的目标检测与跟踪。各摄像机视

角中所获取的检测特征大致可以分为: 点^[6]、线^[8,12]、区域^[13-14]。Mittal and Davis^[5]利用前景区域和对极线相交的线段的颜色信息寻找匹配区域。Hu 等^[8]用主轴的方法可以缓解前景检测误差所带来的影响。

1.3 高度融合

Fleuret 等^[15]将地平面划分成了小的栅格, 每个小栅格的大小由该摄像机中行人所对应的矩形区域的平均大小决定; 随后根据每个栅格中运动目标像素所占的比例计算出了地平面占有率图。Khan 和 Shah^[16]将各单视图的前景似然率图通过基于地平面的单应变换投射至一个参考视角, 然后在参考视角中对其进行叠加, 并利用设定的阈值在参考视角中确定行人在地面的位置。随后, 基于地平面的单应变换被拓展至一系列在一定高度且平行于地平面的虚拟平面的单应变换, 以提高目标检测的可靠性。文献 [17] 同样采用了包括地平面在内的 3 层单应平面通过多摄像机的前景似然率图融合对行人进行定位, 实现稳定跟踪。Eshel 和 Moses^[18]将不同摄像机的各像素根据头顶平面的单应变换映射至虚拟的头顶平面, 然后根据投射至同一点的来自不同摄像机的各像素的灰度信息建立超级像素, 利用超级像素的匹配结果来确定运动目标的位置。Peng 等^[19]将不同摄像机视角中检测出来的前景区域根据地平面映射至虚拟顶视图中进行融合后, 将相交区域根据行人高度所决定的虚拟平面反映射回各摄像机视角, 利用贝叶斯网络进行检测和跟踪。

传统的单应映射方法是一种图像级的变换, 摄像机视图中的每个像素需通过单应变换被逐一投射至参考视角中。为了避免在参考视角图像中由于透视效应产生空洞区域, 一般采用将参考视角中的每一个像素反映射回摄像机视图, 如果反映射的像素点在摄像机视图中, 则用摄像机视图中的像素值填充参考视角中对应像素点的值。这样, 所需要进行的映射变换的像素的数量需由参考视角图像的分辨率来确定。在实际应用中, 常作为参考视角的顶视图由于要覆盖较大的区域通常具有更高的分辨率, 使这种逐像素的映射在计算和传输过程中所需的时间较长, 造成其难以适用于实时检测中。如果采用多摄像机以及多个高度平面单应变换融合的算法, 所需要的时间将会成倍增加。

为了改进上述缺陷, 本文提出了一种基于多边形近似逼近算法的多层单应变换融合算法用于多摄像机目标实时检测, 并采用了两组不同的序列对所提出的算法进行了有效性和可靠性的评估。在与传

统的图像级变换法、矩形框法和轮廓法进行比较后, 实验结果表明所提出的方法在保证可靠性的同时, 除了可以有效地降低所需要的时间外, 还具有时间及精度可控的优点。

2 单应估计

单应关系是基于同一平面 (单应平面) 的两个视角图像之间的映射关系。通过选取其中的一幅图像作为参考视角图像, 可以将另一幅图像投射至参考图像进行图像融合。除了可以选取摄像机捕获的图像作为参考图像外, 还可以选取合成俯视图或空中俯视图作为参考图像以获取更大的视野范围。设 (u_c, v_c) 是在摄像机视角 c 下图像中单应平面上的一点, 其在顶视图中的对应点为 (u_t, v_t) , $\mathbf{p}^c = [u_c, v_c, 1]^T$ 和 $\mathbf{p}^t = [u_t, v_t, 1]^T$ 是这两点的其次坐标表示, 则它们之间的单应变换可表示为:

$$\mathbf{p}^t \cong \mathbf{H}^{c,t} \mathbf{P}^c \quad (1)$$

$$\mathbf{H}^{c,t} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad (2)$$

其中, 符号 \cong 在单应变换中存在一个缩放因子。 \mathbf{P}^c 和 $\mathbf{H}^{c,t}$ 中的上标 c 为摄像机的标号, $c = a, b$ 。由于在本文中采用了虚拟的顶视图作为参考视角图像, \mathbf{p}^t 和 $\mathbf{H}^{c,t}$ 中的上标 t 指顶视图。两幅图之间的单应关系矩阵可根据两幅图像中相关的点、线或椭圆等特征的对应关系计算。本文采用了一系列手工选取的对应点进行摄像机校准, 从而求出单应矩阵。

2.1 地平面单应矩阵估计

由于单应变换是投影变换在一定约束条件下的简化形式, 所以在对摄像机进行标定后, 摄像机的投影矩阵可被用来直接计算基于地平面的单应矩阵。设 $\mathbf{p}^w = [X_w, Y_w, Z_w, 1]^T$ 是三维空间中的一个点, \mathbf{p}^w 映射到图像中的对应点 \mathbf{P}^c 的关系可通过一个投影矩阵 \mathbf{M} 表示:

$$\mathbf{P}^c = \mathbf{M} \mathbf{p}^w = [m_1, m_2, m_3, m_4] \mathbf{P}^w \quad (3)$$

其中 m_1, m_2, m_3, m_4 各是一个 3×1 的向量, \mathbf{M} 是由摄像机标定所得的内外参数构成的 3×4 矩阵。

当 \mathbf{p}^w 和 \mathbf{P}^c 为地平面上一点时, \mathbf{P}^w 在三维空间中可以表示为 $\mathbf{p}_0^w = [X_w, Y_w, 0, 1]^T$, 其中 $Z_w = 0$, \mathbf{p}_0^w 的下标 0 表示地平面。假设有一个由虚拟摄像机产生的顶视图, \mathbf{p}_0^w 在该图中的对应点为 $\mathbf{p}_0^t = [u_t, v_t, 1]^T$, 其中 $u_t = X_w, v_t = Y_w$, 则基于地平面的单应矩阵 $\mathbf{H}_0^{c,t}$ 可由 \mathbf{M} 矩阵降维直接获得:

$$H_0^{t,c} = (H_0^{c,t})^{-1} = [m_1, m_2, m_4] \quad (4)$$

$$P_0^c = H_0^{t,c} p_0^t \quad (5)$$

$$p_0^t = H_0^{c,t} P_0^c \quad (6)$$

2.2 多平面单应矩阵估计

在单应变换中不仅可以采用地平面作为单应平面, 选取平行并高于地平面的一组虚拟平面作为单应平面时, 图像之间的单应变换依旧存在。当摄像机标定后, 多平面单应关系矩阵依旧可以从摄像机投影矩阵直接产生。当不采用摄像机标定时, 多平面单应关系矩阵可根据地平面单应矩阵和垂直方向的消逝点^[16]、四个共线的点的交比关系或4个杆子上具有不同高度标示点求得^[17-18]。除此之外, 多平面单应矩阵还可以通过一系列在头顶和脚底点之间根据高度比例线性内插的点求得。

设 $p_h^w = [X_w, Y_w, Z_w, 1]^T$ 为三维空间中平行并且距离地平面高度为 h 的虚拟平面上的一点, 其中 $Z_w = h$ 。 p_h^w 在头顶的虚拟摄像机所产生的顶视图中的对应点为 $p_h^t = [u_t, v_t, 1]^T$, 其中 $u_t = X_w$, $v_t = Y_w$ 。根据公式(3), p_h^w 在摄像机视角 c 中的对应点可表示为 $p_h^c = [m_1 u_w + m_2 v_w + m_3 h + m_4]^T$ 。由于 $m_3 h + m_4$ 由于中的每个元素都是已知量, 则由顶视图中点 p_h^t 根据高度为 h 的单应平面映射到摄像机视角 c 中点 p_h^c 的映射关系可表示为:

$$p_h^c = H_h^{t,c} p_h^t = [m_1 \quad m_2 \quad m_3 h + m_4] p_h^t \quad (7)$$

对比公式(4)和(7)可知, 基于 h 高度平面的单应矩阵可由基于地平面的单应矩阵与投影矩阵 M 中的第三列元素乘以给定的高度 h 求得:

$$H_h^{t,c} = H_0^{t,c} + [o \mid h m_3] \quad (8)$$

其中, $[o]$ 是一个 3×2 的全零矩阵^[20]。

3 单摄像机前景检测与表示

3.1 前景检测

前景提取的目的是根据背景图像将运动的物体从每帧图像中提取出来, 其提取效果主要由背景模型决定。混合高斯模型可以有效地抑制如树叶摆动等背景元素的轻微变化, 被广泛地应用在前景提取中^[21]。本文采用 K 个高斯分布对每一个像素的彩色值进行建模, 设一个像素在时刻 t 的颜色取值为 $p_t = [R_t, G_t, B_t]^T$, 该像素取该值的概率为:

$$P(p_t) = \sum_{j=1}^K \frac{w_j}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}(p_t - u_j)^T \Sigma_j^{-1} (p_t - u_j)} \quad (9)$$

式中, d 为像素彩色值的维度 (通常 $d=3$), w_j 是权重值, u_j 和 Σ_j 是第 j 个分布的均值和协方差矩阵,

并用 σ_j^2 表示 Σ_j 的迹。 K 个分布根据 w_j/σ_j^2 进行排列, 出现频率高并且方差小的分布被排在前面。最终选取满足权重和小于阈值 Th_w 的前 B 个分布作为背景模型。

$$B = \operatorname{argmax}_b \left(\frac{\sum_{j=1}^b w_j}{\sum_{j=1}^K w_j} > Th_w \right) \quad (10)$$

在 t 时刻, 图像帧 I_t 中的每个像素 $I_t(r, c)$ 与背景模型进行比较, 如果其值与 B 个分布的中心的差值都大于该分布标准差的 Th_B 倍, 则该像素被判断为一个前景像素。 Th_B 可根据经验选取, 通常为 2.5, 且 $j \in [1, B]$ 。

$$F_t = \{ (r, c) : \| I_t(r, c) - \mu_{t-1,j}(r, c) \| > Th_B \sigma_{t-1,j}(r, c) \} \quad j \in [1, B] \quad (11)$$

如果一个像素被判断为背景像素, 即它能与已有的 K 个分布中的一个匹配上, 则需要对与其匹配的分布进行更新。如果背景像素与已有的 K 个分布中都匹配不上, 则需建立新的分布, 并在对权重更新后重新对各分布进行排序。检测出的前景像素在经过连通成分分析、形态学运算、尺寸滤波后形成前景区域。

3.2 前景多边形逼近表示

当前景区域被提取出来后, 需通过基于某个平面的单应变换投射至参考视角中。与传统的需要对图像中每个像素进行映射的图像级变换不同, 本文提出了一种只对前景区域多边形的顶点进行映射的近似算法, 能减少所需映射像素个数或降低计算所需的时间。设 F_i^a 是在摄像机视角 a 中检测出的第 i 个前景区域, 采用 Suzuki^[22] 提出的方法进行轮廓提取, 其轮廓可用 N 个点的集合 C_i^a 表示。为了使轮廓的点集更加灵活可控, 本文采用了 Douglas-Peucker (DP) 算法对前景轮廓进行多边形逼近表示^[23], 所提取的多边形顶点集合用 V_i^a 表示。方法具体如下:

1) 选取 C_i^a 中具有最大距离的两个点作为起始点, 在 C_i^a 中删除这两点的同时将其加入 V_i^a 中, 根据这两点确定一条直线;

2) 根据 V_i^a 中的两点确定一条直线, 计算剩余 C_i^a 中各点与已知直线的距离, 选取距离最大的点。如果该点对应的距离大于调节参数 ε , 将该点从 C_i^a 中删除并加入 V_i^a 中;

3) 根据 C_i^a 中的点根据顺序确定直线, 计算每条直线的两个端点之间的轮廓点距离该直线的距离, 将距离最大且大于调节参数 ε 的点从 C_i^a 中删除并加入 V_i^a 中;

4) 重复步骤 3) 直到距离最大的点所对应的距离小于调节参数 ε 为止。最终得到参数 ε 控制的多边形顶点集合 V_i^a 。

4 基于地平面的多摄像机前景融合

4.1 多边形映射

为了提高单应变换的计算效率, 根据由公式 (4) 计算出的由摄像机视角 a 到虚拟的顶视图 t 的基于地平面的单应矩阵 $H_0^{a,t}$, 每个前景区域近似多边形的顶点被投射至虚拟的顶视图中, 映射后的顶点 $V_{i,0}^{a,t}$ 为

$$V_{i,0}^{a,t} = H_0^{a,t}(V_i^a) \quad (12)$$

由于前景区域的近似多边形顶点是按照顺序排列并映射的, 所以映射后的顶点 $V_{i,0}^{a,t}$ 只需要循序连接起来就能重构出前景映射区域的轮廓, 对该区域的内部进行填充就能得到一个重构区域。与基于图像的前景图映射法相比, 由此得到的重构区域是直接法得到的前景区域近似表达, 所以该方法被称为多边形顶点法。重构前景映射图 $F_0^{a,t}$ 为所有重构区域 $F_{i,0}^{a,t}$ 的集合:

$$F_0^{a,t} = \sum_i F_{i,0}^{a,t} = \sum_i (H_0^{a,t}(F_i^a)) \quad (13)$$

如果将摄像机用一个光源替代, 每一个目标在顶视图的前景映射区域就可看作是在光源作用下物体在地面上的影子, 即由地平面与物体轮廓投影的相交区域。图 1 为基于地平面的单摄像机单应投影示意图。类似的方法可以在文献 [24-26] 中找到。

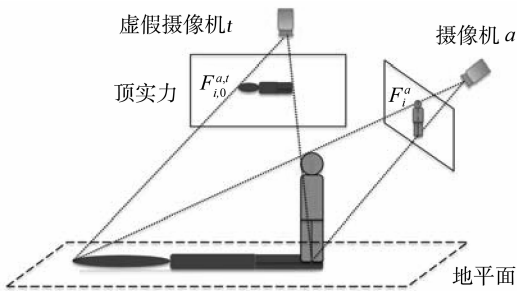


图 1 基于地平面的单摄像机单应投影示意图

Fig. 1 Schematic diagram of the homography projection according to the ground plane

4.2 多摄像机融合

当采用多个摄像机时, 多个摄像机分别检测出来的前景图在投射至顶视图后进行叠加就可得到多摄像机的融合图。由不同摄像机视角产生的前景投影区域可能会在顶视图中相交, 相交区域在融合图

中得到增强:

$$p_0^i = \cap_c F_0^{c,t} \quad (14)$$

式中, c 是摄像机的标号, $c = a, b$ 。

当顶视图中的前景投影区域是来自来源于不同视角的同一运动物体时, 它们的相交区域显示了该运动物体在顶视图中接触地面的位置。图 2 是一个行人的双摄像机融合和相交的示意图, 其中两个前景投影区域相交于的黑色区域是目标站立在地面上的位置, 但其形状、尺寸和目标的横截面并不完全相同。

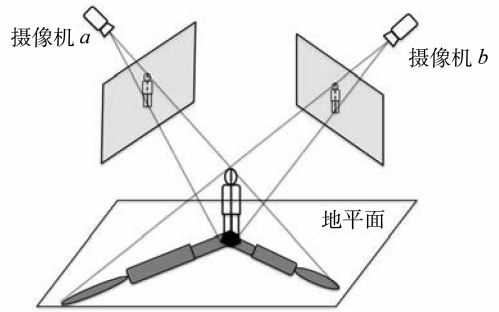


图 2 基于地平面的双摄像机单应投影融合和相交的示意图

Fig. 2 Schematic diagram of the overlaid foreground projections and the intersection region

4.3 多平面融合

为了提高多视角融合目标检测的可靠性, 可采用与地面平行的一系列不同高度的虚拟平面作为单应平面。如采用 h 作为平面下标, 基于多平面的多视角前景融合顶视图可以表示为

$$F^t = \sum_h F_h^t \quad (15)$$

在获取各目标在顶视图中的位置时, 可采用一系列根据经验设定的阈值进行判断。如果融合顶视图 F^t 中的某一个像素值大于给定的阈值 Th_i , 该像素即为前景像素。最后经过连通区域分析、形态学处理和尺寸滤波, 就可将检测出来的前景像素转换成代表目标位置的前景区域, 完成顶视图中的目标检测。

$$P^t = \{(r, c) : F^t(r, c) > Th_i\} \quad (16)$$

5 实验结果及分析

本文采用了两组视频序列对所提出的多视角实时运动目标检测算法进行测试, 并将基于多边形顶点的前景映射法与基于图像、轮廓、矩形框的前景映射法进行比较, 对算法的有效性和可靠性进行评估。实验中同样对算法的运行时间进行了测试, 以

评估其实时性。序列一是在校园中采集的，两个摄像机离运动目标较近，所以检测出的前景区域的尺寸较大。序列二是标准序列，摄像机被架设在路边的建筑物上，所以检测出的前景区域较小。

5.1 校园序列

校园序列中，每个视角视频采用的分辨率为像素，帧率为 15 帧/s。视频中共有 2 790 帧，由于前 660 帧不存在目标或只存在单目标，实验只对后 2 155 帧每隔 15 帧处理一次。顶视图的分辨率为 $840 \times 1\,000$ 像素。

图 3 所示为前景区域提取和前景区域的多边形近似结果，图 3 (a) 和 3 (b) 为原图及所提取的前景图。实验测试了不同的近似距离 ε 来控制顶点的个数，图 3 (c) 和 3 (d) 中采用的 ε 分别为 0 和 2 个像素。其中绿色的线条为前景区域的轮廓，红色的点为多边形近似后的顶点。

多边形近似，平均每个前景区域的轮廓需要用 217.84 个像素表示，而矩形框只需要 4 个像素。采用多边形近似用顶点表示前景区域，随着 ε 的取值由 1 增加到 10，所需的顶点像素个数由 62.73 减小到 8.87 个，但所需的处理时间并没有明显的降低。多边形近似可准确地重建检测出的前景区域，准确度随着 ε 值的减小而提高。采用矩形框法会造成非常多的错误映射像素。

表 1 前景多边形近似、轮廓、矩形框近似性能比较
Table 1 The comparison result of the accuracy and the processing speeds for the polygon approximations with different distance, the contour method and the bounding box method

项目	轮廓	多边形近似距离 ε (像素)				矩形框
		1.0	2.0	5.0	10.0	
前景区域 总个数		667				
总映射点	145 297	41 846	21 601	10 751	5 919	2 668
平均映射点	217.84	62.73	32.39	16.12	8.87	4
总时间/s	1.07	0.69	0.61	0.58	0.55	0.35
$R_{FN}/\%$	0.04	0.53	1.33	4.03	7.28	0.03
$R_{FP}/\%$	0.55	0.67	1.27	3.17	7.36	80.55

经过多边形近似后，多边形的顶点被投射至顶视图，并根据投射的顶点重构前景投影区域。图 4 (a) 和 4 (b) 分为图 3 (b) 和 3 (d) 在顶视图中的映射结果。采用三种方法获取的像素映射后重构出的前景映射区域的准确率比较结果如表 2 所示。多边形近似法可以可靠地重建前景映射区域，当 ε 小于 2 个像素时，重建结果与轮廓法的准确性相似，更能调节 ε 取值以灵活选取所需计算和传输的顶点个数。

表 2 前景多边形近似、轮廓、矩形框映射准确率比较

Table 2 The comparison result of the accuracy for the polygon projection with different with different distance, the contour method and the bounding box method

项目	轮廓	多边形近似距离 ε (像素)				矩形框
		1.0	2.0	5.0	10.0	
$R_{FN}/\%$	3.09	3.02	3.26	4.96	6.16	0.01
$R_{FP}/\%$	2.98	3.07	3.26	4.51	9.69	78.72

当采用多个摄像机时，顶视图中的前景映射区

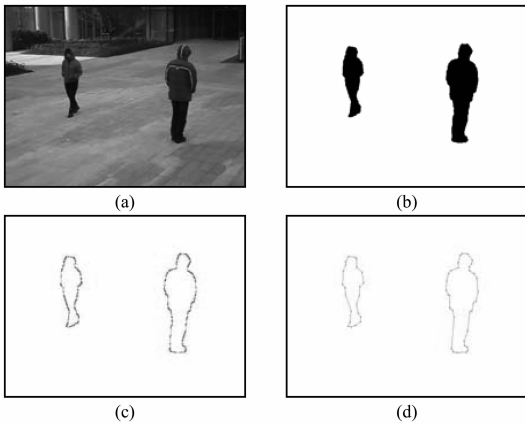


图 3 前景区域检测及多边形近似结果

Fig. 3 The foreground polygon approximation results
(a) 第 1020 帧摄像机视角 a 原图；(b) 前景图；
(c) 前景轮廓及多边形顶点图 ($\varepsilon = 0$)；
(d) 前景轮廓及多边形顶 ($\varepsilon = 2$)

设第 k 帧的原始前景图像为 $F_{temp,k}$ ，经过近似后重构的前景图像为 $F_{poly,k}$ ，

$$R_{FN} = \frac{\sum_k \#(F_{temp,k} \cap F_{poly,k}^c)}{\sum_k \#(F_{temp,k})} \quad (17)$$

$$R_{FP} = \frac{\sum_k \#(F_{temp,k}^c \cap F_{poly,k})}{\sum_k \#(F_{temp,k})} \quad (18)$$

式中， $\#()$ 表示计算非 0 像素的个数， c 表示补集。

表 1 所示为采用不同的 ε 进行前景多边形近似，与前景轮廓（高度近似）和矩形框（非常粗略近似）的时间和准确度比较的结果。如不采用

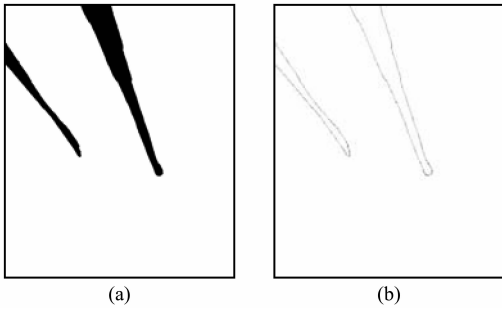


图 4 前景区域图像级映射和多边形近似映射结果比较

Fig. 4 Foreground projection results

- (a) 采用图像级映射法的前景映射图;
(b) 采用多边形顶点法的前景多边形顶点映射图

域叠加后即可得到顶视融合图, 所采用的单应平面可以是地面或与地面平行的一系列虚拟平面。图 5 为两个视角下, 对第 1 275 帧根据不同高度平面进行前景映射至顶视图叠加的结果。所采用平面的高度分别为: 0 m (地平面), 0.75 m 和 1.5 m。

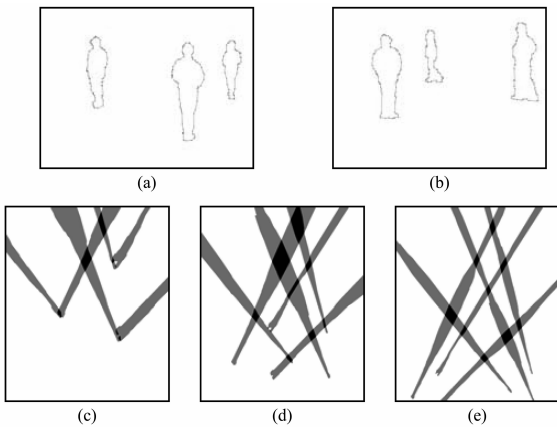


图 5 根据不同高度平面映射叠加的结果

Fig. 5 Fusion of the foreground projections according to the homographies for a set of parallel planes at different heights

- (a) - (b) 两个视角的前景轮廓及多边形顶点图;
(c) 顶视图中根据地平面前景映射融合图;
(d) - (e) 根据高度为 0.75 m 和 1.5 m 平面的前景映射融合图

图像法和多边形近似法的时间评估主要分三部分: ① 两个摄像机的前景检测; ② 两个摄像机的前景投影; ③ 顶视图中前景映射融合。本文主要是对第二部分进行了改进, 两种算法的时间比较结果如表 3 所示。相比于图像级映射所需的 71.56 s 处理 142 帧, 采用多边形近似法后映射只需要 1.03 s, 其计算速度是图像级的 69.47 倍。

表 3 图像法与多边形近似法投影时间比较

Table 3 Execution times for running the bitmap projection and the polygon projection

项目	图像法	多边形近似法	s
前景检测			31.22
		多边形近似	0.69
		顶点映射	0.03
前景投影	71.56	多边形填充	0.31
		总映射	1.03
前景投影叠加			2.08

根据公式 (16), 可计算叠加区域中的相交区域。图 6 所示是根据多高度平面映射叠加的结果 (图 6 (a)), 及取不同阈值时的相交区域结果。选取不同大小的阈值会对影响相交区域的结果, 图 6 (b) 和 6 (c) 的阈值分别取 255 和 210。阈值 Th_i 可根据经验选取, 或根据所采用单应平面的层数 D 自动选取:

$$Th_i = 255D^{-1}(D - 1) \quad (19)$$

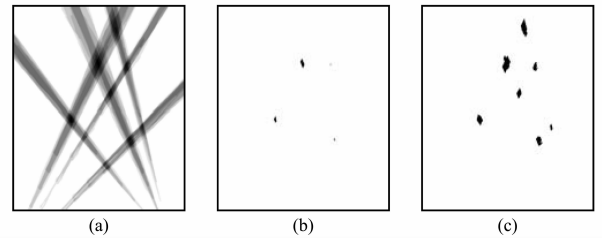


图 6 根据多高度平面映射叠加及使用不同阈值时的相交区域结果

Fig. 6 Overlaid foreground projections from two camera views and intersection regions identified with different thresholds Th_i

- (a) 根据两个视角 6 个高度平面的前景映射融合图;
(b) $Th_i = 255$ 的相交区域; (c) $Th_i = 210$ 的相交区域

5.2 PETS'2001 序列

PETS'2001 序列是由英国金斯顿大学公布的视频监控的一组标准序列, 实验中摄像机视角和顶视图分别选用了 384×288 像素和 500×500 像素的分辨率。图 7 为采用标准序列的两个视角前景多层单应的运动物体检测过程, 其中, (a) (b) 是前景轮廓, (c) (d) 是前景区域, (e) 是前景映射融合, (f) 是顶视图中的运动目标检测结果。由于摄像机架设得较远, 图像中运动物体的尺寸较小, 实验中 ε 的取值变化为 0.5 ~ 2 像素。实验结果显示漏报率由 2.9% 增加至 4.4%, 虚警率则由 4.8% 增加至 5.4%, 同样可以精准地重构映射的前景区

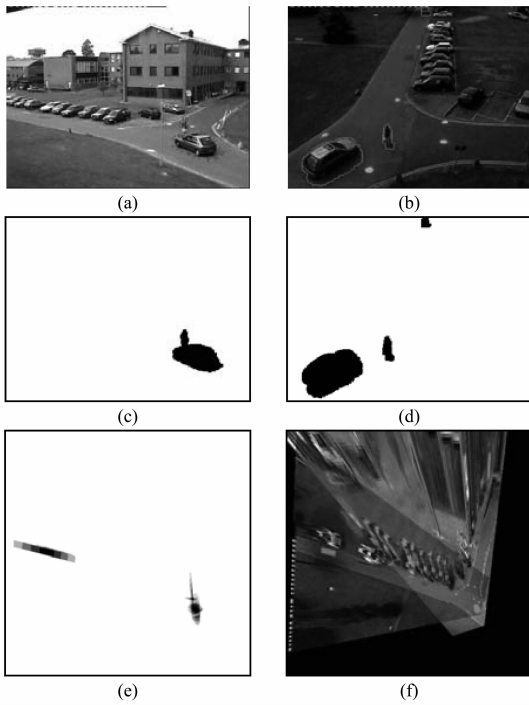


图 7 采用序列 2 的多平面前景单应映射融合运动目标检测结果

Fig. 7 Object detection by foreground fusion for multi-plane homographies using PETS' 2001 datasets
(a) - (b) 前景轮廓; (c) - (d) 前景区域;
(e) 前景映射融合; (f) 顶视图中的运动目标检测结果

域。处理一帧算法所需运行时间为 8.5 ms, 相比基于图像级映射所需的 108.5 ms, 提高 12.8 倍。

综合以上两个测试数据集的实验结果, 本文提出的算法在运动目标尺寸较大或较小的情况下, 均能可靠地重构出映射的前景区域, 完成基于多摄像机多平面单应前景映射融合的目标实时检测。随着参考视角分辨率的增高, 相对于传统的图像级单应变化, 其运算速度的提高会更加显著。

6 总 结

本文提出了一种实时的多摄像机目标检测算法。为了解决摄像机信息高度融合时, 传统的图像级单应变换对像素逐一映射所带来的速度慢的缺点, 本文采用了基于前景轮廓多边形逼近, 只对多边形的顶点进行单应映射, 并在参考视角中根据映射顶点进行前景重构。实验表明, 重构的前景映射图不仅可以有效地逼近图像级单应变换的前景映射图, 而且可以有效地降低运算时间, 使基于多摄像机多平面单应前景映射融合的目标检测具有实时性。与基于轮廓和矩形框的重构方法相比, 本方法具有更强的可控性或较低的错误映射像素。

参考文献:

- [1] WANG X. Intelligent multi-camera video surveillance: A review[J]. *Pattern Recognition Letters*, 2013, 34(1): 3 - 19.
- [2] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述[J]. *计算机学报*, 2015, 38(6): 1093 - 1118.
- [3] 李熙莹, 罗长政, 王镇波, 等. 高饱和交叉口背景提取与更新算法[J]. *中山大学学报(自然科学版)*, 2012, 51(3): 23 - 28.
- [4] DOCKSTADER S L, TEKALP A M. Multiple camera tracking of interacting and occluded human motion[J]. *Proceedings of the IEEE*, 2001, 89(10): 1441 - 1455.
- [5] MITTAL A, DAVIS L S. M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene[J]. *International Journal of Computer Vision*, 2003, 51(3): 189 - 203.
- [6] 夏侯凯顺, 严娟, 叶小鹏, 等. 基于 Kalman 滤波的无线传感器网络多目标跟踪[J]. *中山大学学报(自然科学版)*, 2014, 53(2): 18 - 22.
- [7] BLACK J, ELLIS T, ROSIN P. Multi view image surveillance and tracking[C]//*Proceedings of the IEEE Workshop on Motion and Video Computing*, 2002: 169 - 174.
- [8] HU W, HU M, ZHOU X, et al. Principal axis-based correspondence between multiple cameras for people tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(4): 663 - 671.
- [9] 林青, 范晶晶, 胡波. 多摄像头目标跟踪动态单应投影方法研究[J]. *复旦大学学报*, 2015, 50(4): 501 - 510.
- [10] 高翔, 朱婷婷, 刘洋. 多摄像头系统的目标检测与跟踪方法研究[J]. *计算机技术与发展*, 2015, 25(7): 221 - 225.
- [11] KHAN S, SHAH M. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(10): 1355 - 1360.
- [12] KIM K, DAVIS L S. Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering[M]//*Proceedings of the ECCV. Berlin Heidelberg: Springer*, 2006: 98 - 109.
- [13] ARSI D, HRISTOV E, LEHMENT N, et al. Applying multi layer homography for multi camera person tracking[C]//*Proceedings of the Second ACM/IEEE International Conference on Distributed Smart Cameras*, 2008: 1 - 9.
- [14] ALAHI A, JACQUES L, BOURSIER Y, et al. Sparsity driven people localization with a heterogeneous network of cameras[J]. *Journal of Mathematical Imaging and*

- Vision, 2011, 41(1/2): 39 – 58.
- [15] FLEURET F, BERCLAZ J, LENGAGNE R, et al. Multicamera people tracking with a probabilistic occupancy map[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 267 – 282.
- [16] KHAN S M, SHAH M. Tracking multiple occluding people by localizing on multiple scene planes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 505 – 519.
- [17] 姜明新, 王洪玉, 刘晓凯. 基于多相机的多目标跟踪算法 [J]. 自动化学报, 2012, 38(4): 531 – 539.
- [18] ESHEL R, MOSES Y. Tracking in a dense crowd using multiple cameras[J]. International Journal of Computer Vision, 2010, 88(1): 129 – 143.
- [19] PENG P, TIAN Y, WANG Y, et al. Robust multiple cameras pedestrian detection with multi-view Bayesian network[J]. Pattern Recognition, 2015, 48(5): 1760 – 1772.
- [20] CRIMINISI A. Single-view metrology [M]//Accurate Visual Metrology from Single and Multiple Uncalibrated Images. London: Springer, 2001: 69 – 105.
- [21] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.
- [22] SUZUKI S. Topological structural analysis of digitized binary images by border following [J]. Computer Vision, Graphics, and Image Processing, 1985, 30(1): 32 – 46
- [23] DOUGLAS D H, PEUCKER T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature[J]. Cartographica: The International Journal for Geographic Information and Geovisualization, 1973, 10(2): 112 – 122.
- [24] REN J, ORWELL J, JONES G, et al. Tracking the soccer ball using multiple fixed cameras [J]. Computer Vision and Image Understanding, 2009, 113(5): 633 – 642.
- [25] REN J, ORWELL J, JONES G, et al. Real-time modeling of 3 – d soccer ball trajectories from multiple fixed cameras [J]. IEEE, 2008, 18(3): 350 – 362.
- [26] REN J, XU M, ORWELL J, et al. Multi-camera video surveillance for real-time analysis and reconstruction of soccer games [J]. Machine Vision and Applications, 2010, 21(6): 855 – 863.

(上接第 29 页)

- [6] 杨俊仙, 闫萍. 一类具饱和发生率的时滞 SEIR 传染病模型的分析 [J]. 中山大学学报 (自然科学版), 2015, 54(3): 51 – 55.
- [7] BAIRAGI N, ADAK D. Global analysis of HIV-1 dynamics with Hill type infection rate and intracellular delay [J]. Applied Mathematical Modelling, 2014, 38(21/22): 5047 – 5066.
- [8] URSZULA F, JAN P. A delay-differential equation model of HIV related cancer-immune system dynamics [J]. Mathematical Biosciences and Engineering, 2011, 8(2): 627 – 641.
- [9] LI B, CHEN Y M, LU X J, et al. A delayed HIV – 1 model with virus waning term [J]. Mathematical Biosciences and Engineering, 2016, 13(1): 135 – 157.
- [10] LI D, MA W. Asymptotic properties of a HIV – 1 infection model with time delay [J]. Journal of Mathematical Analysis and Applications, 2007, 335(1): 683 – 691.
- [11] XU R. Global stability of an HIV – 1 infection model with saturation infection and intracellular delay [J]. Journal of Mathematical Analysis and Applications, 2011, 375(1): 75 – 81.
- [12] 杨若晨, 马明菊, 齐逸飞, 等. 含潜伏时滞效应和非线性发生率的 SEIR 模型的长时间行为 [J]. 中山大学学报 (自然科学版), 2015, 54(1): 24 – 29, 36.
- [13] HALE J K, VERDUYN LUNEL S M. Introduction to functional differential equations [M]. Berlin: Springer-Verlag, 1993: 130 – 166.