

# 多目标识别预测

luobai

2023 年 8 月 3 日

Draft

目 录

<b>1</b>	<b>SIMPLE ONLINE AND REALTIME TRACKING</b>	<b>3</b>
1.1	ABSTRACT . . . . .	3
1.2	简介 . . . . .	3
1.3	文献综述 . . . . .	3
1.4	方法 . . . . .	4
1.4.1	Detection . . . . .	4
1.4.2	Estimation Model 评估模型 . . . . .	4
1.4.3	数据关联 . . . . .	5
1.4.4	Creation and Deletion of Track Identities 创建和删除跟踪标识 . . . . .	5
1.5	实验 . . . . .	5
1.5.1	度量 . . . . .	6
<b>2</b>	<b>ExtendedKalmanFilter 仿真及实现</b>	<b>6</b>
<b>3</b>	<b>卡尔曼滤波学习</b>	<b>6</b>
3.1	马尔可夫链 . . . . .	6
3.2	基本动态系统模型 . . . . .	6
<b>4</b>	<b>长宽比与角度的关系</b>	<b>6</b>
<b>5</b>	<b>能量机关</b>	<b>8</b>
5.1	时间轴对齐和击打位置预测 . . . . .	8

Draft

# 1 SIMPLE ONLINE AND REALTIME TRACKING

一篇来自 IEEE 的快速多目标跟踪算法。

## 1.1 ABSTRACT

本文探讨了一种多目标跟踪的语用方法，其主要目的是为在线和实时应用高效地关联目标。检测质量被认为是影响跟踪性能的一个关键因素，改变检测器可以提高跟踪高达 18.9%。尽管只使用了卡尔曼滤波器和匈牙利算法等常见技术的基本组合进行跟踪，但这种方法达到了与最先进的在线跟踪器相当的精度。

此外，由于我们的跟踪方法的简单性，跟踪器更新在 260 赫兹的速率，这是超过 20 倍的速度比其他国家的最先进的跟踪器。

## 1.2 简介

MOT(多目标识别)问题可以看作是一个数据关联问题，其目的是在视频序列中跨帧关联检测。为了辅助数据关联过程，跟踪器使用各种方法来模拟场景中物体的运动和外观。本文采用的方法是通过观察最近建立的视觉 MOT 基准。

首先，包括多假设跟踪 (MHT) 和联合概率数据关联 (JPDA) 在内的成熟数据关联技术重新出现，占据了 MOT 基准的许多顶级位置。其次，唯一不使用聚合信道滤波器 (ACF) 检测器的跟踪器也是排名最高的追踪器，表明检测质量可能会阻碍其他追踪器。此外，精度和速度之间的权衡似乎相当明显，因为最准确的追踪器的速度被认为是太慢的实时应用 (见图 1)。随着传统数据关联技术在顶级在线跟踪器和批量跟踪器中的突出地位，以及顶级跟踪器所使用的不同检测方法的使用，本文探讨了 MOT 的简单性以及它的性能。

为了与 Occam's Razor 保持一致，在跟踪中忽略了检测组件一些外观特征，只使用边界框的位置和大小进行运动估计和数据关联。此外，有关短期和长期闭塞的问题也被忽视，因为它们很少发生，其明确的治疗介绍导致跟踪框架的不必要复杂性。我们认为将复杂性以对象重新识别的形式引入到跟踪框架中会增加大量的开销，这可能会限制它在实时应用中的使用。

这篇论文的主要贡献是：

- ◇ 我们在 MOT 的背景下利用基于 CNN 的检测能力。提出了一种基于卡尔曼滤波器和匈牙利算法的实用跟踪方法，并根据最近的 MOT 基准进行了评估。
- ◇ 代码开源，以帮助建立避免碰撞应用方面的研究试验和吸收的基准方法。

本文组织如下：

- ◇ 第 2 节提供了一个在多目标跟踪领域的相关文献的简短回顾。
- ◇ 第 3 节介绍了拟议的精益跟踪框架
- ◇ 在第 4 节中论证了关于标准基准序列的拟议框架的有效性
- ◇ 第 5 部分提供了学习结果的总结，并讨论了未来的改进

## 1.3 文献综述

传统上，MOT 是使用多目标跟踪 (MHT)[7] 或联合概率数据关联 (JPDA) 过滤器 [16, 2] 来解决的，它们在目标分配存在高度不确定性的情况下延迟了困难的判断。

这些方法的组合复杂性随着被跟踪目标的数量呈指数增长，这使得它们在高动态环境中的实时应用变得不实用。

最近, Rezatofighi et al.[2], 在可视化 MOT 中重新讨论了 JPDA 公式 [16], 目的是通过利用求解方面的最新发展, 利用 JPDA 的有效近似来解决组合复杂性问题。

类似地, Kim 等人 [3] 对每个目标使用外观模型来修剪 MHT 图, 以实现最先进的性能。然而, 这些方法仍然延迟决策, 使他们不适合在线跟踪。

许多在线跟踪方法旨在通过在线学习建立单个物体本身的外观模型 [17,18,12] 或全局模型 [19,11,4,5]。除了外观模型之外, 运动通常被用来协助将检测关联到轨迹 [1,19,4,11]。当只考虑作为二部图匹配模型的一对一对应时, 可以使用全局最优解, 如匈牙利算法 [15][10,20]。

Geiger 等 [20] 的方法在两阶段过程中使用了匈牙利算法 [15]。首先, 通过将相邻帧间的检测关联起来形成小路, 在这些帧中, 几何和外观线索被组合起来形成亲和矩阵。然后, 轨迹相互关联, 以桥接由遮挡造成的破碎轨迹, 再次使用几何和外观线索。

这种两步关联方法限制了这种方法的批处理计算。我们的方法受到了 [20] 的跟踪组件的启发, 但是我们将这种关联简化为一个单一的阶段, 并使用下一节中描述的基本线索。

## 1.4 方法

该方法通过对目标检测的关键部分进行描述, 包括将目标状态传播到未来帧、将当前检测与现有目标相关联以及管理被跟踪目标的生命周期。

### 1.4.1 Detection

为了利用基于 CNN 的检测的快速发展, 我们使用更快的区域 CNN (FrRCNN) 检测框架 [13]。FrRCNN 是一个端到端的框架, 由两个阶段组成。第一阶段提取特征并为第二阶段提出区域, 然后对提出的区域中的对象进行分类。这个框架的优点是参数可以在两个阶段之间共享, 从而创建一个有效的检测框架。此外, 网络架构本身可以交换到任何设计, 使快速实验的不同架构, 以提高检测性能。

在这里, 我们比较了 FrRCNN 提供的两种网络架构, 即 Zeiler 和 Fergus (FrRCNN (ZF)) 的架构 [21] 和 Simonyan 和 Zisserman (FrRCNN (VGG16)) 的更深层架构 [22]。通过这项工作, 我们应用了学习的默认参数为 PASCAL VOC 挑战的 FrRCNN。由于我们只对行人感兴趣, 我们忽略所有其他类别, 只通过输出概率大于 50% 的人检测结果的跟踪框架。

在我们的实验中, 我们发现当比较 FrRCNN 检测和 ACF 检测时, 检测质量对跟踪性能有显著的影响。这是演示使用的序列验证集应用于现有的在线跟踪器 MDP [12] 和跟踪器在这里提出。表 1 显示了最佳的检测器 (FrRCNN (VGG16)) 导致最佳的跟踪精度的 MDP 和提出的方法。

### 1.4.2 Estimation Model 评估模型

在这里, 我们描述了目标模型, 即表示和运动模型, 用于传播目标的身份到下一帧。

我们使用线性恒速模型来近似每个物体的帧间位移, 该模型与其他物体和摄像机的运动无关。每个目标的状态建模如下

$$x = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}]^T \quad (1.1)$$

其中

- $u$  目标中心的水平像素位置
- $v$  目标中心的垂直像素位置

$s$	目标边界框的比例面积
$r$	目标边界框的长宽比
$\dot{u}$	—
$\dot{v}$	—
$\dot{s}$	—

其中  $u$  和  $v$  表示目标中心的水平和垂直像素位置，而比例尺  $s$  和  $r$  分别表示目标边界框的比例 (面积) 和长宽比。请注意，纵横比被认为是常数。当检测与目标相关联时，检测到的包围盒用于更新目标状态，其中速度分量通过卡尔曼滤波框架得到最优解 [14]。如果没有检测相关的目标，其状态是简单的预测，没有修正使用线性速度模型。

### 1.4.3 数据关联

在为现有目标分配检测时，通过预测目标在当前帧中的新位置来估计每个目标的**边界盒几何形状**。然后，将分配代价矩阵计算为每个检测器与现有目标的所有预测边界盒之间的交集距离 (IOU)。

利用匈牙利算法对指派问题进行了优化求解。另外，在目标重叠的检测值小于  $IOU_{min}$  的情况下，施加最小 IOU 来拒绝分配。

发现边界框的 IOU 距离隐含地处理了由于目标通过而造成的短期遮挡。具体来说，当目标被遮挡物覆盖时，只有遮挡物被检测到，因为 IOU 距离适当地有利于具有相似尺度的检测。这允许两个遮挡目标被纠正与检测，而覆盖的目标是不受影响，因为没有任何分配

### 1.4.4 Creation and Deletion of Track Identities 创建和删除跟踪标识

当对象进入和离开图像时，需要相应地创建或销毁独特的标识。对于创建跟踪器，我们考虑任何重叠小于  $IOU_{min}$  的检测，以表明存在未跟踪的对象。

跟踪器初始化使用的几何边界盒与速度设置为零。由于速度在这一点没有被观测到，**速度分量的协方差被初始化为较大的值，反映了这种不确定性**。

此外，新的跟踪器然后经历一个试用期，目标需要与检测相关联，以积累足够的证据，以防止跟踪误报。

如果没有为 TLost 帧检测到轨道，则终止它们。这可以防止跟踪器数量的无限增长和由于长时间的预测而导致的本地化错误，而不需要检测器的修正。在所有实验中，TLost 设置为 1 有两个原因。首先，恒定速度模型不能很好地预测真实的动态，其次，我们主要关注的是帧到帧的跟踪，其中目标的重新识别超出了这项工作的范围。此外，早期删除丢失的目标有助于提高效率。如果对象重新出现，跟踪将在新标识下隐式恢复。

## 1.5 实验

我们评估我们的跟踪实现的性能-在一个由 MOT 基准数据库 [6] 设置的不同的测试序列集，其中包含移动和静态摄像机序列。为了调整初始 Kalman filter 协方差、IOUmin 和 TLost 参数，我们使用了与 [12] 中报道的相同的训练/验证分割。所使用的检测架构是 FrRCNN (VGG16)[22]。来自 [22] 的源代码和示例检测可在线获得。

### 1.5.1 度量

由于很难使用一个单一的评分来评估多目标跟踪性能，我们使用了 [24] 中定义的评估指标，以及标准的 MOT 指标 [25]

MOTA(up)      多目标跟踪精度 accuracy

MOTP(up)      多目标跟踪精度 precision

## 2 Extended Kalman Filter 仿真及实现

## 3 卡尔曼滤波学习

卡尔曼滤波 (Kalman filter) 是一种高效率的递归滤波器 (自回归滤波器)，它能够从一系列的不完全及包含噪声的测量中，估计动态系统的状态。卡尔曼滤波会根据各测量量在不同时间下的值，考虑各时间下的联合分布，再产生对未知变数的估计，因此会比只以单一测量量为基础的估计方式要准。

### 3.1 马尔可夫链

马尔可夫链又称离散时间马可夫链 (discrete-time Markov chain, 缩写为 DTMC)，因俄国数学家安德烈·马尔可夫得名，为状态空间中经过从一个状态到另一个状态的转换的随机过程。该过程要求具备“无记忆”的性质：下一状态的概率分布只能由当前状态决定，在时间序列中它前面的事件均与之无关。这种特定类型的“无记忆性”称作马可夫性质。马尔科夫链作为实际过程的统计模型具有许多应用。

### 3.2 基本动态系统模型

卡尔曼滤波建立在线性代数和隐马尔可夫模型 (hidden Markov model) 上。

其基本动态系统可以用一个马尔可夫链表示，该马尔可夫链建立在一个被高斯噪声（即正态分布的噪声）干扰的线性算子上的。

系统的状态可以用一个元素为实数的矢量表示。随着离散时间的每一个增加，这个线性算子就会作用在当前状态上，产生一个新的状态，并也会带入一些噪声，同时系统的一些已知的控制器的控制信息也会被加入。同时，另一个受噪声干扰的线性算子产生这些隐含状态的可见输出。

## 4 长宽比与角度的关系

对目标物体进行建系

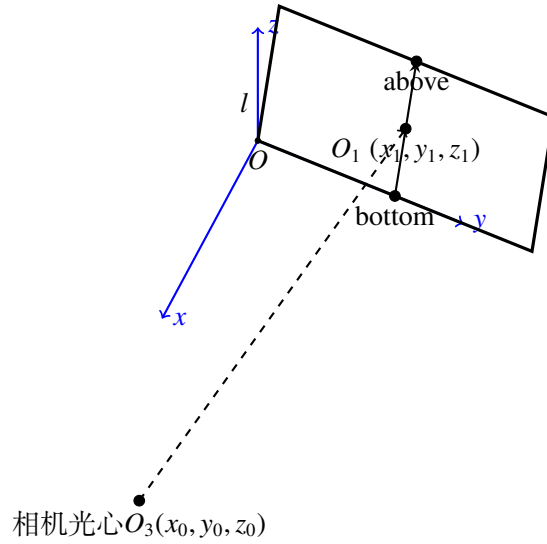


图 1: 示意图 1

如上图所示，此时观测到的角度一定在 90 度左右，其实我们观测到的角度是投影在  $OZY$  平面上的角度。并且假定车中心为  $O_2(x_2, y_2, z_2)$ 。根据投影角度，我们可以大致拟合出其绕  $O_2$  的转动角度。虽然存在透视变化，但基本准确。

假设  $O_3, O_2$  处于一条直线，以  $y$  轴正方向， $x$  轴正方向为正定义旋转角度为从上往下看绕  $y$  轴负方向旋转角度 (0-360)。设转动角度  $\theta$ 。

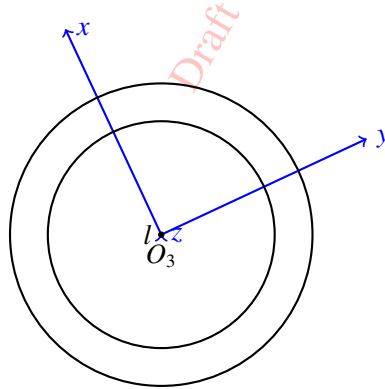


图 2: 示意图 1

$$h = (r_2 - r_1) \tan(75)$$

$$x = (r_2 - r_1) \cos(\theta)$$

$$-\tan \theta_1 = \tan \frac{x}{h} = \frac{\tan(75)}{\cos(\theta)}$$

其中  $\theta_1$  是我们观测到的角度， $\theta$  是转动角度。可得

$$\theta = \arccos\left(\frac{\tan 75}{\tan \theta_1}\right)$$

## 5 能量机关

### 5.1 时间轴对齐和击打位置预测

在之前的

Draft