多目标跟踪算法研究

一、 多目标跟踪算法概述

1. MOT 算法

多目标跟踪，其主要任务为同时跟踪视频画面中的多个目标，为目标分配 ID 并维持其 ID 的长久有效性，得到目标的运动轨迹。需解决的问题包括频繁遮挡、轨迹初始化和终止、相似的外观、多目标间的相互影响等。近些年以来，多摄像机多目标跟踪技术有了不小的发展，国内外的专家学者主要是围绕着 Re-ID（Person Re-identification 也称行人再识别，简称为 Re-ID）、多摄像机之间的交互，摄像机的拓扑结构和时空关系等问题进行了大量的讨论与相关研究。

多目标跟踪算法根据初始化方法的不同可分为 Detection-Based Tracking（DBT）和 Detection-Free Tracking（DFT），前者为基于检测的跟踪，即优先目标检测，再将检测结果与已存在的轨迹进行匹配链接。后者则需要在第一帧手动标记定量目标，在后续帧定位物体。DFT 的跟踪模式仅能跟踪已标记目标，而 DBT 的跟踪模式可以自动发现新的目标，移除已消失的目标，更适应如今多目标跟踪算法的应用场景。传统的基于检测的多目标跟踪算法，如 Nevatia 算法，该算法将 MOT 问题转化为 3 层逐步细化的数据关联问题：最底层通过样本和样本的关联实现简单的目标检测；中间层使用最大化后验概率方法实现轨迹初步拼接；最高层使用 EM 算法优化中间层检测结果得到多目标跟踪轨迹。该算法基于底层目标检测，利用层次递进的思想解决多目标跟踪问题，但检测器较为简单，且场景预设为单出入口，局限性较大。同时算法内部需设定的参数过多，如 EM 算法方差、最大时间差等，实现同等效果较为困难。近年来，随着深度学习的快速进步，目标检测算法的精度不断提升，基于检测的多目标跟踪算法得到了长足的发展。2016 年提出的 SORT 算法凭借其出色的跟踪性能获得了极大关注，标志着利用深度学习目标检测器的多目标跟踪方法逐渐崛起。

基于检测的多目标跟踪可以分为在线跟踪（Online）与离线跟踪（Offline）。在线多目标跟踪是一种逐帧渐进的跟踪方式，与人眼实时跟踪目标过程类似，首先要对每个运动目标进行识别确认（目标检测），然后对其下一步的行动进行预测（轨迹预测），最终根据目标的运动方向（运动模型）、外观形体（外观模型）等特征与之前的轨迹进行关联（数据关联匹配）。可以看出，在线多目标跟踪仅仅是将这一帧的信息与先前的信息或轨迹关联，且必须具备实时性，即运行速度

可以达到视频正常的播放速度（通常算法速度达到 20 FPS 即可算作实时算法）。离线多目标跟踪的输入是一段完整的视频，并已获得了目标检测结果。与在线多目标跟踪算法逐帧渐进方式不同的是，离线多目标跟踪可获取全局信息后再进行匹配关联。可理解为将目标检测算法的结果看作一个集合，将轨迹看作集合的一种划分，多目标跟踪的任务可化为子集优化的任务。前者的优点在于可实时输出，但其易受目标遮挡和检测器漏检、误检的影响。后者的输出从理论上获得了全局最优的结果，但实时性较差。

1.1 在线多目标跟踪算法

1.1.1 SORT 算法

1.1.2 DeepSORT 算法

1.1.3 Fair 算法

1.1.4 JDE 算法

将检测模型和嵌入模型整合为一个模型，即 JDE 模型。

1.1.5 MOTDT 算法

MOTDT 算法整体框架和 Deep SORT 相似。SORT 算法及 Deep SORT 算法进行轨迹匹配时，从检测和跟踪的输出中收集候选项作为集合，输入到匹配机制，导致候选集合中存在冗余候选项。针对该问题，MOTDT 算法通过融合对象分类器和跟踪器置信度制定了统一的轨迹评分机制，将产生的检测框和预测框的标准置信度作为非最大抑制（NMS）输入，从而获得无冗余候选项。引入轨迹评分机制，对 NMS 的每一个输入进行评分，利用卡尔曼滤波预测弥补漏检。数据关联方面，MOTDT 算法为改善拥挤场景中的类别遮挡问题，融合外观表示和空间信息，将现有轨道与所选候选项分层关联。外观表示与 Deep SORT 算法相似，通过深度神经网络训练行人重识别（Re-ID）数据集解决当目标重新出现时的检测问题，即使用神经网络学习得到对象外观信息，提取相应特征向量，再利用获得的特征间距离确定相似性。行人重识别（Re-ID）指在局部位置失去跟踪目标，也可在目标再次出现时将目标与之前的轨迹关联。以往行人重识别的难点在于长距离跟踪情况下，相似度高的行人仍有可能不是同一目标。经验证，在 IDF1 和 IDs 方面，Re-ID 特征优于传统的手工特征。同样，通过从对应的数据集中学习外观表示，例如车辆重新识别，可以轻松将提出的跟踪框架转移到其他类别。分层关联主要体现

在优先将候选集和候选项基于空间信息和外观信息与现有轨迹进行关联匹配，余下的候选项和未关联的轨迹再基于 IOU 进行匹配。

1.2 离线多目标跟踪算法

1.2.1 POI 算法

POI 算法是一种离线多目标跟踪算法，结合了目标检测和基于深度学习的外观特征。基本思路：在每一帧的输入上，用检测算法检测行人的位置，然后利用行人检测框的外观特征进行前后帧行人框的匹配，从而实现对行人的跟踪。该算法使用 Faster R-CNN 作为目标检测算法，选用随机采样动态尺度的多尺度训练策略，此外采用 skip pooling 和 multi-region 策略联合不同尺度和水平层次的特征。经验证，采用该策略可有效降低 FN+FP 的值（即准确率的分母）。嵌入模型训练方面，同时使用 softmax loss 和 triplet loss。2 种损失函数的设定是为了更加清晰地区分目标间的表观特征。相似度的判断以特征间的余弦距离为参考，相关为 1，无关为 0。

1.2.2 IOU 算法

IOU 算法是一种离线多目标追踪算法。该算法指出，随着目标检测算法精度的不断提高，当检测精度与视频帧率较高时，可以结合简单的目标检测算法与 IOU，再通过设定阈值来判断前景与背景即可完成目标跟踪任务。即在高帧率（25 FPS）、高精度的目标检测算法情况下，结合检测与时间步长间的空间重叠完成跟踪。该算法优缺点明显，但由于未引入任何帧间信息、运动模型、外观模型，漏检和错检问题难以解决，若出现频繁遮挡、目标形变的情况，会导致 ID 频繁切换，且极其依赖目标检测算法的性能。但该算法的优点是框架简单、速度快，若设定在某些环境简单的情况下，效果明显。该算法在 MOT16 数据集上速度约 3000 FPS，且该策略对于结合检测和分割有一定启发，如果能找到高速的追踪方法，将检测和追踪进行结合效果更佳。

1.2.3 LMP 算法

LMP 算法的提出虽距今已有一段时间，但表现仍出色。该算法主要针对遮挡影响及行人重识别展开。LMP 算法提出一种新结构，结合深度网络中提取的整体表示特征和从最先进的姿态估计模型中抽取的身体姿态进行判断，从而提高准确率。主要创新点在于：数据关联方面，LMP 算法是在最小代价多分割问题（MP）

基础上改进的，将数据的关联匹配看作一种基于图的分解、聚类问题。通过设置一个基于边的目标函数来选择能最大化相同目标和不同目标概率的分量对，从而完成行人重识别的任务并改善遮挡影响，有效降低 IDs。

1.2.4 基于多线索的多目标跟踪算法

在实际应用系统中，短期线索跟踪的特点是基于附近帧进行预测、更新，只包含当前帧的信息，易受遮挡和相似目标影响，但表现效果较好。与之相对的，长期线索则包含轨迹的运动外观等特征，能应对一定遮挡和相似目标的影响，但表现力不及短期线索。2019 年提出的基于多线索的多目标跟踪算法融合了长期线索与短期线索的优点，并在一个网络结构中突出两者的优点以应对 MOT 场景中的复杂情况。

该系统创新点在于分别设立 SOT 网络与 Re-ID 网络来承载短期线索与长期线索。设定切换感知分类器（SAC）以提升匹配效率，该分类器可利用目标的短期和长期线索、检测结果和切换器来预测检测结果是否与目标匹配。SOT 网络的搭建基于单目标跟踪器 Siam RPN，使用私有行人数据调整网络，根据匹配情况计算跟踪质量，该网络具备部分发现和识别功能，以弥补其他组件产生的纰漏。

2. MTMCT 算法

多摄像机下的视频跟踪本质上是一个多摄像机匹配的问题，即在同一时刻建立不同摄像机视角下运动物体之间的对应关系。多摄像机匹配是计算机视觉一个比较新的课题。在这方面，已有一些研究工作。

2.1 基于区域的算法

基于区域的匹配方法是把人看作一个运动区域，利用运动区域的特征来建立不同视角下人的对应关系。现有的研究利用颜色直方图来估计人的区域颜色分布，通过比较颜色直方图来建立多摄像机的匹配。Mittal 等为人的运动区域建立了高斯颜色模型，然后利用这些模型建立匹配。该算法首先分割每一个图像，然后比较每一对图像的区域，找到分割结果的中心，通过匹配中心点找到场景中可能对人通信的三维点（3D，three-dimension），最后将 3D 点投影到 2D 平面，使用拒绝框架方案对人的 2D 位置给出鲁棒性估计。

在这类方法中，颜色是最常用的一个区域特征。颜色特征虽然比较直观，然而其在匹配方面很不鲁棒。这是因为:首先，基于颜色的匹配依赖于人的衣服的

颜色。当场景中两个人的衣服的颜色相同时，这种匹配方法则很有可能无法区分这两个人，从而产生错误的匹配。其次，光照和视角的变化会影响颜色特征。比如，一个人所穿的衣服前面是白色，而背后是黑色。当两个摄像机分别放置在这个人的前面和后面的时候，这两个视角下观测到的人因其颜色差异很大会被误认为是两个不同的人。

2.2 基于模型的跟踪算法

在多摄像头监控系统中有基于活动模型和基于空间模型的跟踪算法，有些情况下还将这两种算法结合起来使用。已有研究提出了使用网络单位图构建基于网络的活动模型，或基于目标运动轨迹的空间和频率分布的活动模型。尽管这些活动适合单个摄像头场景和视野域重叠的多个摄像头场景，但是不适合非重叠的多摄像头系统，即这些方法没有考虑摄像头之间有非重叠区域或者多摄像头系统中有遮挡区域的情况。之后又继续提出了无监督学习的活动模型和多特征路径模型。利用来自摄像头的空间和时间信息，不受摄像头特征和方向的限制，在非重叠的摄像头区域（即盲区）建立活动路径的内连网，确定各摄像头间的空间关系。但是，由此算法产生的活动路径内连网仍包含一些冗余链。基于空间模型的多摄像头跟踪方法需要计算多摄像头网络的拓扑结构，用来校准场景中的多摄像头。通过使用校准的摄像头可以确定将像素坐标转换成 3D 坐标的变换，也能更精确的确定每一个摄像头中可见地平面在空间上的扩展。

2.3 基于特征的跟踪

基于特征的跟踪包括特征提取和特征匹配两个过程。提取的特征有目标的运动趋势、颜色，形态，位置，动态性等。此前，基于特征的跟踪设计了在视觉监控中使用目标颜色进行跟踪的系统。跟踪系统设计有多个摄像机监视动态场景框架的一部分，首先对来自目标的颜色建模，然后度量目标之间的差值，颜色跟踪补充了空间跟踪，能够在大时隙和空间上未校准的摄像头中使用。在上述所做工作基础上又提出了在非重叠多摄像头系统中实现目标跟踪的外形模型。该模型适合有光照变化的一般场景，是一种基于特征的多摄像头跟踪方法。为了处理目标从一个摄像头进入另一个摄像头所观察到的颜色变化，给出了低维子空间中所有从某一特定摄像头到另一摄像头的亮度转换函数，使用这个子空间计算外形的相似性。方法中已知系统在训练阶段目标间的通信，学习摄像头对之间的亮度转换

函数的子空间；一旦训练完成，利用目标的位置和外形线索使用最大后验概率估计框架来分配通信。

2.4 基于点的算法

多摄像机匹配更为可行的方法可能是基于特征点的方法。即将人看成一系列的特征点，不同视角下人的匹配就转化为特征点的匹配。特征点的匹配是基于一定的几何约束。根据所选用的几何约束的不同，这类方法可以划分为两大子类:三维方法和二维方法。

三维方法有两种策略来建立多摄像机的匹配。一种是先将不同视角下提取的特征点投影到一个三维空间，然后比较这些投影点。其标准是不同视角下相对应的特征点应该对应于同一个三维投影点。另外一种策略是利用极线约束来进行匹配。人的上半身中线的一些点被选为特征点。然后利用极线约束来寻找匹配。在其方法中，只需要对相邻的摄像机进行标定。这两种策略都需要对摄像机进行标定。当监控场景中所使用的摄像机数目很大时，摄像机标定将是一个比较巨大任务。而且，即使对于同一个人，在不同视角下所提取的特征点并不一定对应于同一个三维坐标点。这种情况使得特征点对之间的匹配变得不确定。

为了克服三维方法的缺点，有些研究者提出了利用二维信息来建立多摄像机之间的匹配。当人进入或者离开某个摄像机的可视区域的时候，利用该摄像机在另一个摄像机的图像平面中可视区域边界线的投影来建立匹配。当一个人进入其中一个摄像机的可视区域时，则在其他摄像机的可视区域中离该摄像机投影边界线最近的人被认为是与之对应的。当人在某个摄像机的可视区域的中部出现的时候，则利用两个摄像机图像平面之间的单映关系约束来建立匹配。

二、 在利用再识别方法的多目标跟踪中引入上下文感知

1 介绍

人的再识别(Re-ID)正在被深入研究，但主要是相机间目标关联。采用外观建模的方法解决该问题，其中鲁棒描述符的需求成为优先级。很少有作品将 Re-ID 与摄像头内跟踪相结合在大多数相关的工作中，Re-ID 被用于后处理方案来关联轨迹，而在本文的工作中，它被用于在线方式的上下文感知。

本文利用再识别技术在多目标跟踪器中加入上下文感知，提高其在

线跟踪性能。为了实现这一目标，根据其外观信息的完整性，将目标标记为独立的、遮挡的或被遮挡的。针对每个分类，采用不同的跟踪策略以获得最优结果。在跟踪失败的情况下，提出了一种在线自动重新识别技术，以减少对同一目标的多重身份分配。

半拥挤环境下多目标的在线跟踪是计算机视觉中一个非常活跃的研究领域。一个跟踪器必须克服像外观变化，类内区分，场景遮挡和上述组合等挑战。跟踪也可以扩展成一个摄像头网络，目标沿着不同的摄像头相关联。基于检测的跟踪器由于其改进的性能和准确性而成为流行的跟踪选择。他们使用外观模型通过在每一帧中重新检测物体来找到它的新位置。跟踪器的观察模型基于训练有素的检测器，能够在许多尺度和观察角度上定位一个对象类(行人、汽车等)。跨帧检测响应的关联是基于时空约束的。

在物体间遮挡过程中，为了进一步提高同一类目标之间的可分辨性，针对每个目标分配自适应目标分类器，已被广泛提出。这些方法中的外观模型旨在使同一类的目标可区分。

本文的主要贡献是引入了一个上下文感知的目标标记过程，使用Re-ID 技术来实现跟踪参数的动态调整。根据每个目标的外观信息的完整性将其标记为独立的、遮挡的或被遮挡的。结合检测器和分类器构造的基线跟踪器，使用所提出的上下文感知标记增强。对于每个类别，检测器和分类器的响应采用不同的融合权值。此外，Re-ID 还可以防止目标获取多重身份。每当一个新的目标被引入遮挡区域时，它通常会获得一个新的身份号码(ID)。提出了一种基于 Re-ID 的自动识别技术，用于识别新发现的目标，并在需要时将其与之前的 id 联系起来。提出的方法以因果关系的方式运行，并在每个框架上在线进行关联决策。

2 基线跟踪器

实现了一个基线跟踪器作为跟踪参考，以验证所提出的可附加模块的优点。最初，在每一帧上应用一个对象检测器来构造一个对象定位的响应映射。跨框架的对象关联，然后，促进了在线学习基于分类器。分类单元在每个目标外观上随时间变化。关联是基于这样的假设，即物体

不能在结果帧之间剧烈移动。在我们的实验中，基线跟踪器进行行人检测，压缩跟踪器作为目标特定分类器。

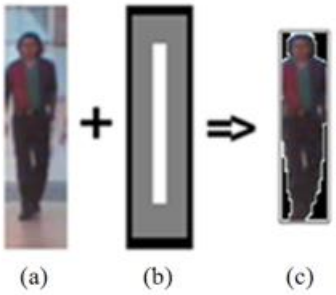

图 2　裁剪帧(a)的分割由 trimap(b)引导，前景像素为白色，前景像素可能为灰色，背景像素为黑色。因此，目标(c)仅用相关像素表示。

探测响应被描述为矩形区域，其中每个区域都由一个边界框描述。为每一个连续帧，对应过程尝试关联根据时空约束将每个区域与现有的轨道之一关联。单目标跟踪是可行的基于检测的关联。然而，在多目标跟踪场景中，关联变得具有挑战性，因为在遮挡期间，单个响应可能属于多个对象，导致歧义。因此，跟踪器必须单独利用每个目标的外观，通过对每个目标使用分类器来提高其类内识别能力。此外，检测响应可能是稀疏的，从而导致跟踪误差。跟踪器由检测器发现的每个新条目初始化。利用两组样本，将内盒区域表征为正，将外盒区域表征为负，提取特征，用于训练朴素贝叶斯分类器。在跟踪过程中，根据预定义的学习率更新模型，以优化轨迹的构建。分类时，搜索窗口设置为覆盖目标周围区域。目标特异性自适应最小化目标中心变化，在检测器响应稀疏时引导跟踪器，并成功解决短期遮挡问题。当对象参与遮挡时，更新被挂起。然而，分类器的性能是基于其有限的搜索窗口大小，这通常适合于在连续帧上定位对象。当对象长时间被遮挡或者模板出现较大的外观差异时，就会暴露出分类器搜索方法的弱点，从而导致模板更新问题。

检测器的预测和分类器的预测，分别是 DP 和 CP，可以通过加权方案相互补偿误差。检测器的稀疏性及其无法处理短期遮挡和分类器的漂移问题可以通过使用一个权重函数将两个预测融合到一个最终的 FP 来克服:

$$FP_k^i = \boldsymbol{w} \cdot \begin{bmatrix} DP_k^i \\ CP_k^i \end{bmatrix}$$

3 相机内 RE-ID

在相关文献中，Re-ID 方法被认为是独立于跟踪的，并在已经剪切目标的数据集上进行测试，并伴随着 Ground Truth 注释，用于从背景中分割前景像素。然而，在自主跟踪框架中，需要一个预处理单元来构建图像库。此外，Re-ID 匹配依赖于目标的外观，因此需要通过对象分割来构造具有代表性的描述符。然而，大多数跟踪器提供对象周围的检测窗口，在这些窗口中，定位可能不在方框的中心，或者方框的大小可能不正确。

在窗口内分割目标的方法有很多。在半自动分割技术上已经有了令人印象深刻的发展，用户交互改进结果。我们提出了完全自动化的分割基于假设，尽管检测器的定位不准确，跟踪器的检测窗口的中心通常包含最相关的信息。因此，分割是有偏的，接受框中心的像素点，拒绝边界的像素点。

预处理单元由裁剪后的独立对象提供，并并行跟踪。分割由初始 trimap $T = T_F, T_{PF}, T_B$ 引导，标记前景 T_F，可能前景 T_{PF} 和背景像素 T_B 的区域(图 2)。为了效率，每个表示的镜头数量是有限的。为了平衡外观随时间的变化，每 n 帧镜头都被聚合。因此，每个轨迹都可以由大量裁剪和分割的图像表示,即一个多镜头表示。这个过程的结果如图 3 所示。

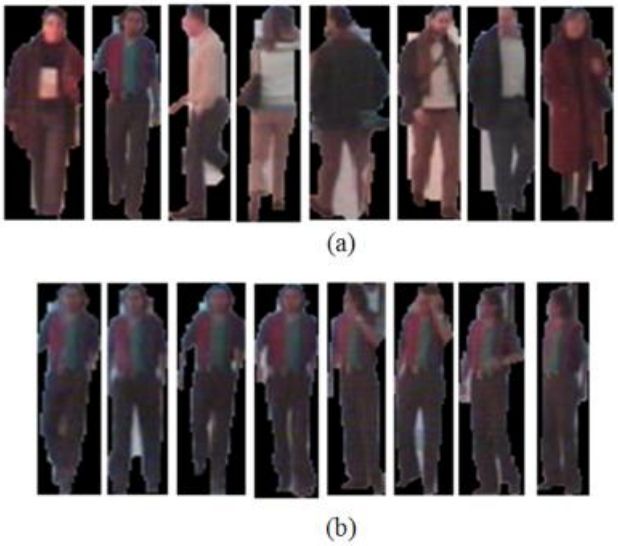

图 3　(a)从一帧中提取的无监督分割镜头。(b)每进入第 n 帧就累计目标镜头，构建一个多镜头表示。

Re-ID 方法将主体视为从一组图像中提取的局部和全局特征集。采用不对称驱动的身体分割方法，在假设行人具有双峰颜色分布(即上衣和裤子)的前提下，通过最大化人体上下 HSV 直方图的差异，将盒子分割为身体的各个部分。与直方图一起，编码纹理信息的最大稳定颜色区域被积累到表示中。从一帧或一串帧派生出一个或多个签名。签名匹配产生的排名结果中，第一个排名位置表示最佳的链接对。匹配基于外观相似度，用目标 A 与匹配的候选 B 之间的距离 d_{ReID} 表示:

$$d_{\mathrm{ReID}}(A,B) = d_{HSV}(A,B) + d_{MSCR}(A,B)$$

4　遮挡处理

被跟踪对象可以完全独立于其他对象，或者遮挡器隐藏其他对象或被遮挡器隐藏。在提出的框架中，提出了在每种情况下应遵循不同的跟踪策略。独立的对象很容易跟踪，因为没有明显的检测稀疏性。当对象相互混淆时就会出现问题。当它们的边界框重叠时，检测物体之间的遮挡(图 4)。

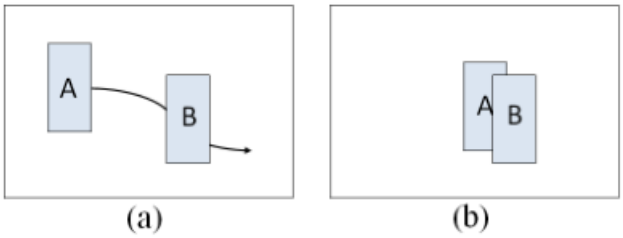

图 4　在(a)中，独立的物体 A 和 B 即将相遇。在(b)中，闭塞开始于 A 从 B 后面经过。

4.1　目标分类

为了对目标进行分类，对其外观完整性进行评估。在遮挡过程中，被遮挡的包围框的内容可能会错过有价值的外观信息，只有最前面的对象具有完整的视觉完整性。基于再识别技术的内在能力，根据其外观相似度对目标进行排序，将目标标记为遮挡物或遮挡物。

假设 K 个物体在第 n 帧参与遮挡。为了对第 n 帧的包围盒进行分类，每个盒子都与其多镜头表示进行比较。由于物体在第 n-1 帧之前是独立的，因此每次比较都会产生盒子在第 n 帧的单镜头和多镜头表现之间的相似度评分 d_{ReID}。总共进行 K 次比较，距离最小的目标被归类为遮挡器。其余 K-1 盒标记为闭塞。

在这一点上，基线跟踪器知道目标分类，现在可以为每个目标组使用不同的策略。对于遮挡器，融合有利于分类器，而对于隐藏目标，融合有利于检测器，因为模板可能不可见。每个目标类别的不同权重可以在表 1 中看到。

Labels/Weights	w_d	w_c
Independent	0.5	0.5
Occluder	0.2	0.8
Occluded	0.8	0.2

表 1　基于目标状态的跟踪器融合权重。

4.2　身份验证

当一个对象获得多个 ID 号时，常见的跟踪错误就会发生。在某些情况下，长期被遮挡的目标丢失，并可能重新出现在分类器的搜索窗口

之外，从而导致初始化一个新目标。由于被遮挡的边界框的持久性，它们的原始边界框不能被终止。跟踪器只在没有检测响应的情况下终止独立的轨迹。为了防止 ID 切换错误和遮挡盒漂移，触发 Re-ID 机制来检查遮挡附近新发现的目标的原创性。

在封闭区域 S 周围，跟踪器对新条目的任何尝试都必须与所有被划分为封闭的候选条目进行比较。比较距离的最小值($d_{minReID}$)表示关联的最佳配对。由于结果是排序的，因此使用全局阈值 th_{ReID} 来确保最小距离在可接受的范围内。

Re-ID 方法使用距离指标进行两两评分，导致效率和简单。然而，由于在特征提取阶段缺乏机器学习算法，并且将关联视为一个排序问题，因此不存在绝对置信度测度。这导致不可避免地使用阈值，阈值是启发式定义的，以便最大化真实匹配。在我们的实验中，所有情况下都使用了 th_{ReID}=0.7 的阈值。为了定义这个阈值，建立了正确和错误匹配距离的双峰分布。通过拟合距离数据上的两个高斯分布并识别它们的交集，设置阈值。

5 实验结果

Tracker	Type	GT	MT	PT	ML
Li et al. [10]	Offline	143	84.6%	14.0%	1.4%
Bak et al. [19]	S.window	140	84.6%	9.5%	5.9%
Baseline tracker	Online	138	85.5%	12.3%	2.2%
Proposed method	Online	138	86.3%	10.1%	3.6%

表 2　不同跟踪结果在 CAVIAR 数据集上的比较。

Tracker	Type	GT	MT	PT	ML
Zhang et al. [12]	Online	19	78.9%	15.8%	5.3%
Badie et al. [20]	Offline	12	50.0%	33.3%	16.0%
Baseline tracker	Online	19	73.7%	26.3%	0%
Proposed method	Online	19	78.9%	21.1%	0%

表 3　对 PETS 2009 S2L1 View 01 序列不同跟踪结果的比较。

GT:地面真实轨迹的数量。

MT:成功跟踪的轨迹占轨迹总长度的 80%以上。

PT:轨迹跟踪占轨迹总长度的 20%-80%。

ML:轨迹跟踪长度小于轨迹总长度 20%的百分比。

跟踪评价结果如表 2、3 所示。对于 Caviar 数据集，比较结果表明，该框架实现了最大限度的 MT。与离线方法相比，我们的系统实现了最多的 MT，但更多的 ML 轨迹。离线跟踪器考虑所有的检测响应给定的帧，可以更好地处理检测稀疏性比在线方法，以一种因果方式决定。这些结果表明，基准跟踪器产生了比较的结果，考虑到最先进的方法，而我们的附加机制增强和改善了拟议框架的整体性能。图 5 描述了输出帧的示例序列。

图 5　　被遮挡的对象用较薄的边框表示

三、　多目标多相机跟踪的状态感知再识别特征

多目标多摄像头跟踪(MTMCT)通过多个摄像头拍摄的视频跟踪许多人。人员重新识别(Re-ID)从与人员查询图像相似的人的图库图像检索。

人员重新识别(Re-ID)与 MTMCT 密切相关:给定人员的快照(查询)，Re-ID 系统从数据库中检索其他人物的快照列表，这些快照通常来自不同的相机和不同的时间，并通过降低与查询的相似性来对它们进行排序。这样做的目的是，数据库中与查询中的人相同(即描述同一个人)的任何快照的排名都很高。

MTMCT 和 Re-ID 有细微但本质上的区别，因为 Re-ID 对查询的距离进行排序，而 MTMCT 将一对图像分类为同一性或非同一性，因此它们的性能由不同的度量标准来衡量:Re-ID 的排序性能，MTMCT 的分类错误率。这种差异似乎表明，用于两个问题的外观特征必须通过不同的损失函数学习。理想情况下，Re-ID 丢失应该确保对于任何查询，目标与它相同的特征之间的最大距离小于与它不相同的特征之间的最小距离。这将保证任何给定查询的正确功能排名。相比之下，MTMCT 损失应该确保任意两个同一性特征之间的最大距离小于任意两个非同一性特征之间的最小距离，以保证身份内和身份间距离的边际。

多目标多摄像头跟踪的目的是从一组摄像头捕获的视频中提取轨迹。近年来，

随着 Re-ID 模型的引入，MTMCT 的跟踪性能得到了显著提高。然而，由于目标的遮挡和方向变化，图像的外观特征往往变得不可靠。在 MTMCT 中直接应用 Re-ID 模型会遇到由于遮挡而导致的 IDS(identity switches)和 tracklet 分片的问题。为了解决这些问题，本文提出了一种新的跟踪框架。该框架将遮挡状态和方向信息应用于 Re-ID 模型，并考虑人体姿态信息。此外，利用融合跟踪特征的轨迹关联处理碎片问题。提出的跟踪器在多摄像头硬序列上达到 81.3%的 IDF1，在很大程度上优于所有其他参考方法。

1 介绍

多目标多摄像头跟踪(MTMCT)是计算机视觉中的一个重要问题，在公共安全领域尤为重要。不同于单摄像头的多目标跟踪，MTMCT 的目标是跨多摄像头的多目标跟踪。摄像机网络具有比单一摄像机更广阔的视野和更广阔的应用前景。然而，MTMCT 除了面临与 MOT 相同的遮挡、姿态方差和背景杂波的挑战外，还面临一些具体的挑战，如相机之间的盲区、视点的变化和照明方差。

特征表示、遮挡处理和推理是 MOT 和 MTMCT 的关键组成部分。外观特征对于保持被跟踪目标的一致性具有重要意义，其中，颜色直方图和 HOG 在以往的工作中得到了很好的研究和应用。但是，颜色直方图和 HOG 对遮挡的鲁棒性不强，不能很好地处理外观方差。近年来，Re-ID 模型作为一种具有鉴别性的外观描述符被广泛采用。此外，人的 Re-ID 与 MTMCT 密切相关，高质量的 Re-ID 特征往往会带来高跟踪性能。然而，Re-ID 训练数据通常是人工标记的，高遮挡的样本总是从训练数据中被丢弃。因此，在拥挤的场景中，直接使用低质量检测器的 Re-ID 特性总是会导致较差的性能。

遮挡可能是 MOT 中最关键的挑战。这是导致 ID 切换或轨迹碎片化的主要原因。直接从目标高度遮挡的检测区域提取特征是不合理的。因此，遮挡感知是特征提取的关键。如果得到遮挡状态，则只保留稳定特征，丢弃被遮挡特征。此外，方位对目标外观有显著影响，这是大多数 ReID 模型所忽略的。

在 Re-ID 任务的训练过程中，一个身份所包含的实例数量是有限的，但跟踪场景中轨迹的长度是不受限制的。外观的变化主要是由于背景的变化，姿态的变化，方向和视点的变化。现有的大多数 ReID 模型不能处理这些问题。因此，跟踪需要对 Re-ID 特征进行后处理。在线跟踪器通过帧间的关联来构建轨迹，通常

只考虑轨迹与检测结果之间的关系。但是，被遮挡目标的检测结果往往不准确，在线跟踪器在这种情况下可能会产生很多碎片化的轨迹。与在线跟踪器不同，有些离线跟踪器首先生成短轨迹，然后链接轨迹以获得最终轨迹。此外，脱机跟踪器通常可以获得更好的性能，因为它们可以提前获得整个序列，并且在关联时 tracklet 包含比检测更多的信息。在本工作中，采用轨迹关联的方法来处理轨迹片段。

本文提出了一种状态感知的 Re-ID 特征，该特征着重于具有额外人体姿态信息的外观表示。具体地说，为了更好地利用 Re-ID 特征，利用人体姿态信息来估计目标状态，包括遮挡状态和方向。为了在跟踪过程中实现稳定、准确的关联，设计了融合跟踪特征作为轨迹的外观表示。提出了一种融合跟踪特征的距离矩阵用于数据关联。为了处理轨迹碎片，提出了轨迹关联方法，包括轨迹纠偏和轨迹聚类。最后通过实验验证了该框架的有效性。

2 MTMCT 总体设计

本节介绍 MTMCT 的总体设计。该跟踪框架由单摄像头跟踪(SCT)和多摄像头跟踪(MCT)两部分组成。本工作中，利用 SCT 跟踪器在单个摄像机中生成轨迹。然后对相机内轨迹进行聚类，最终得到多相机间的轨迹。

2.1 状态估计

利用人体姿态信息估计遮挡状态和方向。遮挡状态和方向的推断详细如下。

$$N_{valid} = \sum_{i=1}^{N_k} \mathbb{1}\{c_i > \gamma_{\text{valid}}\}$$

当N_{valid}大于数阈值(θ valid)时，即大多数关键点可见且目标未被遮挡，则认为 Re-ID 特征有效，否则认为 Re-ID 特征无效。

图 1　　4 个方向:(1)前，(2)后，(3)左，(4)右。蓝色、绿色、红色、黄色四个点分别代表左肩、右肩、左臀、右臀四个关键点。

方位估计是导致同一目标出现外观不一致的重要原因。如图 1 所示，身体关键点集 $K_{body} = \{k_{ls}, k_{rs}, k_{lh}, k_{rh}\}$对应左肩，右肩，左臀，右臀，耳朵关键点集$\{k_{le}, k_{re}\}$对应左耳，右耳。在本工作中，方向分为四种状态 $O=\{o_{left}, o_{right}, o_{front}, o_{back}\}$，方向由深度神经网络(DNNs)推断，其架构如表 1 所示。具体来说，K_{body} 的位置和置信度，耳朵置信度被输入到 DNN 中进行分类任务，因此输入维度为 14。

Name	Input size	Output size
FC1	$4 \times 3 + 2$	128
FC2	128	64
FC3	64	128
FC4	128	64
FC5	64	4

表 1　　定向分类网络的体系结构。采用简单的深度神经网络将输入分为四个方向。使用 5 个 FC(fully connected)层。

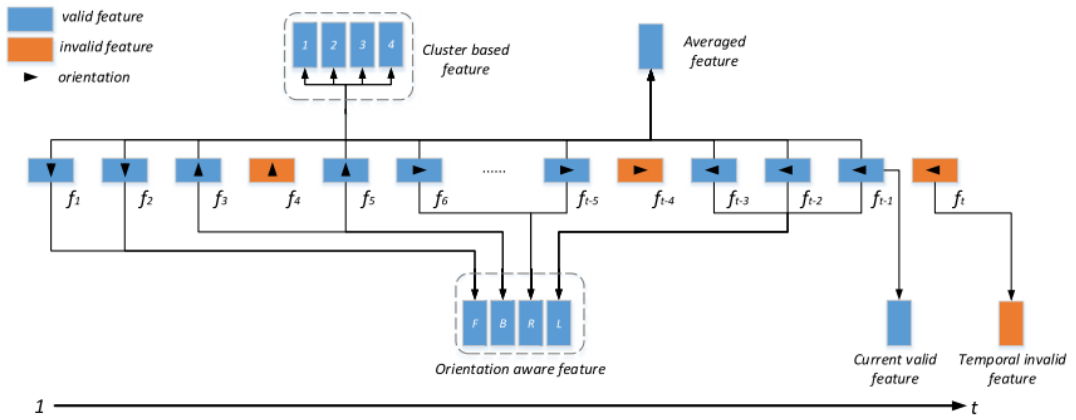

图 2　　融合跟踪特征包括当前有效特征、时间无效特征、方向感知特征和平均特征。已保存的 tracklet 历史 Re-ID 特征显示为 f1,f2,...,ft−1,ft，其中 fi 是来自第 i 帧中匹配检测的 Re-ID 特征。fi 中说明了四种类型的方向，包括前（F）、后（B）、左（L）和右（R）。基于集群的特征显示为 Nc = 4。

2.2 融合跟踪功能

由于轨迹的遮挡和方向的变化，很难对随着轨迹增长的目标外观进行建模。另一方面，Re-ID 模型作为一种高级外观描述符被广泛采用。然而，大多数方法通常以一种简单的方式使用 Re-ID 特性，比如平均这些特性。如图 2 所示，我们采用设计良好的融合跟踪特征 F_{track}，对保存的轨迹的历史 Re-ID 特征进行多种不同的组合，可以更可靠地表示目标的外观。

tracklet 的融合跟踪特征 Ftrack 由 Ftrack={fcurrent, forientation, fcluster, finvalid, favg}五类特征组成，如下所示。

当前有效特性：在某些情况下，目标快速移动，因此它们的规模和姿态变化迅速。我们利用目标历史外观中最新的有效特征使外观模型包含最新的信息。

方向感知特征：方向感知特征由四种不同方向的平均特征组成:方向 $=\{f_{left}, f_{right}, f_{front}, f_{back}\}$。具体地说，方位特征是具有相同方位的所有历史有效特征的平均值。在数据关联中，选择与检测方向相同的方向上的一个元素来计算外观距离。将两个轨迹间的方向距离定义为对应特征在同一方向上的距离的最小值。

基于集群的特性：特征聚类被广泛应用于非监督和半监督 Re-ID 中。本文在基于聚类的特征聚类上采用了一种在线聚类算法，该聚类具有与高斯混合模型相似的初始化和更新策略。我们设置 N_c 为 $f_{cluster}$ 中的集群数量上限，初始化 $f_{cluster}=\varnothing$。

通过计算 Nc×Nc 距离矩阵 $M_{cluster}$ 得到两个轨迹之间的距离 $d_{cluster}$。其中第 i 行第 j 列的值为第 i 个簇中心到第 j 个簇中心到两个轨迹之间的距离。$M_{cluster}$ 中的最小值被选择为 $d_{cluster}$。

时间无效的特征：当 tracklet 与无效特征匹配检测时，由于无效特征的不可靠性，上述三种特征不更新。但是，如果外观特性没有及时更新，就会发生 IDS。因此，采用时间无效特征 $f_{invalid}$ 对无效特征进行更新，使轨迹更加平滑。值得注意的是，$f_{invalid}$ 只保留上一帧的无效特性，如果过期将从 F_{track} 中删除。

平均特性：平均特征 f_{avg} 是对 tracklet 所有有效的 Re-ID 特征的平均特征。

2.3 单相机跟踪

2.3.1 跟踪阶段

为了建模 SCT 跟踪器中轨迹的生命周期，我们定义了四个阶段，暂定、确认、不可见和消失，如图 3 所示。

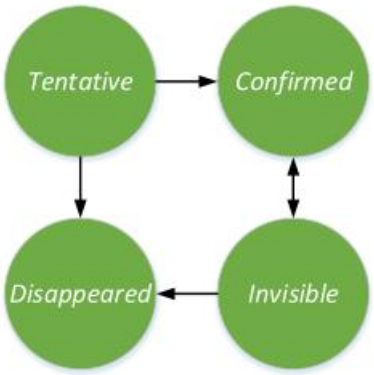

图 3　跟踪轨迹的相位和变换。试探性、确认性、隐形性和消失性是其周期的四个阶段

使用不匹配检测生成新的轨迹，并根据遮挡状态初始化为不同的相位。如果检测被高度遮挡，则跟踪阶段将被初始化为暂定。否则，它将被初始化为 Confirmed。如果在 Confirmed 阶段的 tracklet 错过了 μ_m 次，它将进入 Invisible 阶段。并且如果 Invisible 阶段的 tracklet 错过了 μ_d 次，它将切换到 Disappeared 阶段。如果 tracklet 在数据关联中匹配，则 Invisible 阶段的 Tracklet 将返回 Confirmed 阶段。此外，Tentative 阶段的 tracklet 如果错过一帧，将转为 Disappeared 阶段，如果匹配到有效特征的检测，则转为 Confirmed 阶段。通过这种方式，可以去除误报检测。另一方面，Disappeared 阶段的 tracklet 意味着目标已经消失或已经离开场景，因此将 tracklet 从 tracklet 集中移除。

2.3.2 sct 总体框架

在我们的SCT框架中，由于对象检测的发展，采用了检测跟踪策略。提出的跟踪方法采用在线的方式生成轨迹。具体来说，从头到尾维护一个轨迹集，对每 K 帧轨迹聚类后生成跟踪结果。本文的 SCT 框架可以分为 tracklets 链接和 tracklet 关联两部分，如图 4(1)和(2)所示。(1)显示了在线生成和更新 tracklet 的管道。(2)对轨迹进行后处理，包括轨迹纠偏和轨迹聚类。通过这种方式，可以处理 tracklet 片段。

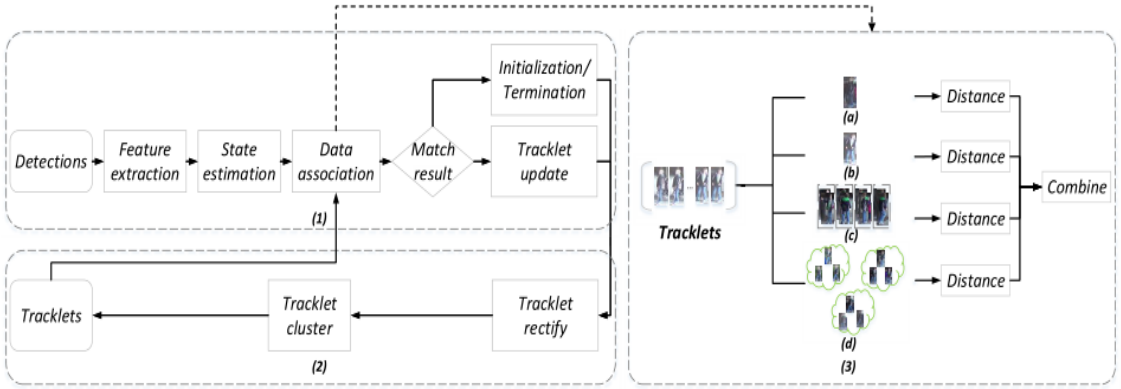

图 4　提出了 SCT 框架。整体跟踪流程如(1)、(2)所示。其中在线生成轨迹的流程如(1)所示，对已有轨迹的后处理如(2)所示。(3)详细计算了已有轨迹与检测之间的距离矩阵。(a)、(b)、(c)、(d)分别为时间无效特征、当前有效特征、方向感知特征和基于聚类的特征。计算来自检测的 Re-ID 特征与来自轨迹的融合跟踪特征之间的距离。然后结合四种特征之间的距离，得到最终的轨迹到检测的距离。

2.4多摄像头跟踪

本文采用距离矩阵 M_{mct} 和贪心算法实现多摄像头跟踪。

首先，我们收集所有摄像机的轨迹并计算 M_{mct}。构造 M_{mct} 后，采用贪心算法进行轨迹关联，直到矩阵的最小距离超过 θ mct。

具体来说，当一个轨迹与其他轨迹相关联时，M_{mct} 中相应的行和列会被更新。另外，假设本文的 SCT 跟踪器足够好，我们只关联不同摄像机的轨迹，因此摄像机内轨迹保持不变。

2.5实现细节

采用 ResNet-34 提取 Re-ID 特征。输入大小为 128×256，从最后一层全连接层提取 128d 的 Re-ID 特征。在训练过程中使用了公开的 Re-ID 数据集，包括 Market-1501、CUHK03、MSMT17、PRW、DukeMTMC-ReID 和额外的私有数据集。Re-ID 模型在 DukeMTMC-ReID 上实现了 78.5 的精度

姿势估计量：本文采用 Alphapos 对人体姿态进行估计。此外，在 DukeMTMCT 数据集上，姿态估计器没有进行微调。

参数设置：参数设置对目标的生命周期进行建模时，μ_m 设为 10，

μ_d设为 300。在状态估计中，γ 有效设为 0.3，θ 有效设为 7。在磁迹关联中，θ 整流 y 和 θ 团簇分别设为 20 和 30。在多相机跟踪中，θ mct 设为 40。

3 实验

在本节中，我们对提出的状态感知 MTMCT 框架进行了实验。验证了工作的有效性，并考察了不同成分的贡献。最后，提交测试集上的结果，并在此基准上与其他最新的跟踪框架进行比较。

3.1 消融研究

消融研究是在训练序列的 SCT 任务上进行的。考虑了方向感知特征、聚类特征和时间有效特征。基线跟踪器仅使用当前有效的特征流进行数据关联，产生跟踪结果的时间间隔 K 设为 10 秒。消融研究结果见表 2。

Method	IDF1↑	MOTA↑	IDS↓
Baseline	77.1	82.8	6409
Baseline + $f_{cluster}$	82.5	82.6	8170
Baseline + $f_{cluster}$ + $f_{orientation}$	85.1	82.8	6564
Baseline + $f_{cluster}$ + $f_{orientation}$ + $f_{invalid}$	85.2	82.8	5466

表 2　消融研究证明了状态感知的 Re-ID 特性的稳步改进。表中显示了 IDF1、MOTA 和 IDS，箭头表示较低或较高的最优度量值。

对比第二行跟踪器和基线，基于集群的特征是必不可少的，它可以提高 IDF1 上 5.4%的性能。可以发现，现有的有效特征不能正确地对目标外观进行建模，基于聚类的特征是表示目标外观的一种有效方法。$f_{cluster}$ 随着轨迹的增长而变得更加稳定，但在一开始对遮挡的鲁棒性不强，这是导致 IDS 增加的主要原因。比第三行和第二行跟踪器，具有方向感知特征的跟踪器性能更好，比 IDF1 提高了 2.6%。我们可以发现，方向特征与聚类特征是互补的。对比第 4 行跟踪器和第 3 行跟踪器，IDF1 改进了 0.1%，IDS 从 6564 减少到 5466，这意味着无效的时间特征有效地减少了 IDS，使轨迹更加平滑。

3.2 与其他先进方法比较

我们在 DukeMTMCT 数据集上将本文提出的跟踪器与其他跟踪方法比较，结果如表 3。

Tracker	*test-easy single*			*test-easy multiple*			*test-hard single*			*test-hard multiple*		
	IDF1	IDP	IDR	IDF1	IDP	IDR	IDF1	IDP	IDR	IDF1	IDP	IDR
BIPCC[35]	70.1	83.6	60.4	56.2	67.0	48.4	64.5	81.2	53.5	47.3	59.6	39.2
MYTRACKER[54]	80.3	87.3	74.4	65.4	71.1	60.6	63.5	73.9	55.6	50.1	58.3	43.9
TAREIDMTMC[22]	83.8	87.6	80.4	68.8	71.8	66.0	77.9	86.6	70.7	61.2	68.0	55.5
DeepCC[36]	89.2	91.7	86.7	82.0	84.3	79.8	79.0	87.4	72.0	68.5	75.8	62.4
MTMC_ReID[57]	89.8	92.0	87.7	83.2	85.2	81.2	81.2	89.4	74.5	74.0	81.4	67.8
MTMC_basel†	91.3	91.8	**90.9**	**87.4**	87.8	**87.0**	83.7	88.8	79.1	75.4	80.0	71.3
Ours	**91.8**	**93.3**	90.3	86.8	**88.2**	85.4	**85.8**	**93.6**	**79.2**	**81.3**	**88.7**	**75.1**

表 3 本文提出的跟踪器与其他跟踪方法比较

如表 3 所示，本文提出跟踪器在 DukeMTMCT 数据集上实现了最先进的性能。由于设计良好的状态感知 Re-ID 特性，我们在测试难度上大大超过了其他所有方法，这进一步验证了所提出的跟踪器在如此拥挤的场景中的鲁棒性。