

基于多相机的多目标跟踪算法

姜明新^{1,2} 王洪玉¹ 刘晓凯¹

摘要 多目标的稳定跟踪是计算机视觉领域的一个具有挑战性的问题. 本文提出了一种基于多相机的多目标定位跟踪算法. 首先, 利用不同高度层上的标志物, 计算基于多层的不同视角间的单应性矩阵. 然后, 利用码本模型对背景进行建模, 检测多个视角的前景似然信息. 最后, 通过单应性变换获得多目标在不同高度层上的定位信息, 利用最短路径优化算法实现跟踪. 与其他算法相比, 本算法不需要计算多相机的隐消点, 降低了算法的复杂度, 提高了算法的准确性. 采用最短路径优化算法, 提高了跟踪算法的效率. 实验结果表明, 本算法对遮挡具有很强的鲁棒性, 并且能够满足实时性要求.

关键词 多目标跟踪, 单应性, 码本, 最短路径优化, 多相机

引用格式 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

DOI 10.3724/SP.J.1004.2012.00531

A Multi-target Tracking Algorithm Based on Multiple Cameras

JIANG Ming-Xin^{1,2} WANG Hong-Yu¹ LIU Xiao-Kai¹

Abstract The reliable tracking of multi-targets is a challenging issue in computer vision. In this paper, we propose a novel multi-target localizing and tracking algorithm based on multiple cameras. Firstly, the view-to-view homographies are computed using several landmarks on different planes. Then, the foreground likelihood map in each view is obtained by using a codebook background modeling algorithm. Finally, we can localize multiple objects at multiple planes and perform tracking by shortest paths optimization. Compared with other popular methods, our proposed algorithm does not require computing the vanishing points of cameras. Therefore, it reduces the complexity and improves the accuracy simultaneously. Adopting the shortest path optimization algorithm can improve the tracking efficiency. The experimental results demonstrate that our method is robust to occlusion and also can achieve real-time performance.

Key words Multi-target tracking, homography, codebook, the shortest paths optimization, multiple cameras

Citation 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

多目标跟踪在视频理解、机器人导航、智能监控等领域都有着广泛的应用^[1-2]. 由于遮挡和光照变化等因素的干扰, 多目标的稳定跟踪成为了计算机视觉领域的一个具有挑战性的问题. 当多目标之间遮挡频繁发生时, 一个前景区域可能属于多个目标. 根据传统的颜色分布、形状等信息, 我们几乎不可能实现对多目标的准确检测和跟踪, 也就无法开展视频内容理解、异常行为检测等进一步的研究工作. 近年来, 很多学者做了这方面的研究, 所提出的

算法可以大致分为两类: 单目跟踪算法和多目跟踪算法. 文献 [3] 提出了基于单个相机利用卡尔曼滤波的多目标跟踪算法, 对目标建立 3D 模型来帮助解决遮挡问题. 文献 [4] 提出了利用蒙特卡洛采样方法进行多目标跟踪. 另外一种比较常见的单目跟踪算法是由 Comaniciu 提出的 Mean shift 算法^[5]. 文献 [6] 提出了一种基于 Mean shift 和粒子滤波的结合算法. 单目跟踪算法虽然处理起来比较简单, 但是由于 3D 空间中的目标在 2D 成像的过程中损失了很多信息, 仅仅从一个视角去检测和跟踪多个目标很难解决多目标之间的遮挡问题. 因此, 多目跟踪算法越来越受到研究人员的重视.

如图 1 所示, 当多个目标在一个视角发生遮挡时, 换个视角可能不遮挡, 如果我们能够融合多个视角的信息, 就可以很好地解决多目标之间相互遮挡的问题. Mittal 等提出了一种多目跟踪算法^[7], 利用贝叶斯框架实现对多目标的跟踪. 文献 [8-9] 提出了利用单应性知识实现对多目标跟踪的算法, 但这些都是建立在相机标定的基础上进行的, 而相机的标定过程比较复杂, 计算量也较大. 文献 [10] 提出了一种可以在人流密集的环境下跟踪多目标头部的算

收稿日期 2011-06-10 录用日期 2011-11-10
Manuscript received June 10, 2011; accepted November 10, 2011

国家自然科学基金 (61172058), 中央自主科研基金项目 (DC110321) 资助

Supported by National Natural Science Foundation of China (61172058) and the Fundamental Research Funds for the Central Universities (DC110321)

本文责任编辑 贾云得

Recommended by Associate Editor JIA Yun-De

1. 大连理工大学信息与通信工程学院 大连 116023 2. 大连民族大学信息与通信工程学院 大连 116600

1. School of Information and Communication Engineering, Dalian University of Technology, Dalian 116023 2. School of Information and Communication Engineering, Dalian Nationalities University, Dalian 116600

法, 该算法提出的方案理论上可行, 但由于人的高度各异, 想准确地跟踪多目标的头部, 需要细分高度层 (在 130 cm 到 200 cm 之间, 每 1 cm 取一层也无法保证把所有目标的高度都取全. Khan 等^[11] 提出了一种不需要完全标定相机的多目标跟踪算法, 该算法虽然不需要相机的完全标定, 但需要计算每个相机的消隐点, 消隐点的计算是个比较复杂的过程, 无法保证计算的绝对准确. 而消隐点的计算误差将会直接影响多目标定位和跟踪的准确性. 另外一个显著的问题是, 该算法的跟踪部分使用图割理论^[12] 对空间 10 层的定位信息进行处理, 增加了算法的计算复杂度, 降低了跟踪效率, 导致该算法根本无法满足实时性的要求.

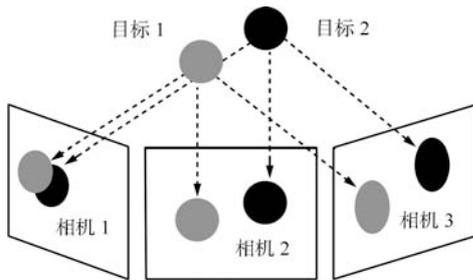


图 1 多目跟踪示意图

Fig. 1 Illustration of multi-camera tracking

基于以上分析, 本文提出了一种基于多相机的多目标定位跟踪算法. 在不同视角放置多个同步相机, 在垂直于地面的方向上选取不同高度的 3 层, 在每一层上设置 4 个标志物, 计算基于每一层的单应性矩阵. 采用码本模型进行背景建模, 通过背景减除法获得多个视角的前景似然信息. 利用基于 3 层的单应性映射融合多相机的前景似然信息, 获得多目标在选定高度层上的定位信息. 然后利用最短路径优化算法处理 3 层定位信息, 融合多层的处理结果, 实现对多目标的立体跟踪. 本算法不需要计算多相机的消隐点, 在很大程度上降低了计算的复杂度, 提高了跟踪算法的准确性. 引入码本模型进行背景建模, 降低了该算法的多目标检测误差, 为准确跟踪提供了必要的前提条件. 利用最短路径优化算法实现跟踪, 显著提高了跟踪算法的效率. 实验结果表明, 本算法能够稳定地跟踪多个目标, 对遮挡和不同视角中的光照变化具有很强的鲁棒性, 同时能够满足实时性要求.

1 运动目标检测

如果运动目标检测的结果存在空洞和噪点过多的问题, 那么将各个视角的前景似然信息映射到参考视角中, 就必然会产生很多错误的定位信息, 进而影响多目标跟踪的准确性. 在相机静止的条件下, 运

动目标的检测方法主要有光流法、帧差法和背景减除法.

背景减除法是比较常用的一种运动目标检测方法, 基本思想是建立背景模型, 将当前帧与背景图像进行差分比较实现对运动目标的检测. 很多文献采用混合高斯模型 (Mixture of Gaussians, MOG) 对背景进行建模, 获取前景似然信息. 混合高斯模型是在像素域的时间尺度上对像素进行分类, 很难将其学习效率控制在既不过检也不漏检的理想状态, 常常会造成误判, 无法解决运动目标空洞和噪点过多的问题.

本文采用码本模型对背景进行建模, 码本模型是由 Kim 提出的^[13], 该模型的基本思想是对图像序列进行学习, 根据每个像素点连续采样值的颜色距离和亮度范围为每个像素点生成一个码本, 然后根据前景点和背景点在图像序列中的分布特性分离出背景码字, 利用背景码字构造出背景, 再利用背景减除法检测运动目标. 该算法抗干扰能力强, 能够有效地克服阴影和空洞的问题, 计算复杂度小, 适合做实时检测.

1.1 码本背景模型描述

假设 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 是一个像素的序列采样值, 其中 \mathbf{x}_t ($t = 1, \dots, N$) 是当前像素 t 时刻的 RGB 向量值. 设 $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_L)$ 为该像素的码本, 码本中含有 L 个码字. 每个码字 \mathbf{c}_i ($i = 1, \dots, L$) 由两部分组成: $\mathbf{aux}_i = (\tilde{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i)$ 和 $\mathbf{v}_i = (\bar{R}_i, \bar{G}_i, \bar{B}_i)$, 其中 \tilde{I}_i, \hat{I}_i 表示码字中最小和最大亮度值; f_i 表示码字出现的频率; λ_i 表示训练中该码字没有出现的最大时间间隔; p_i 和 q_i 分别表示该码字第一次出现和最后一次出现的时间. 此处, 引入两个概念: 颜色失真度和亮度函数:

颜色失真度定义为

$$\text{colordist}(\mathbf{x}_t, \mathbf{v}_i) = \frac{\|\mathbf{x}_t\|^2 \|\mathbf{v}_i\|^2 - \langle \mathbf{x}_t, \mathbf{v}_i \rangle^2}{\|\mathbf{v}_i\|^2} \quad (1)$$

其中, $\|\mathbf{x}_t\|^2 = R^2 + G^2 + B^2$, $\|\mathbf{v}_i\|^2 = \bar{R}_i^2 + \bar{G}_i^2 + \bar{B}_i^2$, $\langle \mathbf{x}_t, \mathbf{v}_i \rangle^2 = (\bar{R}_i R + \bar{G}_i G + \bar{B}_i B)^2$.

亮度函数定义为

$$\text{brightness}(I, \langle \tilde{I}, \hat{I} \rangle) = \begin{cases} \text{true}, & \text{若 } I_{\text{low}} \leq \|\mathbf{x}_t\| \leq I_{\text{hi}} \\ \text{false}, & \text{否则} \end{cases} \quad (2)$$

为了判断前景和背景, 运动目标检测中亮度变化有一个范围, 对于每个码字, 其范围为 $[I_{\text{low}}, I_{\text{hi}}]$.

码本背景建模的详细过程如下:

步骤 1. 初始化, 将每个像素的码本置空, 即 $L = 0$.

步骤 2. 对于训练视频中的每个像素的采样值中的每个 $\mathbf{x}_t = (R, G, B)$ 定义其亮度值 $I = R + G + B$, 并根据以下两个条件找出与其匹配的码字 \mathbf{c}_m :

$$\text{colordist}(\mathbf{x}_t, \mathbf{v}_m) \leq \varepsilon_1$$

$$\text{brightness}(I, (\check{I}_m, \hat{I}_m)) = \text{true}$$

步骤 3. 如果码本为空或者找不到匹配的码字, 则令 $L = L + 1$, 并创建一个新的码字 \mathbf{c}_L :

$$\begin{cases} \mathbf{v}_L = (R, G, B) \\ \mathbf{aux}_L = (I, I, 1, t - 1, t, t) \end{cases} \quad (3)$$

如果有码字 \mathbf{c}_m 满足上述连个条件, 其内容为

$$\begin{cases} \mathbf{v}_m = (\bar{R}_m, \bar{G}_m, \bar{B}_m) \\ \mathbf{aux}_m = (\check{I}_m, \hat{I}_m, f_m, \lambda_m, p_m, q_m) \end{cases} \quad (4)$$

那么把该码字更新为

$$\begin{cases} \mathbf{v}_m = \left(\frac{f_m \bar{R}_m + R_t}{f_m + 1}, \frac{f_m \bar{G}_m + G_t}{f_m + 1}, \frac{f_m \bar{B}_m + B_t}{f_m + 1} \right) \\ \mathbf{aux}_m = (\min\{I, \check{I}_m\}, \max\{I, \hat{I}_m\}, f_m + 1, \\ \max\{\lambda_m, t - q_m\}, p_m, t) \end{cases} \quad (5)$$

步骤 4. 训练结束后, 计算该像素每个码字没有再次出现的最大时间间隔, 对于每个码字 \mathbf{c}_i ($i = 1, \dots, L$):

$$\lambda_i = \max\{\lambda_i, (N - q_i + p_i - 1)\} \quad (6)$$

利用 λ 消除冗余码字, 得到最能代表真实背景的初始码本 $M = \{\mathbf{c}_k | \mathbf{c}_k \in C, \lambda_k \leq T_M\}$ (k 为码字的索引), T_M 通常取训练帧数的一半.

1.2 基于码本模型的运动目标检测

假设在目标检测的过程中, 输入像素为 $\mathbf{x}_i = (R, G, B)$, 其对应的码本为 M . 减背景操作 $BGS(\mathbf{x})$ 可以大致分为以下三步:

步骤 1. 计算当前像素的亮度 $I = R + G + B$, 定义布尔变量 $\text{match} = 0$, 并给阈值变量 ε_2 赋值.

步骤 2. 根据以下两个条件从码本 M 中找出与当前像素相匹配的码字 \mathbf{c}_m , 如果能够找到码字 \mathbf{c}_m , 则 $\text{match} = 1$, 否则, $\text{match} = 0$:

$$\text{colordist}(\mathbf{x}_i, \mathbf{v}_m) \leq \varepsilon_2$$

$$\text{brightness}(I, (\check{I}_m, \hat{I}_m)) = \text{true}$$

步骤 3. 像素 \mathbf{x}_i 属于前景的似然函数为

$$L(x_i) = \begin{cases} 1, & \text{match} = 0 \\ 0, & \text{match} = 1 \end{cases} \quad (7)$$

2 定位算法

2.1 单应性矩阵

设 π 是不通过两个相机任一光心的空间参考平面, 本文选择地平面作为参考平面. 假设在不同视角放置 N 个同步相机, 相机 i ($i = 1, 2, \dots, N$) 和相机 j ($j = 1, 2, \dots, N$) 拍摄的图像 (简称像平面) 分别记为 I_i 和 I_j . 为了保证单应性的存在, N 个相机必须拍摄参考平面上的同一块区域. 令 X 是平面 π 上的任意一点, X 在 I_i 和 I_j 中成像点的坐标分别 $\mathbf{m}_k = (x_k, y_k)$ 和 $\mathbf{m}'_k = (x'_k, y'_k)$, 我们定义一个 3×3 的矩阵: $H_{i\pi j} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}$, 使 $\mathbf{m}'_k = H_{i\pi j} \mathbf{m}_k$. 矩阵 $H_{i\pi j}$ 称为平面 π 诱导的两个相机间 (或两个像平面间) 的单应性矩阵, 或简称为平面 π 的单应性矩阵, 相应的变换称为单应性变换. 利用单应性矩阵 $H_{i\pi j}$, 从一个像平面上的点可以得到另一个像平面上的对应点^[14]. 单应性矩阵 $H_{i\pi j}$ 是一个齐次可逆矩阵.

2.2 基于不同高度层的多目标定位

如果只在参考平面 (地平面) 上对目标定位, 鲁棒性比较差. 当目标跑动或者跳动时, 可能和地面没有接触点, 这时在地平面上检测不到目标, 无法获得目标在地平面上的定位信息. 另外, 在检测运动目标时, 无论采用哪种模型对背景进行建模, 也不能保证检测的区域完全没有缺失, 当鞋的颜色和背景颜色十分接近时, 脚的部分可能会被漏检. 为了提高跟踪算法的鲁棒性, 我们在平行于参考平面不同高度的 2 个平面上也对多目标进行定位, 然后融合 3 层的定位信息, 对多目标进行跟踪. 选取 3 层, 既能获取目标的立体信息, 保证多目标跟踪的鲁棒性和准确性, 又可以满足实时性的要求. 选取层数过多, 对提高跟踪精度没有大的帮助, 反而会导致计算效率严重下降, 无法满足实时性要求.

如图 2 所示, π 为参考平面 (地面), ϕ 为平行于参考平面的任意平面. 由平面 ϕ 诱导的任意两个像平面 I_i 和 I_j 间的单应性矩阵记为 $H_{i\phi j}$ ^[15]. X 在 I_i 和 I_j 中成像点 $\mathbf{m}_k = (x_k, y_k)$ 和 $\mathbf{m}'_k = (x'_k, y'_k)$ 同样可以通过 $H_{i\phi j}$ 建立映射关系, 即为

$$\begin{bmatrix} x'_k \\ y'_k \\ 1 \end{bmatrix} = \begin{bmatrix} h'_{11} & h'_{12} & h'_{13} \\ h'_{21} & h'_{22} & h'_{23} \\ h'_{31} & h'_{32} & 1 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ 1 \end{bmatrix} \quad (8)$$

其中, $H_{i\phi j} = \begin{bmatrix} h'_{11} & h'_{12} & h'_{13} \\ h'_{21} & h'_{22} & h'_{23} \\ h'_{31} & h'_{32} & 1 \end{bmatrix}$.

同理, 可以再选取一个不同高度的平面 κ , 由平面 κ 诱导的任意两个像平面 I_i 和 I_j 间的单应性矩阵记为 $H_{i\kappa j}$.

利用第 1 节的运动目标检测算法得到每个视角的前景似然信息, 选择其中任意一个视角作为参考图像, 利用多层单应性映射关系将其他两个视角的前景似然信息映射到参考图像中, 可以得到多个视角基于不同高度层的融合图像, 融合图像上的高亮区域即为多目标在不同高度层上的定位信息团块 (Blob). 由此可见, 基于多层的单应性矩阵的准确性十分重要, 它直接关系到多层定位信息的准确性, 定位信息如果不准确, 那么跟踪的准确性就无法保证了. 文献 [11] 利用消隐点来计算基于不同高度层的不同视角间的单应性矩阵, 消隐点的计算误差^[16-17]必然导致基于多层的单应性矩阵存在计算误差.

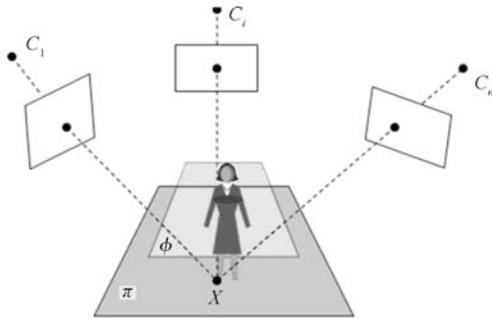


图 2 多层单应性示意图

Fig. 2 Illustration of homography induced by different planes

通过式 (8) 可以看出 $H_{i\phi,j}$ 具有 8 个自由度, 如果能够在一层上找到 4 对对应的特征点, 就可以计算出基于该层的任意两个视角间单应性矩阵. 基于这个特点, 采用一种简单准确的标定方法来计算基于选定 3 层的单应性矩阵. 如图 3 所示, 在拍摄的环境中垂直于地面放置 4 根带底座的杆子, 底座可以保证杆子垂直于地面. 在不同高度的 3 层上分别设置 4 个标志物, 利用标志物在像平面中团块的中心坐标来计算每一层的单应性矩阵. 经实验论证, 该方法不仅避免了检测消隐点的繁琐过程, 而且提高了多层单应性矩阵计算的准确性.



图 3 标定空间多层示意图

Fig. 3 Illustration of calibrating multiple planes in space

3 跟踪算法

本文的跟踪算法基于目标所占位置的时空一致性, 利用最短路径优化算法对多层定位信息进行处理, 实现对多目标的跟踪. 在第 2 节中, 我们得到了多目标在多层上的定位信息团块. 大多数情况下, 目标在相邻两帧的运动距离是非常有限的, 计算相邻两帧任意两个团块重心的距离, 取距离最小的两个团块作为前后两帧的匹配团块. 以一层为例来说明最短路径优化算法步骤:

步骤 1. $t = 1$ 帧, 初始化每个目标的定位信息团块, 并做标记 b_i^1 , 计算每个定位信息团块的重心 $C_{b_i^1}(x_i^1, y_i^1)$.

步骤 2. $t > 1$ 帧, 假设 $t - 1$ 时刻目标 i 的定位团块为 b_i^{t-1} , 其重心为 $C_{b_i^{t-1}}(x_i^{t-1}, y_i^{t-1})$. 在 t 时刻的定位信息团块集中寻找与 $C_{b_i^{t-1}}(x_i^{t-1}, y_i^{t-1})$ 距离最近的团块 b_j^t , 作为匹配团块, 并给 b_j^t 赋予标签 i . 也就是说, 如果满足条件 (9) 和 (10):

$$C_{b_j^t}^* = \arg \min_{C_{b_j^t}} \text{dist} \left(C_{b_i^{t-1}}, C_{b_j^t} \right) \quad (9)$$

同时保证:

$$\underbrace{|A(b_i^{t-1}) - A(b_j^t)| < \varepsilon_1}_A, \quad \underbrace{\text{dist} \left(C_{b_i^{t-1}}, C_{b_j^t} \right) < \varepsilon_2}_B \quad (10)$$

则令 $\text{label}(b_j^t) = i$. 其中, $A(b_i^t)$ 表示 b_i^t 团块的面积.

$$\text{dist} \left(C_{b_i^{t-1}}, C_{b_j^t} \right) = \sqrt{(x_i^{t-1} - x_j^t)^2 + (y_i^{t-1} - y_j^t)^2}$$

ε_1 和 ε_2 是两个阈值, ε_1 由目标定位团块的大小决定, ε_2 由目标运动的速度决定. A 项约束排除了经融合后依然粘连的团块的影响, B 项约束排除了由于前景漏检而将临近目标作为本目标的错误匹配.

利用 3 个高度层上的标志物, 根据式 (8) 同样可以计算出 3 个高度层之间的单应性矩阵. 将其中任意两层的定位信息利用单应性变换映射到另外一层上, 即可找出不同高度层之间的多目标定位信息的对应关系. 融合 3 层的处理结果, 进而实现对多目标的立体跟踪, 我们用立体框来表示跟踪结果.

4 实验结果及分析

为了验证本文提出的基于多相机的多目标跟踪算法的鲁棒性, 分别在室内和室外两种光照条件下利用多台相机从不同视角拍摄多组实验视频进行实验, 鉴于篇幅限制, 本文只列出两组实验结果. 实验视频分辨率均为 320 像素 \times 240 像素, 算法实现基于 Windows 操作系统, 采用 Visual Studio

2008 和 Open CV 2.1 作为软件平台, 计算机配置为 Pentium (R) Dual-Core CPU 2.0 GHz.

图 4 是利用不同算法提取前景似然信息的效果对比. 实验结果表明, 本文引入码本模型对背景进行建模, 能够很好的克服运动目标的空洞现象.

图 5 是对室外拍摄的实验视频进行 3 层定位的实验结果, 图 5 (a) 是 3 台相机从三个视角拍摄的原始视频图像, 图 5 (b) 是 3 个视角的前景似然图像. 我们在实验中均选取第 3 个视角作为参考图像, 基于多层将其他两个视角的前景似然图像映射到参考图像上, 得到基于多层的前景似然融合图像, 每一层

上的高亮区域, 即为每个目标在每一层上的定位信息. 图 5 (c) ~ 图 5 (e) 表示本组视频基于第 1 层 (地平面)、第 2 层和第 3 层的前景融合图像. 多层前景似然融合图像中不可避免的会产生一些次亮点, 为保证跟踪的准确性, 在进行跟踪之前对每层的前景似然融合图像设立一个阈值, 滤除概率较小次亮点的值. 采用连通区域检测算法来分割多目标在多层上的定位信息团块. 图 5 (f) ~ 图 5 (h) 表示第一组视频在 3 层上的定位结果. 图 6 是参考视角对多目标进行跟踪的实验结果. 同理, 可以选用其他两个视角作为参考视角, 图 7 为三个视角在同一时刻的跟踪

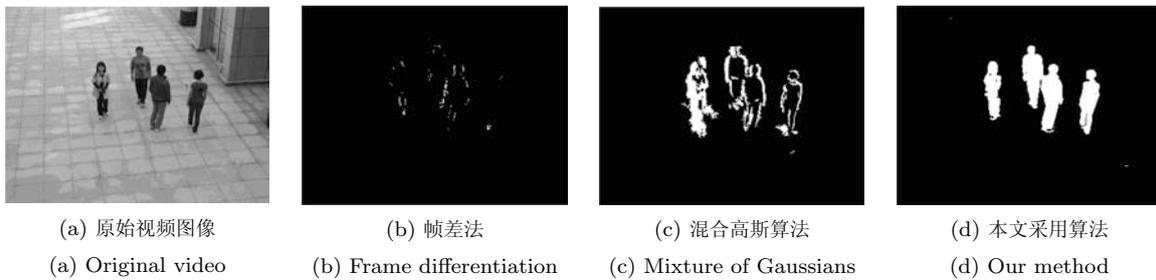
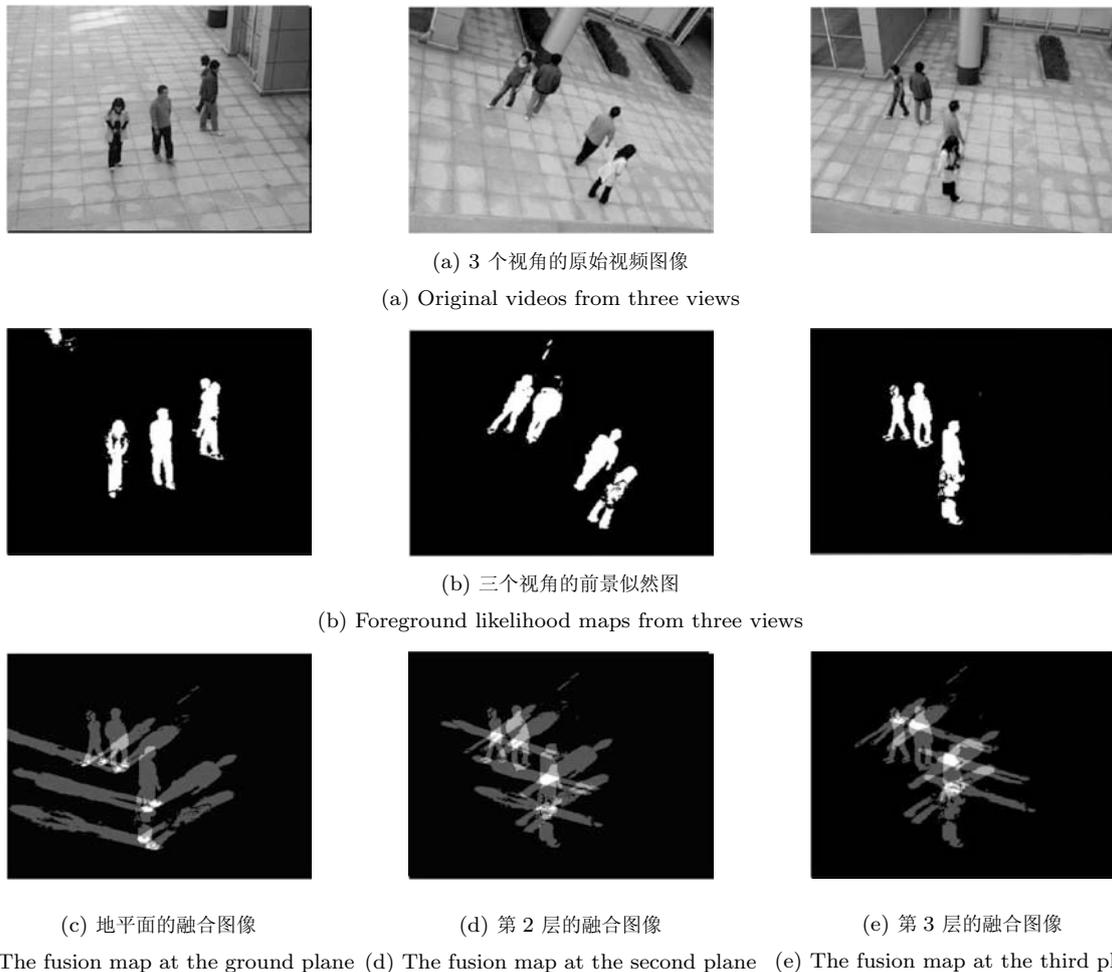


图 4 利用不同算法提取前景似然信息的效果对比

Fig. 4 Comparison of foreground likelihood maps obtained using different methods



(c) The fusion map at the ground plane (d) The fusion map at the second plane (e) The fusion map at the third plane

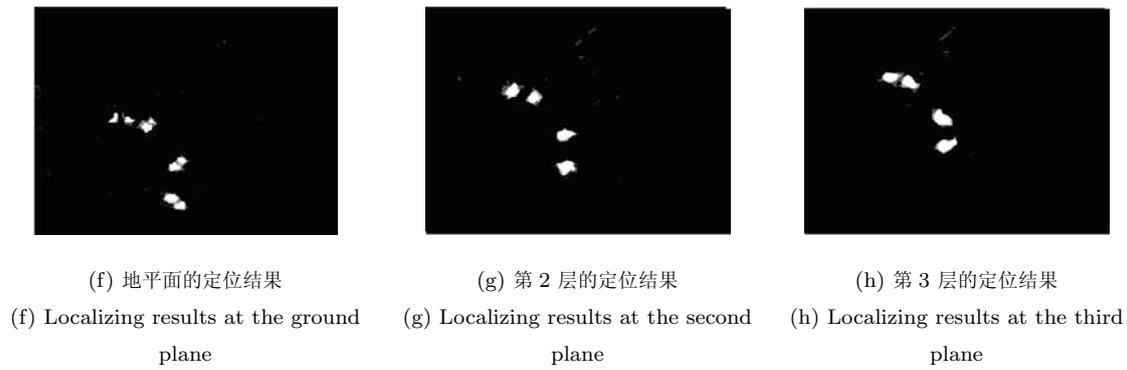


图 5 针对室外环境视频参考视角的多层定位结果

Fig. 5 The results of our localizing algorithm at multiple planes in the reference view for an outdoor data sets

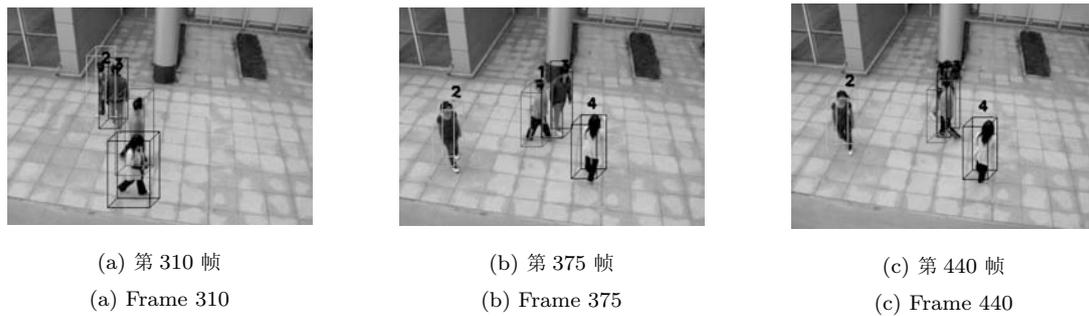


图 6 室外环境参考视角的跟踪结果

Fig. 6 Tracking results of the reference view for the outdoor video sequences

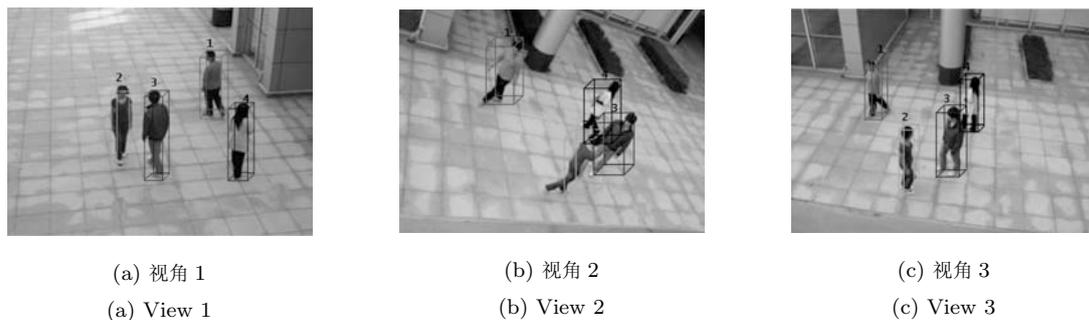


图 7 室外环境三个视角在同一时刻的跟踪结果

Fig. 7 Tracking results of three views at the same time for the outdoor video sequences

结果.

针对本组实验视频, 图 8 为本文算法与 Khan 算法检测误差的比较. 由实验数据可知, 由于本文采用码本模型检测运动目标, 有效地解决了阴影和空洞的问题, 减少了多视角融合后错误的定位信息, 因此目标检测误差比 Khan 算法有所降低, 为后续的定位跟踪提供了良好的前提条件. 图 9 为本文算法与 Khan 算法平均跟踪误差的比较, 从实验数据可以看出, 本文算法的平均跟踪误差也比 Khan 算法低. 由于本文不需要计算相机的隐消点, 从很大程度上提高了计算单应性矩阵的准确性, 进而提高了多

层定位和跟踪的准确性.

在室内利用 4 台相机拍摄实验视频, 图 10 是针对该组视频对 6 个目标进行 3 层定位的实验结果, 图 10(a) 是 4 个视角拍摄的原始视频图像, 图 10(b) 是 4 个视角的前景似然图像. 我们在实验中均选取第 4 个视角作为参考图像, 基于多层将其他两个视角的前景似然图像映射到参考图像上, 得到基于多层的前景似然融合图像, 每一层上的高亮区域, 即为每个目标在每一层上的定位信息. 图 10(c)~图 10(e) 表示本组视频基于第 1 层、第 2 层和第 3 层的前景融合图像. 图 10(f)~图 10(h) 表示第一组

视频在 3 层上的定位结果. 图 11 是参考视角对多目标进行跟踪的实验结果. 图 12 为三个视角在同一时刻的跟踪结果.

5 结论

本文在运动目标检测的基础上, 利用多相机之间的单应性知识融合了不同视角的前景似然信息,

在不同高度的 3 个平面上对多人进行定位. 利用最短路径优化算法处理多层定位信息, 实现对多目标的稳定跟踪. 本文提出基于多相机的跟踪算法不需要标定多相机, 也不需要计算多相机的隐消点, 所以在很大程度上简化了基于单应性的跟踪算法的复杂程度, 提高了单应性矩阵计算的准确性. 从实验结果可以看出, 我们的算法能够满足实时性要求, 在多目

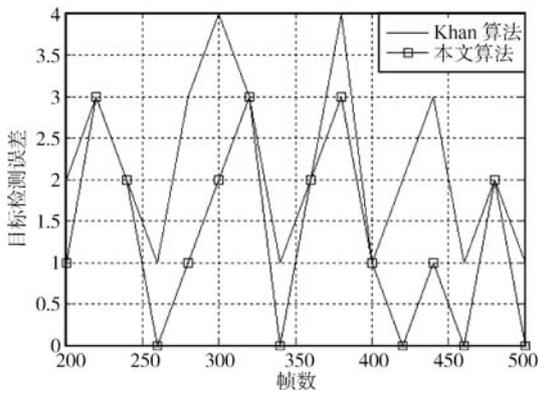


图 8 本文算法与 Khan 算法检测误差的比较
Fig. 8 Comparison of the detection errors of Khan's method and ours

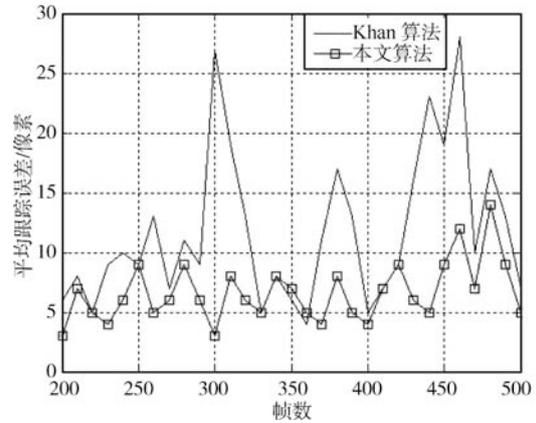
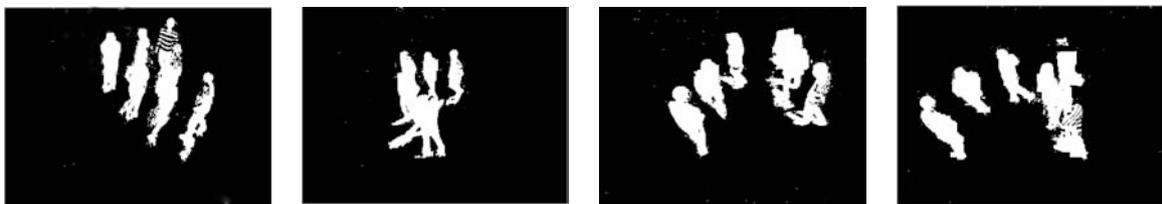


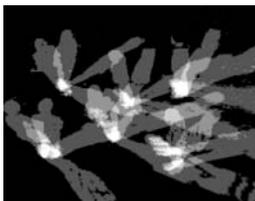
图 9 本文算法与 Khan 算法平均跟踪误差的比较
Fig. 9 Comparison of average track errors of Khan's method and ours



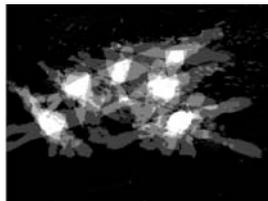
(a) 4 个视角的原始视频图像
(a) Original videos from four views



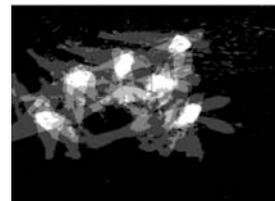
(b) 4 个视角的前景似然图
(b) Foreground likelihood maps from four views



(c) 地平面的融合图像



(d) 第 2 层的融合图像



(e) 第 3 层的融合图像

(c) The fusion map at the ground plane (d) The fusion map at the second plane (e) The fusion map at the third plane

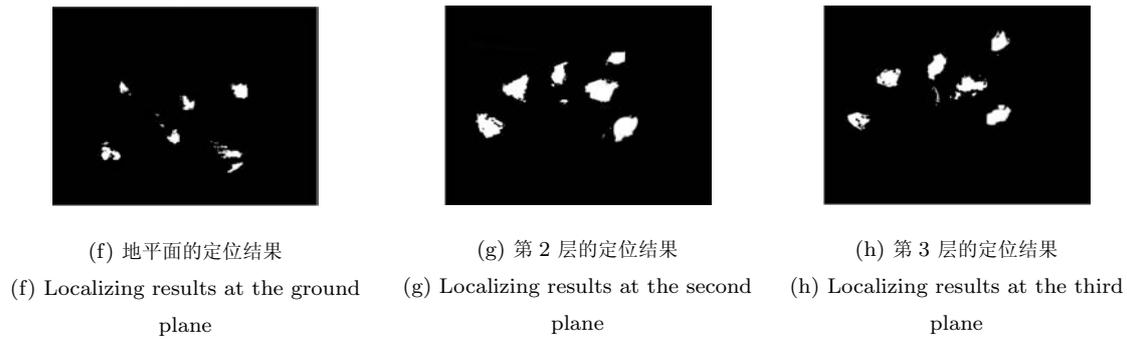


图 10 针对室内环境视频参考视角的多层定位结果

Fig. 10 The results of our localizing algorithm at multiple planes in the reference view for an indoor data sets

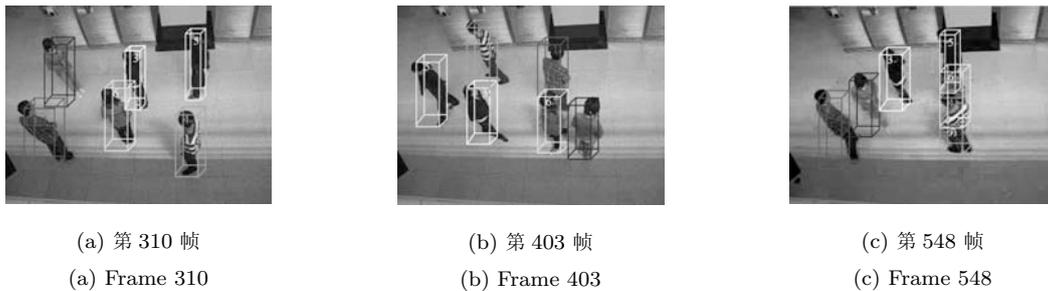


图 11 室内环境参考视角的视频跟踪结果

Fig. 11 Tracking results of the reference view for the indoor video sequences

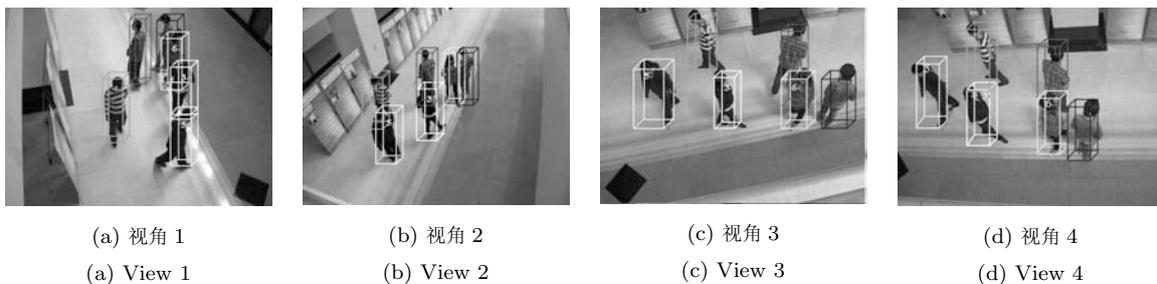


图 12 室外环境三个视角在同一时刻的跟踪结果

Fig. 12 Tracking results of four views at the same time for the indoor video sequences

标之间发生严重遮挡的情况下,依然可以准确地跟踪上多个目标.由于算法不涉及外观模型,所以对光照变化同样具有很强的适应性.未来的研究还可以在很多方面进行拓展,如:能否在运动目标检测的过程中加入颜色信息,能否结合目标的各个关节进行跟踪等.

References

- Hu W M, Tan T N, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2004, **34**(3): 334–352
- Yilmaz A, Javed O, Shah M. Object tracking: a survey. *ACM Computing Surveys*, 2006, **38**(4): Article 13, 1–45
- Zhao T, Nevatia R. Tracking multiple humans in crowded environment. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE, 2004. 406–413
- Hue C, Le Cadre J P, Perez P. Sequential Monte Carlo methods for multiple target tracking and data fusion. *IEEE Transactions on Signal Processing*, 2002, **50**(2): 309–325
- Comanicu D, Ramesh V, Meer P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564–577
- Jiang M X, Li M, Wang H Y. A robust combined algorithm of object tracking based on moving object detection. In: *Proceedings of the International Conference on Intelligent Control and Information Processing*. Dalian, China: IEEE, 2010. 619–622

- 7 Mittal A, Davis L S. M2Tracker: a multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, 2003, **51**(3): 189–203
- 8 Fleuret F, Berclaz J, Lengagne R, Fua P. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(2): 267–281
- 9 Kim K, Davis L. Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 98–109
- 10 Eshel R, Moses Y. Tracking in a dense crowd using multiple cameras. *International Journal of Computer Vision*, 2010, **88**(1): 129–143
- 11 Khan S M, Shah M, Tracking multiple occluding people by localizing on multiple scene planes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(3): 505–519
- 12 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(11): 1222–1239
- 13 Kim K, Chalidabhongse T H, Harwood D, Davis L. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 2005, **11**(3): 172–185
- 14 Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision (Second Edition)*. Cambridge: Cambridge University Press, 2003. 363–406
- 15 Wu Fu-Chao. *Mathematical Methods in Computer Vision*. Beijing: Science Press, 2008. 68–80
(吴福朝. 计算机视觉中的数学方法. 北京: 科学出版社, 2008. 68–80)
- 16 Rother C. A new approach for vanishing point detection in architectural environments. *Image and Vision Computing*, 2002, **20**(9–10): 647–655

- 17 Lv F, Zhao T, Nevatia R. Self-calibration of a camera from video of a walking human. In: Proceedings of the 16th International Conference on Pattern Recognition. Quebec, Canada: IEEE, 2002. 562–567



姜明新 大连理工大学信息与通信工程学院博士研究生. 主要研究方向为计算机视觉. 本文通信作者.

E-mail: jmx@mail.dlut.edu.cn

(**JIANG Ming-Xin** Ph.D. candidate at the School of Information and Communication Engineering, Dalian University of Technology. Her main re-

search interest is computer vision. Corresponding author of this paper.)



王洪玉 大连理工大学信息与通信工程学院教授. 主要研究方向为计算机视觉, 视频编码, 无线传感器网络.

E-mail: whyu@dlut.edu.cn

(**WANG Hong-Yu** Professor at the School of Information and Communication Engineering, Dalian University of

Technology. His research interest covers computer vision, video coding, and wireless video sensor networks.)



刘晓凯 大连理工大学信息与通信工程学院硕士研究生. 主要研究方向为图像处理.

E-mail: xiaokaigirl@mail.dlut.edu.cn

(**LIU Xiao-Kai** Master student at the School of Information and Communication Engineering, Dalian University of Technology. Her main research inter-

est is image processing.)