

指导教师： 杨涛

提交时间： 2015.3.28

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 01

姓名： 范国龙

学号： 2012302419

班号： 10011202



通过约束的序列标记的多目标跟踪

陈胜, 阿兰·弗恩和米哈伊托多罗维奇

{chenshen, afern, [sinisa](mailto:sinisa@eecs.oregonstate.edu)}@eecs.oregonstate.edu

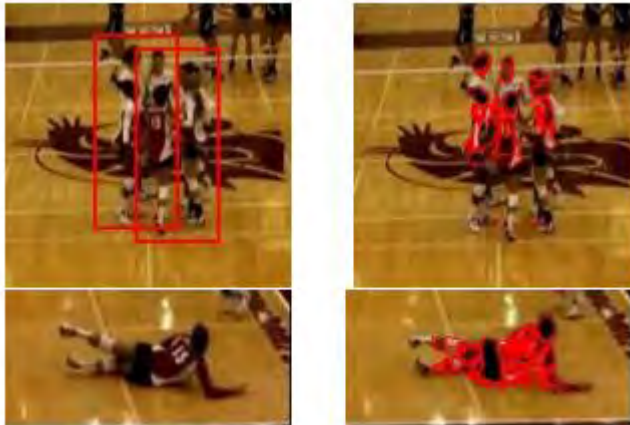
摘要

本文提出了一种新的方法来跟踪在拥挤的场景中的人，那里的人都受到长期的（部分）闭塞，并可能承担不同的姿势和关节。在视频中，检测跟踪器出现不好的表现是因为被检测的人的出现是不可靠的，并且普遍的假设是局部光滑轨迹并不存在。相反，我们使用时间的中级特征（例如，**supervoxels** 或密集的点的轨迹）作为处理阻塞更连贯的时空基础和姿态的变化。因此，我们制定了跟踪标记中层特征的对象标识符，并为执行此标记指定一个新的的方法，称为约束的序列标签（**CSL**）。**CSL** 使用成本函数的顺序指定标签的同时，尊重的含义硬约束计算通过约束传播。这种方法的一个重要特征是，它允许使用灵活的成本函数和约束捕获复杂依赖，不能代表标准网络流量公式。

1.简介:

本文提出了一种新的方法来跟踪多个在现实世界视频中互动的人，被检测的人的出现是不确定的。例如行人在拥挤的场景中的视频，和团队运动，比如篮球和排球。在这些视频中，人们都受到长期的部分或全部遮挡。此外，在体育视频中，玩家可以呈现各种姿势和身体的关节，他们的运动和外观（同队制服）通常类似于那些附近的玩家。这

些挑战使国家的最先进的人探测器不可靠。如图所示。



{图 1: 包围盒的 DPM 检测器[7] (左), 和[20] (右) supervoxels。DPM 往往错过检测玩家在人群中 (顶部) 和意想不到的姿势 (下)。supervoxels 更适合我们的具有挑战性的视频, 因为他们的边界通常是对齐的部分边界促进自下而上的推理。}

我们的实验证明, 流行的 DPM 探测器的响应[7]很嘈杂, 导致国家的基于检测的数据关联的最先进的追踪器性能差。这引起的问题是视频功能可以适用于我们设置接地数据关联。

我们采取时间中等特征 - 即 supervoxels[20], 而密点轨迹[5] - 作为一个更为连贯时空的基础处理闭塞, 姿势的变化, 和非光滑轨迹。例如, supervoxels 的边界通常配合人的轮廓。因此, 跟踪中级特点的正确组合有利于在自下而上的推理闭塞和不同的姿势。如图 1 所示, 这能够帮助消除两个人的部分遮挡的地方。这些中级的功能, 然而, 对常见的网络流量和数据关联使用对象检测相关制剂 (例如, [14, 13, 2, 19, 21, 17]) 提出了根本的挑战。特别是, 一个人要跟踪

通常是由未知数目的中层功能体现，在时间和空间上的分割和合并的。因此，边容量不再能均匀和目标的设置不明确。后处理也需要从流恢复身份。

处理上述问题的方法之一就是采用启发式的简化，直到产生问题符合标准的网络流量制定和采用高效的求解器解决问题。例如，我们可以根据中层特征尺寸假设合并和拆分的地方。然而，信息的丢失和错误运用简化注射会导致无法弥补的精度损失，否定了利用高效的“最优”的解算器。另外，约束和亲和力之间中级的功能的类型，可以用这样的马氏流网络用表示是相当限制性的。我们用这种方法的最初的努力尚未成功。

在这项工作中，既不是试探性地简化中层标签问题的一个有效解的框架，也不是我们开发了一个贪婪的标记的方法，可以直接对原始的（原版）问题操作。特别是，我们的方法通过依次标记中级视频特征与对象标识符来进行跟踪，在硬约束下。我们的新的算法，称为约束的序列标签（CSL），使用一个灵活的成本函数来顺序分配标签当遵循硬约束的影响时（例如，一个人不可能在不同的位置）。这种方法的关键优势是它在成本函数的形式上限制更少，允许它捕捉中层特征间高阶的依赖关系。我们描述了获取限制和成本函数的监督学习算法。

重要的是，虽然我们的方法本质上是贪婪的，但它配备了强大的理论保证。我们证明，如果有一个成本函数（在所考虑的成本函数空间中），支持精确的标记给定的约束条件，然后学习算法实现有限时

间收敛到这样一个成本函数，其中收敛速度提高了约束的“力量”。我们的实验证明这些假设在具有挑战性的现实问题中是满足的，具有我们能够学习高质量的约束和成本的功能，支持精确的 CSL。特别是，我们目前实验使用的运动和步行的数据表明国家的最先进技术有显著改善。

2.准备工作:

现有的多目标跟踪通常使用点，tracklets，或对象的检测[21, 2, 19] 17,12。前一部分强调了挑战的对象检测具有显著的闭塞和姿态变化。同样地，兴趣点是不适合设置的特性，因为它们不捕捉对象空间的凝聚力，因此需要使启发式的假定对象的运动，大小和形状来解决闭塞[18, 23]。

对于数据关联，视频功能之间的依赖关系，通常由一个网络捕捉，并且通常简化为在第一或第二阶马尔可夫依赖性[14, 4]。一些方法试图编码硬约束，从而促进模棱两可的情况下的数据关联。然而，视频功能之间的亲缘关系是典型的结合硬约束成一个 NP 硬的优化问题[3,15]。这，相反的，需要一个硬约束到连续域的松弛，这往往会导致无意义的结果（例如，两个不同的对象的检测可被分配了相同的对象 ID），并要求启发式后处理。相反，我们的 CSL 办法直接利用这样的硬约束来约束传播的，而我们的实验证明是有效的。

CSL 是最近成功由顺序的分类方法来解决结构预测问题激励的（例如[10]），其中分类器被用于顺序标签结构化对象。然而，CSL 通过纳入硬约束来克服潜力短视。

3.约束的序列标记:

我们的输入跟踪是一组 N 个中级，时空特征（例如 supervoxels） $\{S_1, S_2, \dots, S_n\}$ 捕捉前景对象在视频中的活动（前台或显着性检测）。目标是通过标签 $l_i \in \{1, \dots, K\}$ 标记每个标签 s_i ，这表明它的 k 对象的特征对应。为简单起见，我们首先假设 k 是已知的，放弃这一假设后，我们假设初始帧内可见目标对象是标记（手动或通过一个检测器），且中级功能处在足够精细的分辨率中，通常会遵守对象的时空范围。

CSL 操作有分配标签和删除标签功能。为了这个目的，我们从功能标签集定义一个部分标记为 l 的映射，使得 $L(S_i) \subseteq \{1, \dots, K\}$ 是不排斥 s_i 的标签。如果 $L(S_i) = \{L\}$ 然后 s_i 被分配标签 L 。我们说 s_i 已有部分标签 if $|L(S_j)| > 1$ 。CSL 迭代细化部分的标签，这是初始化，所以 $L(S_i) = \{1, \dots, K\}$ 如果 if 不从第一帧开始（即所有的标签都是可能的），否则 $L(S_i) = \{L\}$ ，其中 L 是 s_i 在第一帧中指定的标签。每个细化迭代使用成本函数和不等式约束来优化部分标记直到达到完全指定的标记。我们在下面描述了这一过程，首先忽略约束，然后将约束具体化，算法 1 给出了 CSL 的伪代码。需要注意的是关于学习算法 1 的 6-12 行，在第 4 节中阐述。

一个成本函数标记。我们假设一个成本函数 $C(S_i, l | L)$ ，成本分配给标记 S_i 取决于 $l \in L(S_i)$ ，其中 L 是当前部分被提炼的标记。鉴于目前部分标记 L ，每次迭代中首先考虑的一组活动的特点（见下文）表示 $A(L)$ 作为分配一个新的标签的可能性。成本函数 C 是用来选择一个积极的特征 $s_i \in A(L)$ 和相应的标签 $l \in L(S_i)$ 通过设置 $L(S_i) \leftarrow \{l\}$ 分配给它。特别是，对 (S_i, l) 被选择作为最小成本对考虑所有活性的功能,并标记这些功能。这些步骤是在 4,5 行和 14 的伪代码中描绘的。

```

input : A set of mid-level features  $\{S_i\}$ 
        Object labels for first frame
        Weight Vector  $w$ 
        Set of inequality constraints  $\Sigma$ 
        Learning Option learn with ground truth
        labeling  $L_{gt}$ 

output: A labeling  $L$  or weight vector  $w$ 
1 Initialize partial labeling  $L$  according to label information in first
  frame
2 while there exists  $S_i$  such that  $|L(S_i)| > 1$  do
3    $L = \text{ConstraintPropagation}(L, \Sigma)$ 
4    $D = \{(S, l) : S \in A(L), l \in L(S)\}$ 
5    $(S_i, l) = \arg \min_{(S, l) \in D} C(S, l | L)$ 
6   if learn and  $L_{gt}(S_i) \neq l$  then
7      $C^+ = \{(S, l) \in D | L_{gt}(S) = \{l\}\}$ 
8      $(S^+, l^+) = \arg \min_{(S, l) \in C^+} C(S, l | L)$ 
9      $C^- = \{(S, l) \in D - C^+ | C(S, l | L) < C(S^+, l^+ | L)\}$ 
10     $w = w + \sum_{(S, l) \in C^-} \frac{\Psi(S, l, L)}{|C^-|} - \sum_{(S, l) \in C^+} \frac{\Psi(S, l, L)}{|C^+|}$ 
11     $L(S^+) = \{l^+\}$ 
12  end
13 else
14    $L(S_i) = \{l\}$ 
15 end
16 end
17 if learn then Return  $w$ ;
18 else Return  $L$ ;

```

算法 1: 学习选择约束的序列标记算法。当标示学习被设置，则该算法进行学习权重矢量 w 的一个运行。

直观地说，我们希望 C 分配最低的成本来正确的配对可以在 l 的上下文中被最确切的标记。在这项工作中，我们将使用线性成本函数表示为 $C(S_i, l | L) = W \cdot \Psi(S_i, l, L)$ ，其中 W 是一个 m 维的权

重向量和 $\Psi(S_i, L, L)$ 返回一个描述符向量，代表相应的决策信息。在这项工作中使用的描述符是在本节和 4 节结束时给 w 提出了学习算法。

它仍然是指定活动的特点 $A(L)$ 。如果它的标签是局部的,则 s_i 是积极的, 在时间上接近一个已标记的特征 (见图 2)。限制我们对积极的特征的注意,降低了成本函数的评价次数,可导致显著的加速。

不等式约束和传播。利用成本函数的顺序使简单的跟踪问题等的所有标签是有效的,例如那些不涉及重大的闭塞。然而,更困难的问题,我们发现它很难学或手工编码的成本函数,实现国家的最先进的性能。然而,更困难的问题,我们发现它很难学及手工编码成本函数,实现国家的最先进的性能。这是由于短视,贪婪的本性所决定的顺序。幸运的是我们发现这个可怜的性能可以通过选择与约束推理相结合,克服贪婪。第 4 节提供的这种观察的理论表征。我们的 CSL 的方法是使用一组和每个视频相关不等式约束 Σ 。每个约束的形式为 $S_i \neq S_j$, 表明功能不能被分相同的标签。部分标记我违反了约束的 $S_i \neq S_j$ 如果 $L(S_i) = \{i\}$ 和 $L \in L(S_j)$ 或反之亦然。如果它不受任何约束违反 L 符合 Σ 。

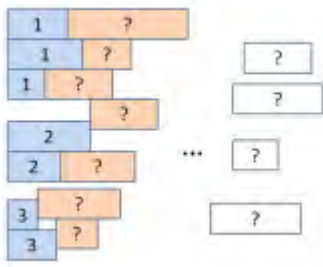


图 2: **supervoxels.Blue** 矩形序列标号的例子代表已经标记

supervoxels, 邻近的阴影长方形表示, 将考虑标注活跃 **supervoxels**, 剩下的矩形将不被考虑, 直到他们变得活跃。

结合约束 Σ 和成本函数的基本方法是限制成本函数的标记选择是不会导致违反约束, 鉴于目前部分标记 L 而以这种方式使用的限制, 可以通过降低成本函数评价次数提高效率, 我们的研究表明, 精度提高是适度的。这种方法的缺点是, 它只考虑贪婪违反约束, 而不是进一步深远意义的制约。例如, 假设目前 $L(S_i) = \{1\}$ 和 $L(S_j) = \{1, 2\}$, 我们有约束的 $S_i \neq S_j$ 。我们可以推断, S_j 标签 2 因为不能 1, 这个新的信息, 我们可以消除标签 2 从任何特征与 S_j 不等式约束的可能性, 等等。这种迭代计算的过程所隐含的后果是约束的过程称为约束传播[6], 是艺术状态约束的满意度和优化算法的一个极其重要的组成部分。鉴于当前部分标记 L 和约束 Σ 我们用 $\text{constraintpropagate}(L, \Sigma)$ 表示约束传播过程, 它返回一个精致的符合 Σ 的部分标记 L 。具体来说, 我将 L 最小细化 L (最大限度的接近), 这是一致的。 $(L, \text{constraintpropagate}\Sigma)$ 是简单的不等式约束的实现。从我反复发现违反约束的 $S_i \neq S_j$, $L(S_i) = \{我\}$ 和删除标签 $L(S_j)$ 得到一个新的标签。进程终止时, 没有发现违规行为。

CSL 在每次迭代 (第 3 行) 的开头通过调用程序集成约束传播。

因此, 在在第一次迭代标签信息将从第一帧的已知的标签传播。在后

面的迭代中，标签信息将被传播来响应标记，由成本函数先前的迭代决定。这种传播可以排除许多可能的标签可能以其他方式被认为是在当前或后面的通过成本函数的迭代。以这种方式，该限制有提高效率，因为效果的成本函数的调用数量可以减少和增加准确度，因为其中的一些去除成本函数调用可能导致错误的标签分配。我们的结果表明，该改进是显著。

处理未知的对象。以上对象 k 的数量假设在许多体育项目域被熟知。然而，**CSL** 很容易修改，以处理未知 K ，例如行人跟踪。上述 **CSL** 介绍是不变的，所不同的是 k 被初始化为所述在第一帧标记对象的数目。在迭代标签组被放大 (K 增加) 只要约束传播的结果中删除所有的标签一些特征 S (即 $L(S) = \emptyset$) 的可能性，表明使得 S 不应该被贴上任何当前的对象。每当这种情况下，一个新的对象被添加 (k 递增) 到设置的任何部分标记功能的可能的标签和 S 被分配了新的标签。当得知限制在 Σ 中是嘈杂的，我们可能会错误地删除一个正确的标签从 $L(S)$ 造成 $L(S) = \emptyset$ ，这将影响性能。我们的实验结果表明，这是罕见的，而且约束提供了压倒性的积极影响。

成本函数 **Descriptors**。给定一个局部标签 L 时，描述符函数 $\Psi(s_i, L, L)$ 的定义为：

$$\Psi(s_i, L, L) = [n_u, n_d, n_f, d_l, d_v, d_c]$$

第一种类型的描述符 n_u, n_d 和 n_f 在 s_i 的附近的未标记的特征数目，在附近的不同的标签数量，数量和帧分别不覆盖。这些描述符提供的相关措施来达到自信的一个特征如何可以被标记，它允许一个偏爱的

事物的标记时间更长的特点适应它周围的环境。第二组描述符的 DL, DV, DC 编码的特点, 在位置对象模板之间常见的视觉差异, 光流和颜色直方图。标签 I 对象模板为标记 i 联合所有 s_i 正如 $L(S_i) = \{L_i\}$ 。位置和光流, 模板被保持每一帧, 和相异度是通过平均该功能和在每个帧的重叠模板之间的欧氏距离来计算的。对于颜色直方图我们保留一个模型与相异是直接的特征与模板之间的距离。为了考虑颜色和运动的对象的之间部分的差异, 对于颜色和光流模板使用 k 均值集群化到 3 份。给定一个特征与模板, 它们相异颜色和运动根据最近的簇来计算。我们在每个 CSL 迭代后的更新相应对象的模板。需要注意的是通过引入对象模板, 成本函数变得非常高位, 相比于其中只有低阶的关系被认定的最先前的工作。

4.成本函数的学习:

现在我们描述一种学习算法用于调谐我们的线性成本函数 C 的权重 ($s_i, i | L$) = $W \cdot \Psi(S_i, L, L)$, 并给它收敛条件。我们假设标记的训练视频和提供地面真实标签分配 LGT 的功能的相关约束。这目标是学习一组权重, 使 CSL 实现了高准确度的训练数据。

这种学习的问题提出了至少有两个挑战在标准的监督学习上。首先, 训练数据是不明确的, 因为它不表示的标记所决定的确切顺序。第二个挑战是, 由成本函数作出的决定的顺序不是独立的, 因为每个决策取决于当前局部标签 L 时, 这是基于以前的决定产生。

我们的算法, CSL-Learn, 通过直接整合学习到 CSL 来应对这些挑战。外循环迭代通过每个训练视频, 可能执行多次更新的权重向量

w ，并结束当迭代的任一指定的次数达到或准确性是完美时。对于每一个培训视频，学习算法执行“学习增强” CSL 的方法通过算法 1 中给出的标志方法学打开。

该算法的行为很喜欢 CSL，直到成本函数使用 w 表明不正确的标签决定（第 6 行）。在这一点上的权重向量在今后的类似的决策点阻止错误的方式下更新（见下文）。另外，不正确的决定是不用于更新部分标记的，而是以最少的费用正确标记决定用于更新 L （第 11 行）。

它仍然以指定的权重向量更新当不正确的标签 L 被分配给特征 s_i 给定的标记 L 。我们定义 C^+ 是一组涉及活性特征的所有正确标签的决定及 $(S^+ L^+)$ 可以以最少的费用决定 C^+ 。我们还定义 C^- 是所有不正确标识组,且比 (S^+, L^+) 具有较小成本, 即其优于所有正确对不正确的对。权重向量是使用以下 Perceptronstyle 规则更新:

$$w = w + \sum_{(S_i, l) \in C^-} \frac{\Psi(S_i, l, L)}{|C^-|} - \sum_{(S_i, l) \in C^+} \frac{\Psi(S_i, l, L)}{|C^+|}$$

此规则更新 w ，以便增加所有对在 C^- 中的平均成本（高级不正确对），并减少了 C^+ 中正确对的平均成本。

学习收敛。我们分析现实的学习设定，其中可以找到一个权向量即正确的标注所有的训练数据。特别是，作为最感知式学习算法，我们结合训练过程中的造成的错误数目，直到加权矢量正确标记所有训练数据。我们必须在这里假设每个训练例子的不等式约束是与地面真相标签一致的，否则收敛不一定是可能的。

作为经典的感知算法，我们收敛担保是保证金的概念的术语。给定一组训练录像设 L 是该组中的部分标号，且与不等式约束相一致。给定一个约束集 Σ ，一个训练集的权重矢量 w 的 Σ -constrained margin 的最小值以上的所有 $w \in L \cdot \Psi(S^-, 1^-, L) - w \cdot \Psi(S^+, L^+, L)$ 的 $L \in L$ ，其中 (S^+, l^+) 的任何正确的候选标签和 $(S^-, 1^-)$ 是任何不正确的候选标签。当 w 有积极的有限余量那么使用它在 CSL 的限制将正确标注训练集。我们现在表明，一个切缘阳性 w 的存在足以收敛。下面 R 是一个常数，使得对于所有可能的特征矢量 $\Psi(S, L, L)$ 和 $\Psi(S', L', L)$ ，我们有 $k\Psi(S, L, L) - \Psi(S', L', L) \leq R$ 。

定理 1 给定约束 Σ 和任何训练集，使得存在一个权重矢量 w 与 Σ 受限余量 $\gamma > 0$ 和 $\|w\| = 1$ ，CSL-Learn 将收敛到一个权重向量能正确标记所有训练样例，使得不超过 $(R/\gamma)^2$ 权重的更新。

因此，一个更大的余量意味着一个更好的错误约束。证明呈现在补充材料中。我们涉及约束的利润。

命题 1 对于任何约束集 Σ 和 Σ' ，使得 $\Sigma \subseteq \Sigma'$ 中的任何权重矢量 w 的 Σ' 约束余量不小于 Σ 受限余量。此外存在一个训练集和约束集 $\Sigma \subseteq \Sigma'$ ，使得有一个权重矢量 w 与正 Σ' 约束余度和不存在与正 Σ -权重余量的权重向量。

综合以上结果，我们可以看到使用的限制在学习的实用性。首先，添加约束决不降低裕度且经常增加它，这意味着错误的范围的永远不

会恶化，往往还能得到提高，这表明学习变得容易。第二，有问题的地方的限制是必要的，以获得一个正余量，从而保证收敛。

5.代表和学习的限制：

我们的目标是在中等特征不等式约束的视频中学习一个约束发生器。约束应该在一定意义上有高精度度，如果 $S_i = S_j$ 是生成的，其与基础事实具有高概率一致性。此外，我们希望约束设定得尽可能大，维持高的精度，以最大限度地约束传播时间。

我们的约束的生成在精神上和以前的工作类似，即所用点轨迹[9]之间的约束。在这项工作中，每个视频帧的连接的部分分别计算和一个不等式约束被列入轨道之间，如果他们属于不同的组件。我们发现，我们的特性接近经常产生错误的限制，因为在我们的视频中它是不寻常的用于单个对象跨越在一帧中的多个连接元件。出现这种情况，例如，由于在我们的前景显着掩模的缺陷。

幸运的是，它往往仍然是，当一个对象不跨越多个部件，这些部件将满足一定的空间布局属性的情况。例如，对于人跟踪，如果我们考虑一个边界多边形为两个分量，某些类型的边界的多边形不可能对应于单个个体。基于这种认识，我们训练一个 SVM 分类器来预测在一个帧的两个组件是否对应于不同的对象。要做到这一点，我们使用标记的训练数据中提取对组件横跨视频帧和标记它们的正，如果该分量对应于不同的对象，并以其他方式分配一个负的标签。然后，我们

计算出对功能和训练的 SVM 分类。在我们的实验中，我们发现，一个简单的组四个特征是足以实现良好的结果，其中包括：部件的中心的高度和宽度包含组件的最小边界框之间的水平和垂直距离。由于了解到 SVM 分类，我们再调整其预测的阈值，使其达到较高的精度。

鉴于这种分类，我们再生成特征不等式约束 $s_i \neq 8.7$ ，如果在某些视频帧 S_i 和 S_j 的都包含在组件中，则 SVM 判定来自不同的对象。在我们的测试数据中，这种方式可以实现精确度超过 97% 的具有挑战性的排球视频，同时保持召回超过 44%，这给出了很翔实的约束集。图 3 给出了约束生成处理的一个例子。

6.实验结果:

数据集。我们评估使用基准数据集的排球和篮球视频，包括在导言中提到的挑战。还比较了基于状态检测的方法艺术，我们报告的结果在广泛使用的对人探测器有效的行人基准视频 `pets2009-s2l1` [8]。排球数据集[1]包含 38 个全部大学生打排球视频（每 200–800 帧，720×480）。排球视频是非常具有挑战性的，因为许多球员不同的身体关节间的遮挡，和非光滑的快速运动。提供六个接近队员的地面实况边界框，而对方球员没有注释或用于评估。我们扩大了地面实况，包括精细注释播放器像素、面具，这些帧中的每个玩家是提供 5 帧。我们的训练和通过 3 倍交叉验证测试报告平均横跨褶皱。虚构的数据[9]包含 18 个视频篮球动作（每帧 441×50-80，180）。

地面实况玩家面具提供每 7-8 帧。我们采用留一交叉验证评估。采用了 PETS2009-S2L2 用于成本函数的 PETS2009-S2L1 训练和限制。

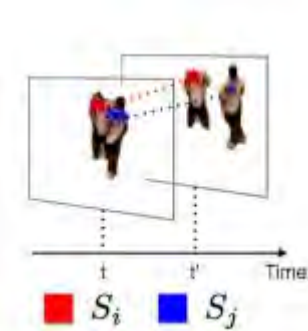


图 3: 说明确定两个 supervoxels 是否可以有相同的标签

不等式约束。帧 t , S_i 和 S_j 属于一个前景连接组件, 所以框架不会产生一个不等式约束。在帧 t' 中, SVM 分类决定了两个连接组件不能属于某一个人, 生成之间的不等式约束 S_i 和 S_j

实施细节。我们同时使用 supervoxels (CSL-VOX) 和密集的点迹(CSLDPT)来测试 CSL.Supervoxels 生成的叶级 supervoxels 是 [20] 层次的视频分割方法。CSL-VOX 唯一的输入参数是 supervoxels 的总数量, 这是通过改变所述视频分段参数来控制。我们遵循[9]来生成致密的点的轨迹。对于排球, 我们获得通过基于高斯混合颜色模型背景减法前景的前景功能。对于篮球的视频中, 我们遵循[9], 并在基于密集的点迹的显著性运动轨迹上提取前景功能。对于 PETS2009-S2L1, 我们用一个训练有素的背景模型以及显著的运动估算前台功能。公平与以检测为基础的方法 PETS2009 相比较, 我们使用检测输出来初始化的第一帧, 而不是手动初始化。

评价指标。我们使用标准的 CLEAR MOT 评价指标: 漏检 (MD), 假阳性 (FP), ID 开关 (ID-SW) 和准确性 (ACC)。此外, 为了在虚构上比较公布的结果[9], 我们使用指标从这些工作中: 每个对象的集群错误 (PRCE), 召回和跟踪时间。

基线。在排球上，我们从共同框架创建基准：1) NCuts [16]具有不同数量的簇 6 的（簇的真数）和 12，以及计算在色彩方面逐对相关度，空时的位置和 **supervoxels** 的光流。NCuts 也使用我们的不等式约束分类相同的响应被并入说明书中的亲和基质零值的条目，2) 使用公开可用的代码，采用基于流的连接人的检测，来检测网络流量 [14]。我们训练一个人在我们的领域和使用的检测器的检测作为输入。我们还更新了代码中使用的外观为主的亲和力，从而提高成绩，3) 网络流量*，以前的网络流量的方法，只是适用于地面实况边界框，而不是真正的检测。这是 Oracle 基准（因为完善的检测时）旨在评估理想化的性能与完美的输入，4) 像素的使用我们的方法基于帧片段（CSL 的框架，交叉 **supervoxels**）代替时间 **supervoxels**。对于其他两个数据集的比较，我们对比报告的国家最先进的结果。

定量结果。标签.1 显示了排球结果。首先，我们观察到，在所有的指标上，无论 CSL-VOX 和 CSL-DPT 显著胜过所有非 Oracle 的方法。而且我们看到，CSL-VOX 优于 CSLDPT 小幅，主要是因为 **supervoxels**

提供更稳定的亲和力估计。这显著改善了超像素显示了使用更连贯的时间中等特征的效用，而不是在 CSL 基于帧功能的效用。使用时间延长功能可以更显著的约束少短视的标签。以网络流量的对比表明，一个国家的最先进的检测基础技术在我们的域名面临严峻的挑战，由于在检测上做到足够准确的获取是很困难的。相反，通过中层的特点，CSL 的方法不太容易闭塞。出人意料的是，CSL 的方法和

oracleNetwork Flow*方法具有可比性，这允许欺骗和利用地面实况检测。值得注意的是，CSL 在编号开关上是显著更好的。需要注意的是由于使用完美的检测 oracle 方法一定会达到完美的 MD 和 FP 分数。

标签.2 表明 FIGMENT CSL 在密集点轨迹图划分与各种后处理的步骤的基础上优于国家最先进的两种方法[9,5]。我们推测其中的一个原因是，在现有的方法减轻硬约束在相似性矩阵中的亲和力，这可导致无意义的分区，所以必须启发式后处理。 CSL-VOX 再次优于 CSL-DPT 表明 supervoxels 在这一领域是更有效的中等特征。

标签.3 表明 CSL 的结果和在 PETS2009S2L1 上的两个顶级表现的以检测为基础的方法。CSL 达到相当的性能，这表明即使是在结构域，检测也是更可靠的，CSL 仍然是有竞争力的。The 略差精度性能主要是由于不完善约束分类，这有时排除了对某些 supervoxels 的标记，然后将它们分配新的标签。

Method	MD	FP	ID-sw.	Acc
CSL-VOX	2.69	2.49	0.41	94.41
CSL-DPT	5.32	4.67	0.87	89.14
NCut(6)	65.45	65.15	34.84	-65.44
NCut(12)	43.63	80.30	35.03	-58.93
Superpixel	5.15	5.15	37.27	52.43
Network Flow	56.36	3.33	0.91	39.40
Network Flow*	0	0	3.64	96.36
Uniform	2.99	2.89	0.45	93.67
SL	37.58	36.97	6.06	19.39
Uniform-SL	63.03	63.03	13.03	-39.09
CSL-noProp	20.91	20.60	16.06	42.43

表 1: 排球的数据集。网络流

量*使用地面实况检测。

Method	PRCE	Recall	Tracking time
CSL-VOX	17.10	82.89	89.73
CSL-DPT	19.28	43.62	79.41
[9]	20.32	31.07	75.13
[5]	86.42	0.46	1.03

表 2: 对虚构的数据集的结果。价格:

每个玩家面具错划为像素的百分比; 回忆: 每个玩家面具回忆像素的百分比; 跟踪时间: 帧, 帧数召回在 20% 以上。

Method	Rec.	Prec.	ID-sw.	MOTA
CSL-VOX	98.28	91.07	6	89.78
CSL-DPT	97.64	90.45	8	88.13
[11]	94.03	92.40	10	84.77
[22]	96.45	93.64	8	90.3

表 3: 结果 PETS2009 S2L1。

Number of supervoxels	accuracy	time
500	-96.56	23s
1000	-91.20	76s
1500	84.66	124s
2000	92.28	291s

表 4: 跟踪精度和我们用不同数量的

supervoxels 输入给 300 帧视频的运行时间的结果。

敏感性分析。标签.4 评估敏感性的唯一输入参数 - 中级要素的数量 (如 supervoxels 在这种情况下) - 对排球的数据集。正如预期的那样, 增加 supervoxels 提高精度和增加运行时间。

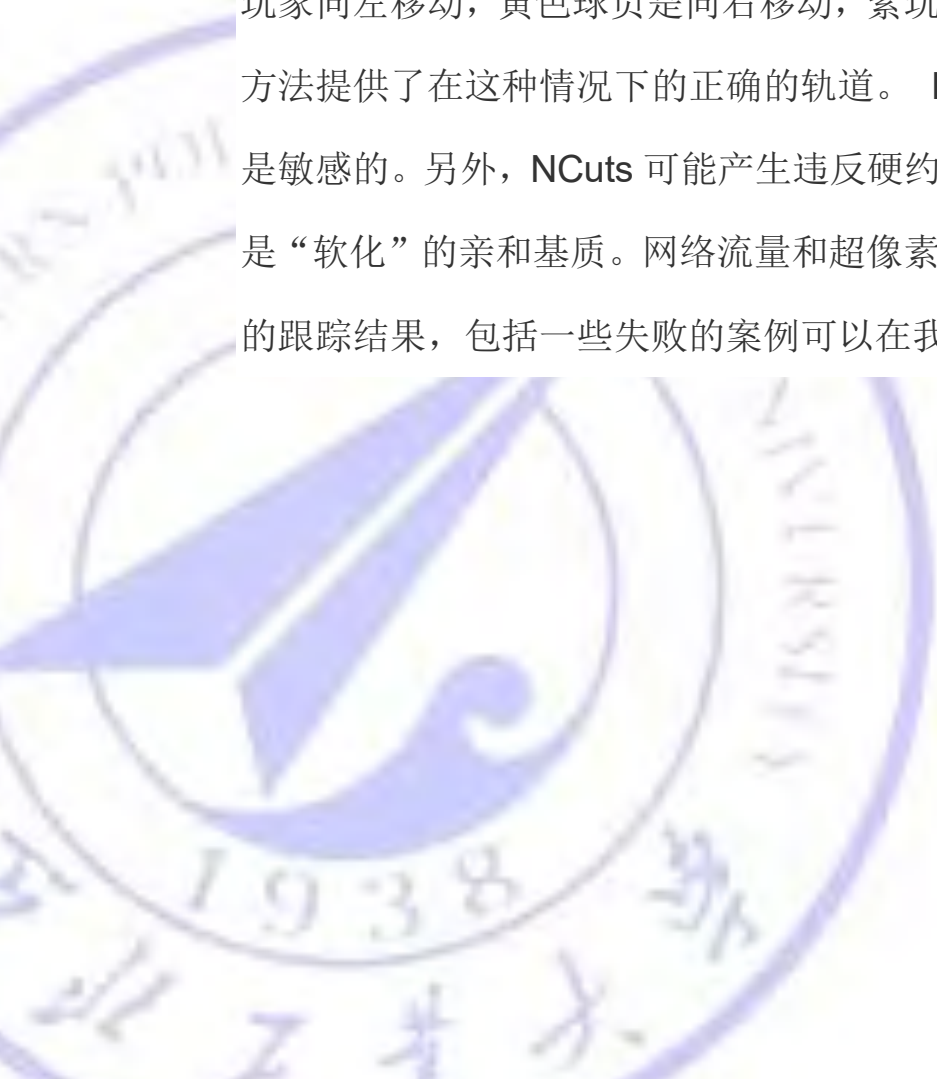
标签.1 还评估学习和使用限制使用几种 CSL 变种 (全部采用 supervoxels) 的重要性: 1) 统一是没有权重学习的 CSL, 所有的成本函数的权重设置为 1, 2) SL 是纯粹的顺序标记 (CSL 无约束) 与权重的学习, 3) 统一-SL 是没有权重学习的 SL, 并把所有的权重设置为 1, 4) CSL-noProp 是 CSL 无约束传播, 但不使用约束分配无穷成本来做立即违反约束的标号。

我们看到, 制约和约束传播是至关重要的。尤其是, SL (无约束) 比 CSL 显著差。此外, CSL-noProp 能够使用限制来改进 SL, 但是, 它不是和使用传播的 CSL 一样有效。第二个观察是, 虽然权重学习

能够提高性能（CSL 与统一），我们看到在这种情况下提高是很微小的，这可能是由于这样的事实，我们的特征被设计成相当的信息和还使约束发挥主导作用留下很少改进的余地。然而，我们看到，在学习算法上有提高的空间时是相当有效的。特别是，我们看到，学而不约束（SL）比没有学习且没有限制（统一-SL）好得多。

运行时。产生 supervoxels 和密集的点的轨迹的平均运行时间大约分别为 28S/帧和 24 秒/帧。鉴于中等特征，表 4 显示了 CSL 的运行是相当快的。此外，使用约束传播相比没有传播时，CSL 大约快 10 倍。这是因为约束传播设置大约 9/10supervoxels 的标签避免了很多成本函数评估的。

定性评价。图.4 说明了 CSL 跟踪的结果，在一个排球视频其中橙色玩家向左移动，黄色球员是向右移动，紫玩家保持静止。只有我们的方法提供了在这种情况下正确的轨道。NCuts 对簇的数目的选择是敏感的。另外，NCuts 可能产生违反硬约束的解决方案，因为它们“软化”的亲基质。网络流量和超像素困惑的两侧的球员。更多的跟踪结果，包括一些失败的案例可以在我们的补充材料中找到。



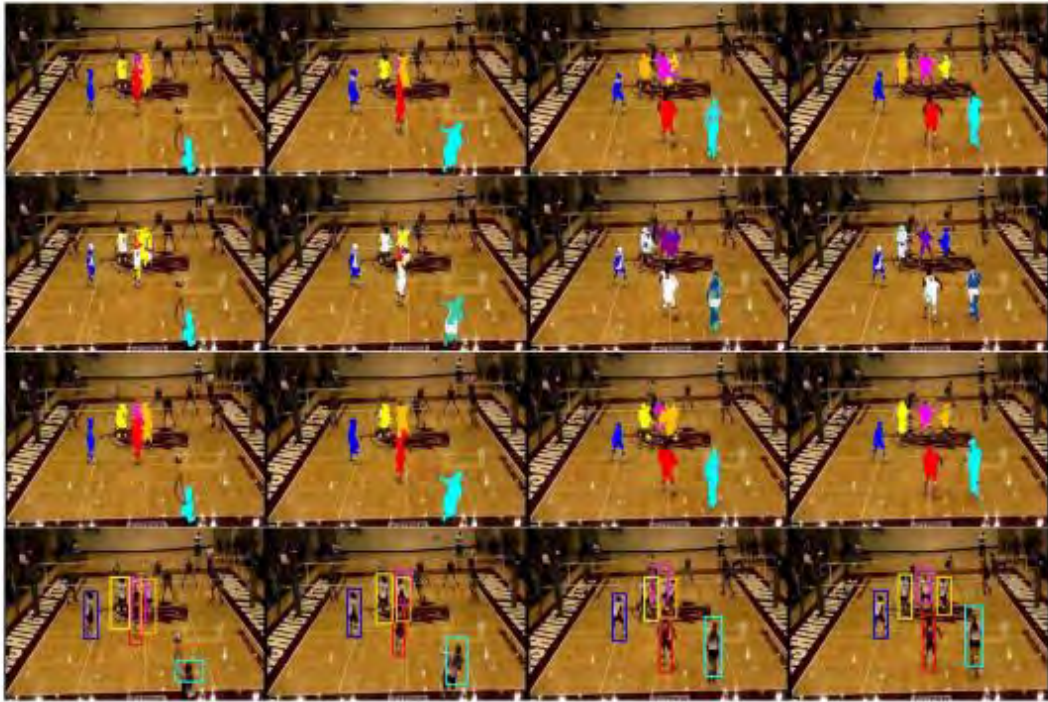


图 4: 排球的数据集: 第一行 (CSL-VOX), 第二行 (北方工业大学), 第三排 (超像素), 第四行 (网络流量*).

7.结论:

基于标记中层、时间功能例如 **supervoxels**, 我们提出来在显著闭塞时多目标跟踪的新方法。这个标签的问题放宽了现有的数据关联方法的共同假设。我们的一个主要贡献是约束顺序标记 (CSL) 的方法, 这利用硬约束和一个柔性成本函数来进行精确的标识。我们提供了学习约束和成本函数的算法, 并证明了目标函数学习在有限时间里收敛到一个精确的解时, 当这样一个解存在时。我们在排球, 篮球, 和行人跟踪实验结果表明, 该方法在当人探测器不可靠时是优于国家的最先进的探测器的, 即使检测是精确时于国家最先进的探测器也是可比的。

鸣谢

这部分工作由美国国家科学基金会资助支持的 IIS1219258, IIS1018490 和 DARPA MSEE FA8650-11-1-7149。

工具书类

- [1] M. Amer and S. Todorovic. Sum-product networks for modeling activities with stochastic structure. In CVPR, 2012.
- [2] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In CVPR, 2008.
- [3] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In CVPR, 2012.
- [4] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. In PAMI, 2011.
- [5] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In ECCV, 2010.
- [6] R. Dechter. Constraint processing. Morgan Kaufmann, 2003.
- [7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. PAMI, 32(9):1627 - 1645, 2010.
- [8] J. Ferryman and A. Shahrokni. Pets2009: Dataset and challenge. In Performance Evaluation of Tracking and Surveillance, International Workshop, 2009.
- [9] K. Fragkiadaki and J. Shi. Detection free tracking: Exploiting motion and topology for segmenting and tracking under entanglement. In CVPR, 2011.
- [10] Hal Daum´e III, J. Langford, and D. Marcu. Search-based structured prediction. MLJ, 75(3):297 - 325, 2009.
- [11] J. Henriques and J. Caseiro R., Batista. Globally optimal solution to multi-object tracking with merged measurements. In ICCV, 2011.
- [12] Y. Huang and I. Essa. Tracking multiple objects through occlusions. In CVPR, 2005.
- [13] Y. Li, C. Huang, and R. Nevatia. Learning to associate: HybridBoosted multi-target tracker for crowded scene. In CVPR, 2009.

[14] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In CVPR, 2011.

[15] Z. Qin and C. R. Shelton. Improving multi-target tracking via social grouping. In CVPR, 2012.

[16] J. Shi and J. Malik. Normalized cuts and image segmentation. IEEE TPAMI, 22(8):888 - 905, 2000.

[17] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah. Part-based multiple-person tracking with partial occlusion handling. In CVPR, 2012.

[18] B. Wu and R. Nevatia. Tracking of multiple, partially occluded humans based on static body part detection. In CVPR, 2006.

[19] J. Xing, H. Ai, and S. Lao. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In CVPR, 2009.

[20] C. Xu, C. Xiong, and J. J. Corso. Streaming hierarchical video segmentation. In ECCV, 2012.

[21] B. Yang and R. Nevatia. Online learned discriminative part-based appearance models for multi-human tracking. In ECCV, 2012.

[22] A. R. Zamir, A. Dehghan, and M. Shah. Gmcp-tracker: Global multiobject tracking using generalized minimum clique graphs. In ECCV, 2012.

[23] L. Zhang, Y. Li, and R. Nevatia. Global data association for multiobject tracking using network flows. In CVPR, 2008.

