

基于视图不确定关系模型的地面目标位置估计和跟踪

季莎¹, 段宏博²

(1. 西安明德理工学院 通识教育学院, 西安 710124;

2. 哈尔滨石油学院 数理部, 哈尔滨 150028)

摘要: 为了实现地面上目标的位置估计和跟踪, 提出了一种摄像机平面视图与地面位置估计之间关系的建模理论, 推导了正态随机变量在投影下的变换; 首先分析了随机变量(摄像机图像平面上的位置估计和地面上的位置估计)在投影变换下的变换方式, 将变换后的随机变量中的 Cauchy 分量的强度与图像平面上目标的真实位置到地面无穷远处的直线投影之间的距离联系起来, 表明了当某些几何性质满足时, 投影变换会将正态分布映射为正态分布; 其次采用无迹变换来计算得到变换后的随机变量的矩; 最后采用得到的建模相关性设计了一种用于多个摄像机位置估计的最小方差估计器, 并应用于跟踪地面上动态环境中的多个目标; 实验结果表明, 提出的模型不仅具有较好的组合位置估计能力, 而且还能够利用这种模型得到的最小方差估计器有效地呈现和跟踪地面目标。

关键词: 摄像机网络; 位置估计; 随机变量; 视图; 单应变换; 最小方差估计器; 目标跟踪

Position Estimation and Tracking Based on View Uncertainty Relational Model for Targets on A Ground Plane

Ji Sha¹, DUAN Hongbo²

(1. College of General Education, Xi'an Mingde Institute of Technology, Xi'an 710124, China;

2. Department of Mathematics and Physics, Harbin Petroleum Institute, Harbin 150028, China)

Abstract: In order to achieve the position estimation and tracking of targets on a ground plane, a theory for modeling the relation between the camera plane view and position estimation on a ground plane is proposed, and the transformation of normal random variables under projection is deduced. Firstly, the transformation mode of random variables (position estimation on camera image plane and position estimation on a ground plane) under a projective transformation is analyzed, the strength of Cauchy component in the transformed random variable is associated with the distance of the true location of the target in the image plane from the projection of the line at infinity of the ground plane, which shows that when some geometric properties are satisfied, a projective transformation maps normal distributions to normal distributions. Secondly, the unscented transformation is used to calculate the moments of the transformed random variables. Finally, the obtained modeling correlation is used to design a minimum variance estimator of the position estimation by multiple cameras, and the tracking of multiple targets is applied in a dynamic environment on a ground plane. The experimental results show that the proposed model not only has a good combined position estimation ability, but also can effectively present and track targets on a ground plane by using the obtained minimum variance estimator of the model.

Keywords: camera network; position estimation; random variable; view; homography transformation; minimum variance estimator; target tracking

0 引言

在监控应用中, 通常会有多个摄像机观察地面上的运动目标, 利用多个摄像机获取的观测数据对目标进行跟踪是人们广泛关注的问题^[1-2]; 地面上目标的位置估计和跟踪是一个重要的推理问题。在许多监控中, 目标移动限制在一个平面(即地平面上)。地平面的存在提出了一个强约束

条件, 使得许多诸如跟踪、登记、测量和结构估计等实际问题需要条件良好的解决方案。

已有不少采用单个或多个摄像机来跟踪地面上目标的算法。为了对人体目标进行准确可靠的持续跟踪, 文献[3]提出了一种融合主颜色特征、纹理特征和时空拓扑特征的目标跟踪算法。该算法将人体区域分割成上、中、下3个目标子块, 分别利用最近邻聚类算法提取每个目标子块

收稿日期: 2022-12-06; 修回日期: 2023-01-10。

基金项目: 西安明德理工学院科研项目(2019XY06L14)。

作者简介: 季莎(1985-), 女, 硕士, 讲师。

通讯作者: 段宏博(1986-), 女, 硕士, 讲师。

引用格式: 季莎, 段宏博. 基于视图不确定关系模型的地面目标位置估计和跟踪[J]. 计算机测量与控制, 2023, 31(9): 283-289, 296.

的主颜色信息,并计算主颜色匹配率。同时根据目标关联信息的累计统计信息,采用增量学习建立和更新摄像机网络的时空拓扑关系。实验表明,该算法能有效地对非重叠视野多摄像机网络中出现的人体目标进行连续跟踪;文献[4]提出的算法利用图像和地平面之间的单应性线性化得到的观测噪声模型,采用卡尔曼滤波跟踪地面上的位置和速度;文献[5-6]提出的多摄像机跟踪算法采用将来自背景减法的输入投影到地面上的单应变换,数据关联和目标定位是通过每个摄像机的投影之间的一致性(共识)来实现的。这些算法的不同之处在于它们处理来自背景减法的输入以及随后跟踪目标的方式。例如,文献[5]提出,在每个视图中提取每个被分割的人的垂直轴,并采用他们在地面上投影的交点来定位这个人,然后采用粒子滤波器对交点进行滤波;而文献[6]提出的算法通过将每个视图的前景似然图像投影到一个参考视图来获得一致性,联合似然图像中的峰值则采用图形切割算法进行阈值化和分割;文献[7]提出了一种基于 SIFT (scale invariant feature transform) 特征匹配的运动目标检测和跟踪算法。在目标检测阶段,首先提取两帧带检测图像的 SIFT 特征点并进行特征匹配,然后计算两帧图像之间的几何变换矩阵,从而实现图像的几何对齐。再将几何对齐后的两幅图像进行差分,并在差分图像中寻找 SAD 最大值区域作为运动目标区域。在目标跟踪阶段,将已检测到的目标作为跟踪样本,与后检测到的目标区域进行 SIFT 特征匹配,结合论文提出的跟踪样本集更新机制实现目标跟踪;尽管文献[8-9]研究了如何确定单应变换空间上的测度和密度,但并没有研究如何将随机变量考虑进一个单应变换中,特别是在视觉应用的背景下;文献[10]提出了一种基于 Grassmann 流形和投影变换群的双模视频目标跟踪算法。算法运用 Grassmann 流形状态空间的内蕴几何特性,建立目标观模型,采用投影变换群建立目标的几何形变模型,将目标的状态变化看作是流形上的点沿着测地线移动,并设计了双模粒子滤波算法完成目标跟踪。实验结果表明,该算法对目标在经历大幅度非平面的几何形变、光照变化及部分遮挡情况下,都能够实现稳定的跟踪;文献[11]探讨了多摄像机视频监控系统中的多目标跟踪问题。将目标跟踪问题分为两部分:一是单摄像机下多目标跟踪的数据关联问题,二是多摄像机之间的航迹关联问题。针对传统多目标数据关联算法关联错误率高、实时性低的情况,将图像中目标的全局运动特征、HSV 颜色特征、LBP 特征和光流特征用于多假设数据关联算法(MHT)的关联矩阵的计算,提出了基于多特征融合的数据关联算法(MC-MHT),实验证明该算法在关联性能和实时性上都有所提高;文献[12]提出了一种复杂环境下的多目标实时跟踪方法,包括第一视频采集模块、第二视频采集模块、信息传输模块、数据库、背景剪出模块、视频处理模块、显示模块和控制器与报警模块。在进行静态背景剪出的过程中,通过采集大量监控区域的视频帧图像,通过对大量视频帧图像进行

整合处理,减少人物活动以及偶然事件对背景剪取的影响,同时对监控区中出现的有一定活动范围的物体进行分析,降低在追踪过程中该活动物体产生的噪声,从而降低分析难度,提高分析跟踪的准确度;还通过对全局监控区域建立坐标系,对目标对应的坐标点进行读取,根据目标所处坐标对目标进行接力跟踪,防止在对目标的追踪过程中丢失目标,以及无法及时完成接力;文献[13]提出了一种基于核相关滤波(KCF)的分步关联框架。算法首先采用基于卷积神经网络的目标检测器检测目标,获得准确的检测结果。然后,为了更好地预测目标的运动状态,通过加权融合三种特征的跟踪结果,为每个目标建立一个基于 KCF 算法的快速跟踪器。为了有效地降低碎片化轨迹的数量,算法通过跟踪片的置信度分步关联轨迹,并在遮挡的情况下,利用在线随机蕨重新检测目标。最后利用关联成功的检测信息自适应更新 KCF 算法中的尺度,实验结果表明,与现有算法相比,提出的算法能够在各种复杂的条件下,表现出强大和高效的跟踪性能。

上述这些算法主要侧重于跨视图的数据关联和跨遮挡的鲁棒跟踪,以及当背景较差时。而且全部算法都是同等处理来自不同视图的输入。

图 1 所示为某大学校园里某建筑物周围地面上的 4 个视图,地面上的同一目标以不同分辨率成像到不同的摄影机上,即摄像机放置以获得地面上不同的视图。显然,不同的视图以不同的分辨率解析目标,而相应地对目标在地面的位置估计就存在差异。



图 1 4 个摄像机观察一个以平面运动为主的场景

给定一个在全部 4 个摄像机上同时观察到的目标,如何估计它在地面上的位置并实现对目标的跟踪就是本文要研究的问题。对此,本文提出了摄像机平面视图与地面位置估计方差之间关系的建模理论。首先分析了随机变量(摄像机图像平面上的位置估计和地面上的位置估计)在投影变换下的变换方式,表明了当某些几何性质满足时,投影变换会将正态分布映射为正态分布;其次采用无迹变换(unscented transformation,也称 UT 变换)^[14]来计算得到变换后的随机变量的矩;最后采用得到的建模相关性设计了一种用于多个摄像机位置估计的最小方差估计器,并应用于跟踪地面上动态系统环境中的多个目标;实验结果表明,

本文提出的模型不仅具有较好的组合位置估计能力, 而且还能够利用这种模型得到的最小方差估计器有效地呈现和跟踪地面目标。

1 摄像机地面投影问题

图 2 所示为 3 个摄像机 A、B 和 C 观察一个平面 P 的示例, 其中摄像机 B 的图像平面平行于平面 P, 而摄像机 A 和 C 的图像平面是垂直于平面 P 的, 还在每个摄像机的图像平面上显示了表示该摄像机的图像平面上分布的等误差轮廓。摄像机与平面 P 之间的单应变换分别为 H_{AP} 、 H_{BP} 和 H_{CP} 。在这种情况下, H_{BP} 不是完全投射的, 它只是一个仿射变换, 相比之下, 而 H_{AP} 和 H_{CP} 引起了强烈的视角扭曲 (视图失真)。当将 B 投影到平面上时, 期望 B 上的密度保持其原始形式 (相同的等误差轮廓)。

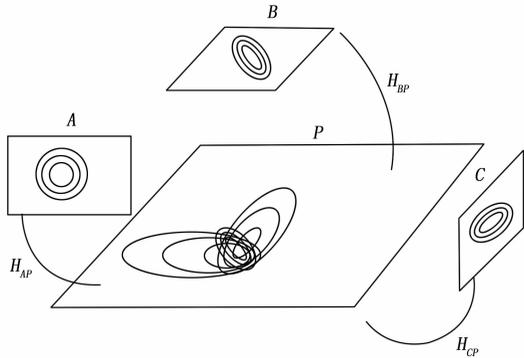


图 2 摄像机图像面上的密度及其向平面的变换示意图

假设一个场景安装有 M 个摄像机, 标记为 $1, 2, \dots, M$, 令 \mathbf{u}_i ($i=1, 2, \dots, M$) 为每个摄像机的图像平面上的位置估计。通过应用第 i 个摄像机与地面的对应单应变换, 就得到地面上的位置估计 $\mathbf{x}_i = H_{iP}\mathbf{u}_i$ 。假设 \mathbf{x}_i ($i=1, 2, \dots, M$) 为 M 个摄像机同时观察地面上的目标的估计值, 则希望将它们进行融合。

从图 2 可以清楚地看到, 即使 \mathbf{u}_i 是相同分布的, 但各个 \mathbf{x}_i 也不是相同分布的, 因此采用样本均值 $\mathbf{x} = (1/M) \sum_{i=1}^M \mathbf{x}_i$ 可能不是最佳的 (如在最小方差意义上)。一个好的估计器或融合方案应恰当利用各个 \mathbf{x}_i 的分布, 这就需要研究在投影变换下随机变量是如何变换的。

2 视图不确定关系的建模理论

2.1 单应变换下的随机变量分布

用下划线表示齐次坐标中的向量, 用大写字母表示矩阵和随机变量。令 $\mathbf{H} = [h_1, h_2, h_3]^T$ 为矩阵, 它定义从图像平面坐标到地面坐标的单应变换 (在两个平面上的某种坐标系选定下)。给定图像平面上一点 $\mathbf{u} = (u, v)^T \in R^2$, 就可以得到地面上相应的点 $\mathbf{x} = (x, y)^T$ 为:

$$\underline{\mathbf{x}} = \mathbf{H}\underline{\mathbf{u}} \quad (1)$$

\mathbf{u} 和 \mathbf{x} 之间的关系是齐次形式的线性关系。当 \mathbf{u} 和 \mathbf{x} 为有限点时, 式 (1) 可以改写为一个非线性方程:

$$\begin{cases} x = \frac{h_1^T \underline{\mathbf{u}}}{h_3^T \underline{\mathbf{u}}} \\ y = \frac{h_2^T \underline{\mathbf{u}}}{h_3^T \underline{\mathbf{u}}} \end{cases} \quad (2)$$

令 $Z_U = (Z_u, Z_v)^T$ 为随机变量, 它是对一个摄像机的图像平面上的位置的不确定性的建模, $Z_x = (Z_x, Z_y)^T$ 为采用式 (2) 通过变换 Z_U 得到的随机变量, 即:

$$\begin{cases} Z_x = \frac{h_1^T Z_U}{h_3^T Z_U} = \frac{h_{11}Z_u + h_{12}Z_v + h_{13}}{h_{31}Z_u + h_{32}Z_v + h_{33}} \\ Z_y = \frac{h_2^T Z_U}{h_3^T Z_U} = \frac{h_{21}Z_u + h_{22}Z_v + h_{23}}{h_{31}Z_u + h_{32}Z_v + h_{33}} \end{cases} \quad (3)$$

式中, $\mathbf{h}_i = (h_{i1}, h_{i2}, h_{i3})^T, i=1, 2, 3$ 。

为了更进一步处理, 需要知道 Z_U 的分布。在没有噪声的情况下, 图像平面上的位置为 $\mathbf{u} = (u, v)^T$, 然而, 来自于以下两个源的噪声破坏了 \mathbf{u} 的观测:

1) 成像: 在图像平面上采样以生成帧, 会在每个像素的单一像素区域引入一个均匀噪声, 这个误差对于成像模式是主要的;

2) 建模误差: 一个更重要的误差来源是假设平面运动。通常我们研究的是平面上忽略尺寸大小的点 (理想质点) 的位置, 但在实际中, 目标是有形状大小的, 且研究的点与目标上的某个参考点是相关联的。此外, 这样的点还受到自遮挡和建模误差导致的视差的影响, 与成像传感器引入的噪声不同, 建模误差的统计特征对于本文研究的问题是特定的。

在实际中, 图像平面上的密度可能是多模态的, 通常采用非参数方法 (如核和粒子滤波等) 进行建模。然而, 由于单应变换式 (3) 的非线性, 对于复杂的图像平面密度来说, 解析处理和推理就变得十分困难。因此, 假设图像平面上的状态空间就是位置 (因此是二维的), 且 Z_U 是双变量正态分布的:

$$Z_U \sim N(\mathbf{m}_0, \mathbf{S}_0) \quad (4)$$

$$\begin{cases} \mathbf{m}_0 = \begin{pmatrix} m_0^u \\ m_0^v \end{pmatrix} \\ \mathbf{S}_0 = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \end{cases} \quad (5)$$

还假设协方差矩阵 \mathbf{S}_0 是一个常数, 独立于均值 \mathbf{m}_0 。

由于 Z_U 具有正态分布, 则 $\mathbf{h}_i^T Z_U$ ($i=1, 2, 3$) 是单变量正态分布的, 即:

$$\mathbf{h}_i^T Z_U \sim N(m_i, \sigma_i^2) \quad i=1, 2, 3 \quad (6)$$

$$\begin{cases} m_i = \mathbf{h}_i^T \mathbf{m}_0 \\ \sigma_i^2 = h_{i1}^2 \sigma_u^2 + h_{i2}^2 \sigma_v^2 + 2h_{i1}h_{i2}\rho\sigma_u\sigma_v \end{cases} \quad (7)$$

此外, 当 $i=j$ 时, $\mathbf{h}_i^T Z_U$ 与 $\mathbf{h}_j^T Z_U$ 之间的相关系数 ρ_{ij} 为:

$$\rho_{ij} = \frac{h_{i1}h_{j1}\sigma_u^2 + h_{i2}h_{j2}\sigma_v^2 + (h_{i1}h_{j2} + h_{i2}h_{j1}\sigma_v^2)\rho\sigma_u\sigma_v}{\sigma_i\sigma_j} \quad (8)$$

由式 (3) 和 (7) 可见, Z_x 和 Z_y 均由正态密度比^[15]得到。首先考虑 Z_x 的分布形式。 Z_x 可以写成 $\mathbf{h}_1^T Z_U$ 和 $\mathbf{h}_3^T Z_U$

的比值 (见式 (3))。尽管 Z_x 为两个非标准相关正态随机变量的比值, 但其形式可以由两个独立标准正态随机变量的比值得到。具体来说, Z_x 可以写成:

$$Z_x \sim c_x + d_x W_{a_x, b} \quad (9)$$

式中, c_x 和 d_x 为标量, $W_{a_x, b}$ 为独立标准正态比, 其形式为:

$$W_{a_x, b} = \frac{Z_1 + a_x}{Z_2 + b} \quad a_x > 0 \quad b > 0 \quad (10)$$

式中, Z_1 和 Z_2 为独立的标准正态随机变量。

c_x, d_x, a_x 和 b 的表达式如下:

$$c_x = \rho_{13} \frac{\sigma_1}{\sigma_3} \quad (11)$$

$$d_x = \sqrt{1 - \rho_{13}^2} \frac{\sigma_1}{\sigma_3} \quad (12)$$

$$a_x = \frac{|m_1/\sigma_1 - \rho_{13}m_3/\sigma_3|}{\sqrt{1 - \rho_{13}^2}} = \frac{|(h_1/\sigma_1 - \rho_{13}h_3/\sigma_3)^T m_0|}{\sqrt{1 - \rho_{13}^2}} \quad (13)$$

$$b = \frac{|m_3|}{\sigma_3} = \frac{|h_3^T m_0|}{\sigma_3} \quad (14)$$

Z_x 的分布严重依赖于 $W_{a_x, b}$ 的分布, 因为其分布可以从 $W_{a_x, b}$ 的缩放 (d_x) 和平移 (c_x) 得到。 $W_{a_x, b}$ 的分布可以表示为两个密度的加权和, 其中一个密度 f_1 为 Cauchy, 另一个密度 f_2 有一个解析表达式。 $W_{a_x, b}$ 的分布如下:

$$f_{W_{a_x, b}}(t) = p_x f_1(t) + (1 - p_x) f_2(t) \quad (15)$$

f_1 为柯西密度:

$$f_1(t) = \frac{1}{\pi(1+t^2)} \quad (16)$$

式中, $0 < p_x < 1$ 是一个与柯西项相关的权重:

$$p_x = e^{-1/2(a_x^2 + b^2)} \quad (17)$$

分量 f_2 的形式为:

$$\begin{cases} f_2(t) = \frac{q \int_0^q e^{-1/2(x^2 - q^2)} dx}{\pi(1+t^2)(e^{1/2(a_x^2 + b^2)} - 1)} \\ q = \frac{b + a_x t}{\sqrt{1+t^2}} \end{cases} \quad (18)$$

Z_y 密度的推导类似, 分别用 $h_1^T Z_U$ 和 $h_2^T Z_U$ 替换相关的统计量, 密度的形式保持不变, 即两个分量的混合, 其中一个为柯西。此外, 对于一般的问题设置, Z_x 和 Z_y 是相关的, 这和柯西分布的其他矩一样是未知的, 这时, 分析 Z_x 和 Z_y 的密度形式及其对单应变换 H 的依赖关系以及目标在平面上的真实位置 x_0 是非常重要的。

Cauchy 分量的存在意味着 Z_x 和 Z_y 的均值、方差和一般的高阶矩都没有定义, 这导致了弱/强的大数定律不适用的情况。然而, 当 Cauchy 分量很弱 (即 p_x 和 p_y 很小) 时, 就可以忽略 Cauchy 分布的存在而对密度进行近似。首先研究当 Cauchy 分量的概率可以忽略不计时的几何设置。

引理: 在 Z_x (Z_y) 的分布中, Cauchy 分量的强度 p_x (p_y) 依赖于图像平面上的均值 m_0 与图像平面上无穷远处的直线 l_∞ 的投影的距离, 即 $l_\infty^T m_0$ 。

证明: 单应变换 H 可以分解成下列变换:

$$H = H_A H_P = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \quad (19)$$

式中, H_A 为仿射变换, H_P 表示采用齐次形式的 $l_\infty = (l_1, l_2, l_3)^T$ 的变换的全部投影部分。这意味着 $h_3 = \lambda l_\infty$ ($\lambda \neq 0$ 为比例因子), $m_3 = h_3^T m_0 = \lambda l_\infty^T m_0$ 为点 m_0 到 l_∞ 的距离。 m_0 越靠近 l_∞ , $b = |m_3|/\sigma_3$ 就越小。随着 b 的增大, $e^{-1/2b^2}$ 就减少。由式 (16) 可以看出, 随着 b 的增大, Z_x 中的 Cauchy 分量的强度 p_x 减小。 Z_y 的讨论类似。

推论: 当成像点离无穷远直线足够远时, Z_x 和 Z_y 分布中的 Cauchy 分量强度就可以忽略不计, 或

$$m_3 = \frac{h_3^T m_0}{\sigma_3} \gg 1 \quad (20)$$

强度 p_x 和 p_y 分别依赖于 a_x 和 a_y 。式 (13) 表明, a_x 与成像均值 m_0 离直线 $h_1/\sigma_1 - \rho_{13}h_3/\sigma_3$ 的距离成比例。

可见, 当柯西分量的强度可以忽略不计时, 总的分布可以用正态密度以极高的精度来近似, 其参数可以通过数值方法得到。

在投影变换下, 正态密度映射到正态密度 (假设式 (20) 的要求满足) 意味着在投影变换下随机变量的变换可以以逐点方式建模为仿射变换, 其参数依赖于均值 m_0 。由于参数对均值 m_0 的依赖性, 故仿射不会延伸到区域上的变换。然而, 考虑到这种映射的平滑性, 一般而言, 局部仿射近似在小邻域上仍然有效, 只要成像的区域离无限远直线的投影足够远。文献 [16] 中就采用了这种局部仿射模型用于几何分组。

上述理论的主要结果是在式 (20) 的条件满足下, 建立起了变换随机变量的矩的存在性。然而, 这些矩的解析计算并不容易, 本文采用近似方法来计算 Z_x 的矩。

2.2 采用无迹变换计算矩

计算 Z_x 的均值和协方差的一个有效近似方法是采用无迹变换。采用无迹变换, 通过称为 sigma 点的点和相关的权重的确定性选择来计算均值和协方差。

变换后的随机变量的矩计算如下。首先, 生成 sigma 点 v_i ($i = 0, \dots, 2n_u$) 和相关权重 w_i , 其中 n_u 是 Z_U 的维数, 在本文设置中 $n_u = 2$ 。sigma 点的生成采用以下选择方案:

$$\begin{cases} v_0 = m_0 \\ v_i = m_0 + (\sqrt{(\kappa + n_u)S_0})_i \quad i = 1, \dots, n_u \\ v_{n_u+i} = m_0 - (\sqrt{(\kappa + n_u)S_0})_i \end{cases} \quad (21)$$

式中, κ 是一个常数标量, $(\sqrt{(\kappa + n_u)S_0})_i$ 是 $(n_u + \kappa)S_0$ 的矩阵平方根的第 i 行。权重取值如下:

$$w_i = \begin{cases} \kappa/(\kappa + n_u) & i = 0 \\ 1/2(\kappa + n_u) & i \neq 0 \end{cases} \quad (22)$$

每个 sigma 点都采用单应性传播:

$$x_i = \frac{1}{h_3^T v_i} \begin{pmatrix} h_1^T v_i \\ h_2^T v_i \end{pmatrix} \quad i = 0, \dots, 2n_u \quad (23)$$

Z_x 的均值 μ_0 和变量协方差矩阵 \sum_0^X 的估计值为:

$$\begin{cases} \hat{\boldsymbol{\mu}}_0 = \sum_{i=0}^{2n_u+1} \omega_i \boldsymbol{\chi}_i \\ \hat{\Sigma}_0^X = \sum_{i=0}^{2n_u+1} \omega_i (\boldsymbol{\chi}_i - \hat{\boldsymbol{\mu}}_0)(\boldsymbol{\chi}_i - \hat{\boldsymbol{\mu}}_0)^T \end{cases} \quad (24)$$

κ 的值选择很重要。考虑到随机变量 Z_U 是二元的, 我们选取 $\kappa=3-n_u=1$, 均值矩阵和变量协方差矩阵的估计精度可以达到二阶, 比简单的一阶线性更精确。

下一节将通过应用每个摄像机上的无迹变换学习得到的变量协方差模型, 来得到一个用于融合多个摄像机的位置估计值的最小方差估计器。

3 采用多个摄像机的位置估计

假设有 M 个摄像机, 摄像机视图和地面之间的单应变换矩阵为 \mathbf{H}_i ($i=1, 2, \dots, M$), 本节提出一种融合位置估计的算法。令 U_i 为随机变量, 它对第 i 个摄像机的图像平面上的目标位置进行建模, 假设随机变量 $\{U_i\}_{i=1}^M$ 是统计独立的, 对于成像 (传感器) 噪声来说, 这一假设是合理的。

U_i 的分布来自于跟踪算法或检测算法的输出, 但我们只对分布的均值 m_i 和协方差 S_i 感兴趣。基本跟踪器采用卡尔曼滤波器^[17]或粒子滤波器^[18]来获得均值和协方差的估计作为输出; 采用 2.2 给出的无迹变换来得到随机变量 \mathbf{X}_i 的均值 $\hat{\boldsymbol{\mu}}_i$ 和变量协方差矩阵 $\hat{\Sigma}_i^X$ 的估计值, \mathbf{X}_i 是根据第 i 个摄像机的估计对地面上的位置进行的建模。现在来得到地面上位置的最小方差估计。假设估计值 $\hat{\boldsymbol{\mu}}_i$ 是无偏的 (或 $E(\hat{\boldsymbol{\mu}}_i) = \boldsymbol{\mu}$ 为真实目标位置), 则最小方差估计值 $\hat{\boldsymbol{\mu}}_X = (\hat{\mu}_x, \hat{\mu}_y)^T$ 为:

$$\hat{\boldsymbol{\mu}}_X = \sum_{i=1}^M (\hat{\Sigma}_i^X)^{-1} \sum_{MV} \hat{\boldsymbol{\mu}}_i \quad (25)$$

$$\hat{\Sigma}_{MV} = (\sum_{j=1}^M (\hat{\Sigma}_j^X)^{-1})^{-1} \quad (26)$$

在线性类估计器中, 式 (25) 定义的是最小方差意义上的最优估计值。最后, 最小方差估计器的方差就可由式 (25) 计算得到。

现在, 给定观察地面的一组摄像机的单应变换矩阵, 就可以计算并绘制出最小方差估计器的方差作为地面上实际均值 (的函数)。

3.1 动态系统

在大多数情形下, 位置跟踪是跟时间有关的 (而不仅仅是静态估计)。下面构建一个离散时间动态系统用于地面上的目标位置跟踪。状态空间由地面上的位置和速度构成。令 \mathbf{x}_t 表示时刻 t 的状态空间, 即 $\mathbf{x}_t = [x_t, y_t, \dot{x}_t, \dot{y}_t]^T \in R^4$ 。采用恒定速度模型的状态演化方程为:

$$\mathbf{x}_t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \boldsymbol{\omega}_t \quad (27)$$

式中, $\boldsymbol{\omega}_t$ 为噪声过程。观测模型采用第 2 节得到的均值和协方差模型。观测向量 $\mathbf{y}_t \in R^{2M}$ 正好是采用无迹变换从每个摄像机估计出的位置均值, 观测模型为:

$$\mathbf{y}_t = \begin{bmatrix} \hat{\boldsymbol{\mu}}_1 \\ \vdots \\ \hat{\boldsymbol{\mu}}_M \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \cdots & & \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_t + \mathbf{A}(\mathbf{x}_t) \boldsymbol{\Omega}_t \quad (28)$$

式中, $\boldsymbol{\Omega}_t$ 是一个具有单位变量协方差矩阵的零均值噪声过程, $\mathbf{A}(\mathbf{x}_t)$ 建立整个噪声的变量协方差矩阵, 定义为:

$$\mathbf{A}(\mathbf{x}_t) = \begin{bmatrix} \sum_x^1(\mathbf{x}_t) & \cdots & 0_{2 \times 2} \\ \vdots & \ddots & \vdots \\ 0_{2 \times 2} & \cdots & \sum_x^M(\mathbf{x}_t) \end{bmatrix} \quad (29)$$

式中, $0_{2 \times 2}$ 为全部元素为零的 2×2 矩阵, $\sum_x^i(\mathbf{x}_t)$ 为当地面上的目标的真实位置为 \mathbf{x}_t 时 \mathbf{Z}_X^i 的变量协方差矩阵。

式 (28) 的观测模型是一个完整的观测器模型的多视图扩展, 它包含 2 个重要的特性。

- 1) 不同视图的观测值的噪声特性是不同的, 变量协方差不仅依赖于视图, 还依赖于目标的真实位置 \mathbf{x}_t , 这种依赖关系被编码在 \mathbf{A} 中;
- 2) \mathbf{x}_t 的最大似然估计就是最小方差估计器。

现在, 采用依赖于数据集性质的 Kalman 滤波或粒子滤波, 就可以实现跟踪目标。实验中将给出这两种方法的实例。

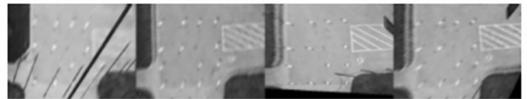
4 实验结果

4.1 静态估计的方差映射

利用从摄像机网络中每个视图计算得到的单应变换矩阵, 并对每个摄像机采用无迹变换, 就能计算出地面上感兴趣区域的变量协方差矩阵。图 3 所示为图 1 的 4 个摄像机



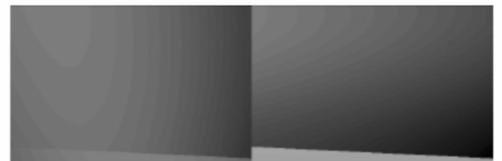
(a) 摄像机视图



(b) 从 (a) 摄像机对应视图生成的地面顶视图



(c) 每个摄像机在地面上的 Z_x 和 Z_y 的方差估计



(d) 沿两个轴的最小方差位置估计器的方差

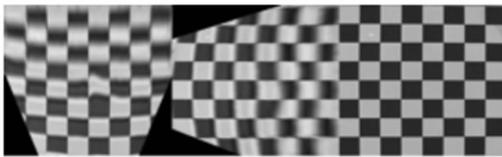
图 3 图 1 所示摄像机设置的方差估计

网络的结果，这些方差被绘制成地面上的函数。可以看出，对于相应的视图，当一个目标接近无穷远处的直线时，方差随之增大。最小方差估计器（式（25））的方差也可以计算为目标在地面上的真实位置的函数。这种情形在摄像机放置问题中可能很有用，在这种情形下，给定一组摄像机和地面上感兴趣的区域，目的是放置更多的摄像机，以提高跟踪目标的性能。

图 4 所示为采用高度不对称摄像机设置的 3 个摄像机成像一个棋盘得到的类似结果，其中两个摄像机放置在离地面非常近的地方，且沿棋盘的两个正交轴，每个摄像机只能准确地估计一个方向的位置。与最右边列相对应的摄像机仅用于为 4.2 节的跟踪实验提供地面的真实情形。通过结合同时来自两个摄像机的估计值，利用变量协方差映射，就可以得到在两个方向上具有低方差的估计值（图 4（e））。



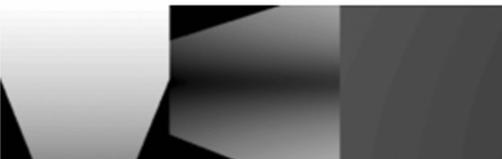
(a) 每个摄像机的视图



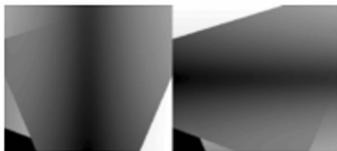
(b) 由每个摄像机估计的单应变换构造的顶部视图



(c) 地面上的 Z_x 的方差估计



(d) 地面上的 Z_y 的方差估计



(e) 仅采用与左列和中列对应的摄像机的最小方差估计器的方差

图 4 3 个摄像机成像一个棋盘得到的类似结果

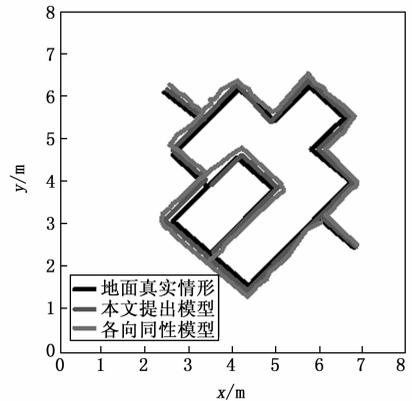
4.2 多摄像机跟踪

4.2.1 图 4 所示 3 个摄像机设置下的跟踪

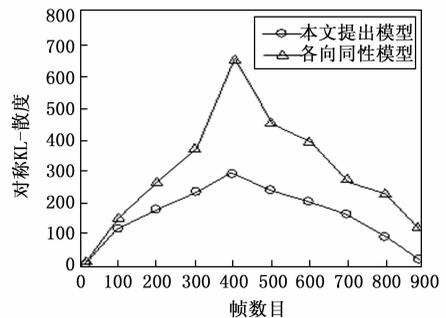
下面给出一个实验来说明本文模型在多摄像机跟踪中

的重要性，特别是当摄像机的视图高度不对称时（见图 4）。研究的区域还是棋盘，采用激光指针来创建目标，采用基于颜色的分割方法来检测激光指针创建的目标，与右列对应的摄像机仅用于提供地面真实情形。

比较 2 种系统的跟踪：一种系统的观测模型采用 3.1 节的观测模型，另一种系统对全部摄像机同等处理，在摄像机上采用各向同性模型。2 种系统都采用卡尔曼滤波跟踪器。从数据中得到的状态方差用作地面真实情形（即由第 3 台摄像机完成）。为了定量评价，计算 2 种系统的输出（均值和变量协方差）与地面真实情形之间的对称 KL-散度（KLD, kullback leibler divergence）^[19]。得到的跟踪结果如图 5 所示。从图 5（a）可以看到，采用本文模型得到的跟踪结果与地面真实情形非常接近，而且非常平滑，而采用各向同性模型的跟踪结果明显偏离真实情形，且存在较大的不规则轮廓；从图 5（b）可以看到，采用对称 KL-散度的地面真实情形输出时，本文模型的 KL-散度明显小于采用各向同性模型的 KL-散度，这说明本文模型能更好地呈现地面的真实情形，这说明本文模型对各个摄像机获取到的数据有较好的融合能力。



(a) Kalman 滤波跟踪输出



(b) 采用对称 KL-散度的地面真实情形输出的比较

图 5 2 种系统的跟踪比较

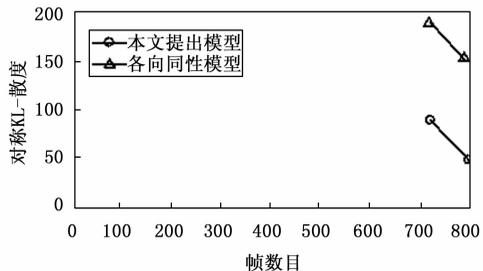
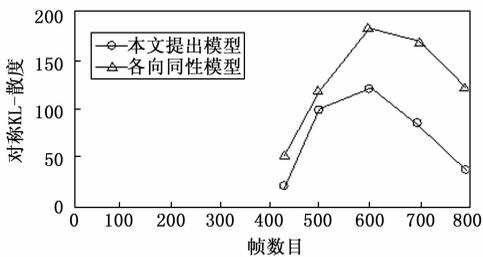
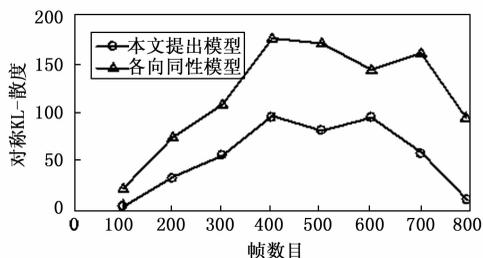
4.2.2 图 3 所示 4 个摄像机设置下的跟踪

本节的多目标跟踪是用来测试本文所提出模型对于图 3 所示的真实摄像机放置的有效性。提取每个摄影机上每个背景减去的一小片的最底层点，并投影到世界平面上，该数据与跟踪器的关联采用传统的联合概率数据关联滤波（joint probabilistic data association filtering, JPDAF）^[20]，

并与来自每个摄像机的数据分别进行关联, 使用标记获取地面真实情形。仍然对上节的两个观测模型进行比较, 即一个采用本文提出的模型, 另一个采用跨视图的各向同性建模。最后, 采用粒子滤波进行跟踪, 并对由于遮挡而丢失的数据点进行跟踪。图 6 所示为实验得到的跟踪结果。实验中对 800 帧视频进行了跟踪测试, 在 100、430 和 720 帧处依次引入 3 个目标, 如图 6 (b) 所示, 每个图形分别对应 3 个不同的目标。从图 6 (b) 可见, 采用跨视图的各向同性建模相应的 KL-散度值要高于采用本文模型相应的 KL-散度值, 即本文提出的模型一致地得到了较低的对地面真实情形的 KL-散度, 说明本文模型能更好地呈现地面的真实情形。



(a) 用彩色和高缩放最佳观看在不同时刻的跟踪输出快照



(b) 采用对称KL-散度对地面真实情形进行跟踪的评价

图 6 4 个摄像机数据集上 3 个目标的跟踪比较

5 结束语

本文提出了一种摄像机对地面目标的位置估计和跟踪的建模, 推导了正态随机变量在投影下的变换。具体而言, 将变换后的随机变量中的 Cauchy 分量的强度与图像平面上目标的真实位置到地面无穷远处的直线投影之间的距离联系起来, 阐明了该结果对于位置估计和跟踪的相关性; 一个可能的扩展是采用核方法来对密度进行建模, 但分析推理变得更具挑战性, 未来打算研究这种模型在摄像机放置和评价中的应用。

参考文献:

- [1] 张兴国. 地理场景协同的多摄像机目标跟踪研究 [D]. 南京: 南京师范大学, 2014.
- [2] 阳小燕, 任家富, 刘长江. 基于视野分界线的多摄像机运动目标跟踪算法 [J]. 计算机工程与应用, 2016, 52 (9): 208 - 212, 238.
- [3] 王建功, 林国余. 基于多特征融合的多摄像机人体跟踪方法 [J]. 吉林大学学报 (信息科学版), 2014, 32 (6): 675 - 683.
- [4] LIU Z B, HUANG J X, HAN J W, et al. Human Motion Tracking by Multiple RGBD Cameras [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27 (9): 2014 - 2027.
- [5] JIANG X Y, KORNER M, HAASE D, et al. A graph-based MAP solution for multi-person tracking using multi-camera systems [C] // In Proceedings of 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 2014: 343 - 350.
- [6] MOUSSE M A, MOTAMED C, EZIN E C. People counting via multiple views using a fast information fusion approach [J]. Multimedia Tools and Applications, 2017, 76 (5): 6801 - 6819.
- [7] 康晓梅, 穆柯楠, 康 贤. 基于 SIFT 特征匹配的运动目标检测及跟踪方法 [J]. 电子设计工程, 2018, 26 (1): 174 - 177.
- [8] SZPAK Z L, CHOJNACKI W, ERIKSSON A, et al. Sampson distance based joint estimation of multiple homographies with uncalibrated cameras [J]. Computer Vision and Image Understanding, 2014, 125 (2): 200 - 213.
- [9] MAYBANK S J. A Fisher-Rao Metric for Curves Using the Information in Edges [J]. Journal of Mathematical Imaging and Vision, 2016, 54 (3): 287 - 300.
- [10] 谢英红, 庞彦伟, 韩晓微, 等. 基于 Grassmann 流形和投影群的目标跟踪 [J]. 仪器仪表学报, 2016, 37 (5): 1140 - 1147.
- [11] 廖逸琪. 多摄像机下多目标跟踪相关技术研究 [D]. 南京: 南京理工大学, 2016.
- [12] 连国云, 孙宏伟. 一种复杂环境下多目标实时跟踪方法 [P]. CN110047092A. 2019-07-23.
- [13] 周海英, 杨 阳, 王守义. 基于核相关滤波器的多目标跟踪算法 [J]. 激光与光电子学进展, 2018, 55 (9): 296 - 302.

(下转第 296 页)