问题：

1.P1原文提到

一、问题分析

1背景：在摄像机集群中的实时视频分析具有海量的数据和巨大的资源耗费，由于需要找到不仅跨帧，还有跨相机的关联性，比无状态的单相机视频分析在计算和网络上有更高的要求。最近的视频分析系统大多着重于优化单视频分析的成本、精度权衡，而他们在单个流上做的优化却是相互独立的。这导致跨摄像机视频分析的成本会随着整体相机数量和目标的存在时间而增加。

2 解决思路：利用相机间的**时空相关性**来降低网络和计算成本。之前也有利用时间、空间相关性的系统（如变色龙），但其主要还是降低单视频分析的成本，并没有对视频本身基于跨相机关系进行任何优化，每个视频流的分析是独立于其他视频的。

突破点：跨摄像机应用程序的一个关键特征是，在任何给定的时间内，感兴趣的对象只出现在摄像机的一小部分中。

问题：变色龙中利用的空间相关性，让相似的相机共享相似的配置，难道不能算一种跨相机的资源优化，一种联系吗？

3.挑战：

①在未标记视频数据上，自动获取时空相关性是非常昂贵的。

②为最大化收益，需要具有必要系统支持的清晰抽象。

③任何时空分析都必然存在误差，这些误差将导致物体丢失，必须进行有效的修正。

4.应对：提出Spatula的工作步骤（分为3个阶段）

1)在离线剖析阶段，从无标签视频数据构建跨摄像机时空关联模型，对历史交通模式中观察到的位置进行编码。这是一个昂贵的一次性操作，需要用脱机跟踪器检测实体，然后将它们聚合到跨相机相关性的配置文件中。

2)在推理时，Spatula使用这个时空模型来过滤掉与查询身份的当前位置(相机)不相关的相机，因此不太可能包含它的下一个实例。

3)偶尔，这种过滤会导致Spatula错过查询检测。在这些情况下，Spatula对最近过滤的帧(它存储的帧)执行快速重放搜索，发现错过的查询实例，并优雅地恢复到实时搜索中。

5.实验说明时空相关性：DukeMTMC数据集

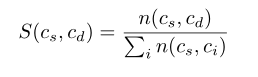
二、系统设计

**时空模型**描述了相机之间的时空相关性，可通过应用程序进行查询。在高级别上,与两个摄像头,一个可以查询模型cs和cd,和一个时间窗口,它将返回对象离开cs怎么可能出现在cd(即空间相关性),如果它出现在cd的时间内可能会出现窗口(例如,时间相关)。

**前向和回放分析**对直播视频进行实时推理(即转发)，对历史视频进行推理(即回放)。这两种能力都是联合操作的，重放搜索是时空剪枝的固有需求:由于较弱的空间/时间相关性而忽略摄像机将不可避免地引入假负，而这是搜索所有摄像机的基线已经避免了的，所以Spatula提供了重放搜索的抽象，允许对一些历史视频(被忽略的)进行**比实时更快**的搜索，以进行纠错。

问题：在S, T的定义中，总人数是指查询总持续时间内的总人数吗？时间窗口t1,t2是怎么确定的呢？

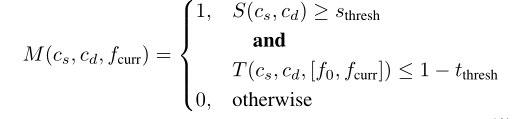
空间相关性：



时间相关性：



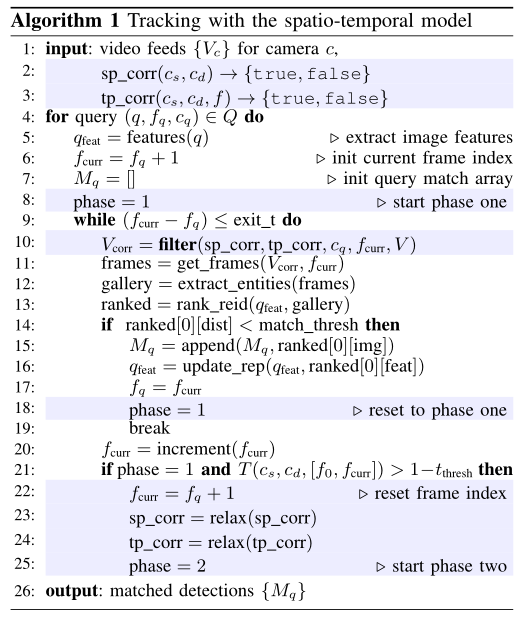
时空相关性模型：



为什么这里要反着定义t(th)？不能是T>=t(th)=98%吗？

注意：这里Sth和tth阈值的设置是似然阈值，从根源上排除了这些异常的相机和时间窗口，因此可能存在对异常情况的遗漏。（如，物体偏偏就出现在了异常的帧中）

三、目标追踪算法的分析



问题：这个算法形式上是写在一起的，但实际上其中不同部分是由不同的对象执行的？

目标：flag all subsequent frames, on all cameras, where q appears.

首先明确，这里的phase=1标志目前系统正工作在第二阶段：利用时空模型过滤视频流，在缩小的搜索空间上进行搜索；phsase=2标志目前系统正工作在第三阶段：没有找到目标，进行快速的重放搜索。Phase就是一个标志位，表示当前正在进行的搜索循环是正常搜索还是重放搜索。

1~3 输入：相机集群中各相机对应的视频流，判定相机间时空相关性的两个函数（函数式编程）

问题：这里的输入是一个相机的多个视频流，还是全部相机的各自单个视频流？

自答：可能是一个相机的多个视频流，因为后文说了物品识别模块是运行在相机上的。

4~8 对于一次查询请求中的目标q（首次出现在相机cq的第fq帧），提取其识别特征，帧指针移到下一帧（因为当前在当前帧q的唯一准确位置已经是已知的初值了，必然下一时刻应该在接下来的帧中寻找q），初始化检测结果数组，开始阶段一的算法。

如果fq未知，则首先运行时空相关性模型找出来。

问题：Pc,w是如何计算出来的？

9 进入搜索循环。exit\_t是一个人为预先设定好的时间阈值。当当前帧与起始帧间的时间差大于这个阈值时，我们就认为目标已经离开了搜索范围，于是停止搜索。否则，就在阈值内按一定增量逐帧进行目标检测。

10 利用时空相关性过滤得到一个大大缩小的搜索空间Vcorr，里面包含的全部是与相机cq的第fq帧时空相关的视频源。

11 从这些视频源中获取其当前时间的帧

12 从这些帧中尝试提取检测的目标

13 利用原始目标的特征对这些检测结果进行排序。注意，ranked是一个二维数组，其第一维是按照相似度排序的序号，第二维有以下分量：dist相似度，img检测目标图像，feat该目标图像的特征。

14~18 选择与原始目标最相似的结果，即ranked[0]，若其间的差异位于人为规定的阈值内，则认为是我们需要的目标，并加入结果数组中，更新特征，帧指针。同时由于没有进行重放搜索，故当前仍处于phrase1。

问题：目标是标记所有相机上所有出现了检测对象的后续帧，那19行的break是什么意思？如果break了，岂不是意思就是对于当前请求，只标记了出现检测对象的后续第一帧，算法就结束了？我认为算法有问题，应该改为：19行与20行互换位置，然后把break改为continue。这样一来，在标记某个相机的后续一帧后，会继续往后标记剩下的后续帧，直到超出阈值，认为目标已经逃离，这样才符合作者的描述。

21~25 如果在第一阶段，没有找到符合精度阈值要求的目标，且按照计算，当前帧与原始相机原始帧的时间相关性达到要求，在当前帧理应出现目标，则说明算法可能出现了误差，在利用时空相关性剪枝时减去了可能出现目标的相机源和时间窗口。故开始进行重放搜索，先将帧指针回拨至初始帧，放宽时空相关性（即增大搜索空间），重新开始搜索循环。同时将phrase标志位设为2，表示当前位于重放搜索工作状态。

由于此时进行回放搜索，表明相对于请求我们已经落后了，产生了一定延迟，故加快搜索速度十分必要。Spatula采取的加速方法是：1.适当调整帧率和分辨率等旋钮。这里可以利用之前论文的方法。如VideoStorm等；2.利用空闲的机器空闲的资源并行处理该工作。

26 算法结束，返回所有包含目标的后续帧。（这些帧可能位于不同的时空相关相机上）

四、局限性

1. 时空模型具有似然阈值，这使得它容易丢失“异常值”，即罕见的空间运动模式或运动速度的时间异常。虽然重放搜索检测到这样的实例，并在很大程度上能够纠正他们，但它不是完美的。确保不遗漏异常值是未来工作的一部分。

问题：什么叫似然阈值？

2.时空相关性可能随着时间的变化而变化(例如，道路施工可能会阻塞繁忙路段，这可能会降低两个相机之间的相关性)。Spatula可以根据追踪那些在正常搜索中失踪而在重放搜索中被发现的目标的数量来决定是否触发时空模型的重新分析。

五、系统实现

Spatula系统中，集群中的相机各自运行着视频分析模块，而一个笔记本运行着控制器，控制器负责生成和维护相机间的时空相关性模型。控制器和各相机之间只有控制信息的交换，没有任何视频数据的交换。

问题：这样高度解耦，可以避免视频流在网络中的传输，节省了很多网络资源。但这需要相机具有足够的资源和能力运行视频分析模块。但现在大部分的相机还不具备这种能力，是否可以考虑结合Reducto的工作？

控制信息包括：控制器向相机发送触发请求，让相机开始在其流中检测目标对象，相机会在检测完成后发送反馈信息，如是否检测到。

六、评估

We report the following four metrics which are computed over the entire query set. (i) Compute cost – Number of video frames processed, aggregated over all queries {qi}. (ii) Network cost – Average network bandwidth usage of transmitting encoded videos required by search algorithms. (iii) Recall (%) – Ratio of query instances retrieved to all query instances in dataset, qj i . (iv) Precision (%) – Ratio of query instances retrieved to all retrieved instances, rj i . (v) Delay (sec.) – Lag between position of tracker and current video frame, in seconds, at the end of a tracking query.

问题：1.为什么分开考虑召回率和精确率而不是考虑使用F1分数作为评估标准？用F1分数是不是更好。下面在考虑带宽预算的情况下与基线比较，都使用了F1分数

2.网络成本中，视频是在什么之间传送的？这里的搜索算法指的是上面的Tracking算法吗？

3.既然Spatula的最佳阈值取决于数据集，即视频内容，是否有必要动态地根据视频内容改变其阈值设置，选择当前最佳的阈值？

4.跳帧与Spatula利用时空相关性降低计算成本是正交的，是否意味着我们可以将之前的研究成果与这篇里的相结合，进一步降低计算成本？