

指导教师：_____

提交时间： 2015/3/29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 第 1 次作业

姓名： 李亮亮

学号： 2012302485

班号： 10011205



采用弱监督系统学习在一段时间中对不可区分的半透明物体进行追踪

Luca Fiaschi¹, Ferran Diego¹, Konstantin Gregor¹, Martin Schiegg¹, Ullrich Koethe¹, Marta Zlatic² and Fred A. Hamprecht¹

¹HCI University of Heidelberg, Germany, <http://hci.iwr.uni-heidelberg.de>

²HHMI Janelia Farm, USA, <http://janelia.org/>

摘要

在视频的许多帧中会出现重复很多难以辨别的半透明变形目标，我们用弱监督系统学习的方法来追踪并且消除这些特性。对于这个具有挑战性的问题，我们提出了一种新的处理遮挡的方式，这种通过共同优化多帧之间由潜在的强度流引起的复杂运动和非刚性形变来处理遮挡。这些流是用户不能分类的隐藏变量。相反，我们利用结构化学习公式针对弱用户注解找到了这个模型最好的超参数。

我们公开的方法是评估一个具有挑战性的对大量果蝇幼虫进行跟踪的数据集。我们追踪许多幼虫，尽管它们非常地难以分辨并且在持续很久的相互遮挡中个体的特性极小。

1. 简介

近几年,杰出的研究已经涌现出由生物学中的婚姻行为总结出的多目标追踪算法,如[5, 14, 6]。报道中的方法是在对果蝇幼虫社会动态进行调查时激发出来的,而果蝇是生物界中一种流行的模型。这些实验要用单摄像机来可视化的追踪在孔板上运动的大量相互干扰的幼虫个体。而要在控制实验条件下促进发现并分出独立的幼虫个体,最大的挑战是在几帧之间如何分辨和追踪相互接触并遮挡的小虫子。在这种情况下,有两方面因素妨碍了这个利用对象的外观特征(如:

[9, 30, 23, 11]) 和运动模型 (e.g. [18, 7]) 来处理相互遮挡追踪算法的设想:

- 幼虫是半透明的,非常地难以区分,以及它们都没有特定的形态。
- 幼虫在互相接触的时候会表现出不规律的运动。

在没有假设这些被追踪的目标可区分的前提下,我们的方法就可以处理复杂的运动和非刚性的形变。这种方法的核心,我们用模拟观测到相互遮挡的图像强度作为每个个体对象的混合潜强度。因此,我们通过共同优化大量强度流的运动来分辨对象的身份。这就致使一种新的灵活的模型,这个模型参数由一般的低水平线索均衡后导出。相比人工地来协调这些参数,我们这种方法和 [9, 30, 23, 11, 18, 7] 的一个重要的不同是我们开创了结构化的有潜变量 [28] 的学习方式从注解中学习最佳的权重。我们的主要成果包括:

- 第一个以分辨半透明相互遮挡的可变形个体对象为目标的多目标追踪问题的求解公式。
- 一种弱监督学习策略,花费最小的用户成本来用参数表示相应的能量术语。
- 低密度追踪特性很少的果蝇幼虫的解决方案。我们在我们为以后研究¹做的一个公共具有挑战性数据集上测试我们的方法。

¹Download at <http://hci.iwr.uni-heidelberg.de/Benchmarks>

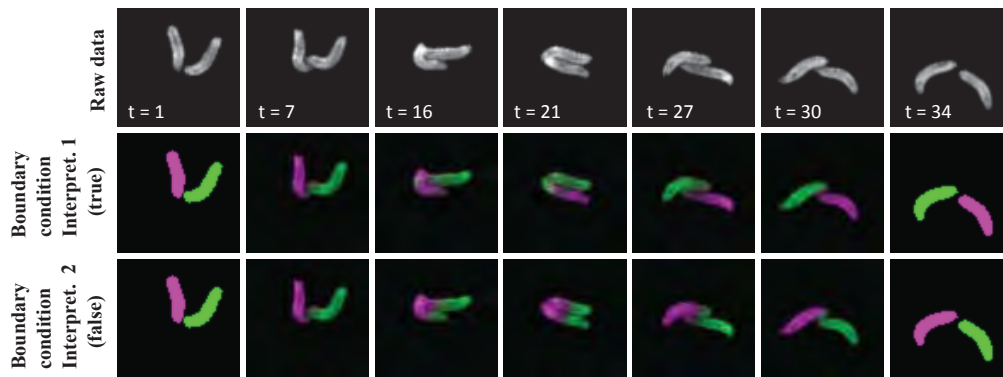


图 1. 上面一行: 从原始数据中选择出来的子帧. 中间和下面一行: 这个序列的两种可能的解释. 对象的解释是特征按照边界特征进行编码(这里用不同深度的颜色来表示,第1帧和第34帧). 在这种模型下 (Eq. (1)), 推论允许我们作出潜变量情形的判断 (用不太饱和的颜色来表示) 并且给与当前边界条件相联系的解释以有力的支持. 给定一些练习数据, 还是按照边界条件, 学习算法在找模型参数以便使正确的解释比其它错误的解释可能性大些.

2. 相关工作

用单一摄像机追踪多目标在计算机视觉领域已经有很长的一段历史了, 不过大多数都是专注于对行人和车辆的追踪. 这些处理相互遮挡的方法主要利用被追踪对象的两个关键特征: (i) 它们是可区分的, 例如行人穿着不同的衣服或汽车的不同颜色, 和/或 (ii) 它们有着简单可预测的运动轨迹 (比如可以很好地用线性运动模型来描述). 这两种特征被好多结合运动和外观的模型用来一帧一帧地追踪, 如[18]那样, 而且最近被用于全局数据关联技术. 这些以共同优化候选的身份信息为目标根据时间整合局部外观和运动得来的线索, 如[11, 30, 7, 23]. 混在一起的对象一般被当做错误的推测来处理. 在[2, 9]中, 作者为了明确重合在一起的信号提出了一种全局的推理计划, 用源于遮挡前后的外观和运动特征信息来匹配目标. 类似于这种方法, 我们之前的工作[8]依靠我们所展示的幼虫间遮挡可以被可靠地检测到. 和[9]不同的是, 我们避免了依靠可被区分的对象特征, 而是寻找使得潜在强度流表现出的情形与观测到的情形一致的解释.

动物追踪与分割: 最近生物学中对图片进行自动分析的需求提出了一系列新的问题, 在自然图片中追踪罕见的生物群落. 事实上, 一方面生物学所关心的对象是高度可形变的, 经常是难分辨的, 有时还是半透

明的; 另一方面, 这些对象所表现出的动作是随机出现的, 而且它们还有可能会分成不同的几部分. 在生物图像分析中, 大多数研究都在明确地一帧一帧的追踪像细胞那样的滴状物 [15, 3, 22]来解决重叠和形变. 在[15]中,相重合的细胞用水平集与运动过滤的结合来分辨, 而[3]用的是一帧一帧局部轮廓匹配和[22]混合高斯模型. 出色的工作靠粒子滤波器追踪对象的轮廓来追踪小鼠[5], 但这种方法局限于小数目的人工设计形状模型的仿射变换. 粒子滤波器也已被用于蚂蚁的追踪 [14], 但是在大多数的软件包中用的都是简单的独立匀速的模型, 比如用于果蝇的Ctrax [6]. 最近的一项对虫性对象追踪的回顾[10]. 据作者所讲, 所有的这些软件都是在不拥挤的条件下工作的, 而且还没有处理遮挡问题. 轨迹在相互遮挡的情况下会终结而后再重新启动, 而身份错误要手动来编辑[6]. W'ahlby [26]已经提出了一种分割虫子重叠的方法. 然而, 这种方法依赖于有很小概率并排躺在一起的长对象骨架和外表所产生的候选片段.

结构化的学习和运输理论: [18, 5, 11, 30, 15, 3, 9, 23, 7]中的方法需要手动设定的参数. 我们的方法依赖Lou *et al.*在[17]中计划用弱监督学习系统[28]来确定细胞追踪[16]的联合数据模型的参数的这一直觉. 沿着[17], 我们展示了如何从弱用户注释中学习出能量参数.

但是，我们的应用以及能量公式和[16, 17]是不同的。这些工作中，在重叠区域所丢失的对象的身份信息只有在一帧一帧的分析中才被考虑到。

我们模型的想法也是来自文献地球移动距离(EMD) [21]。EMD找出了源和水槽两个块之间的最小花费。这个距离在追踪里面有好几个应用，如[20, 19, 27]。Ren *et al.* 在[20]里面介绍了用于单目标追踪的EDM并且用它来对观测到的强度流来建模。Oron *et al.* [19]介绍了EMD花费可以在线更新，Wang *et al.* [27]打算从数据里面学习出这种算法的值。我们延伸了这一想法并且允许这些块儿可以有不同的颜色，我们连带地使好多差不多相似的流之间的花费最小化，但这和多物品留不太一样[1]。这些都是用于在同一网络中研究将多可种区分产品的运输总成本最小化的操作。

3. 问题的求解公式及模型

我们的想法在图. 1中进行了描述。当两个难以区分的对象重叠又再次分开时，我们对于它们的身份评估有两种可能的解释。直觉上，假设每一个对象都有一种唯一的颜色，但我们只有一个单色的传感器。在这种情况下，每一个像素上的颜色都是潜在的变量，然而像素的灰度是可观测的数据。根据当前的这种解释给出一种边界条件（即每一个个体在相互遮挡前后指定身份），我们的模型是一个允许将估计到的所有潜在变量作为最小化能量方案的能量函数。这种学习方法的目的是找出模型的参数使得正确的解释得到最低的能量。

我们的模型是EDM在大量果蝇幼虫被大量不同的色块流代替后的普遍化。这些色块流在几帧中被下面这些点所表达的花费共同优化：

- 为了获得活动最贴近的解释，色块的移动应该尽可能的小，还应该满足边界条件。
- 每个对象里面的每种颜色的空间分布应该是平滑的。
- 一个像素中所有颜色的强度加起来应该接近这个像素观测到的颜色强度。
- 颜色只能随着时间近似地保持，这会导致图像中的所有强度波动，泊松分布和传感器的噪声和传

感器的饱和度也会造成影响。

系统地来阐述这个能量函数，我们制定了两种设计选择。第一个，用多个特征得到的花费的加权和，我们确保这个模型具有充分的表现力来允许为每个测试样例的正确方案设定最低的能度。第二个，我们约束了我们的公式使之对潜变量的凸起达成协议，从而允许了有效的优化。

3.1. 准确的公式

公式域 (1) 中的能量函数给出了前几段概念的表达（参见表1）。它取决于两类变量：流数 $\{f\}$ 块数 $\{m\}$ 。如果我们在时间和空间上对视频中的每一个像素用 $i = 1, \dots, N_P$ 来进行编号，而其中一个色块变量 m_i^k 代表第 i 块像素幼虫 $k = 1, \dots, N_L$ 的色块强度和。这些块可以被观测到关于问题的变量或者潜变量。特别地，每个孤立对象中的像素块是可被观测到的变量，这是由于在学习中我们可能会要求用户用轻轻的划一下来标出完全独立的对象。相反的，对象发生相互重叠的像素与好多的潜在的变量都有关系。而为了区别观我们测到的变量和潜变量，我们用独立的记号 \dot{m}_i^k 和 $\overset{\circ}{m}_i^k$: $\{\dot{m}\} \cap \{\overset{\circ}{m}\} = \emptyset$ 和 $\{\dot{m}\} \cup \{\overset{\circ}{m}\} = \{m\}$. 我们预留了记号 m_i^\dagger 来表示测到的像素 i 的灰度值，我们收集了容器 $\dot{m}, \overset{\circ}{m}, f$ 中的集合 $\{\dot{m}\}, \{\overset{\circ}{m}\}, \{f\}$ 。

$$\begin{aligned}
 E(\underbrace{f, \overset{\circ}{m}}_{\text{latent variables}}; \dot{m}, w) = & \underbrace{\sum_{i=1}^{N_P} \left| \sum_{k=1}^{N_L} m_i^k - m_i^\dagger \right| d_{ii}^1}_{\text{data fidelity term}} w^1 \quad (1) \\
 & + \underbrace{\sum_{i=1}^{N_P} \sum_{k=1}^{N_L} \left| \sum_{j \in \mathcal{N}_i^{st}} f_{ij}^k - m_i^k \right|}_{\text{outgoing flow conservation term}} w^1 \\
 & + \underbrace{\sum_{j=1}^{N_P} \sum_{k=1}^{N_L} \left| \sum_{i \in \mathcal{N}_j^{st}} f_{ij}^k - m_j^k \right|}_{\text{incoming flow conservation term}} w^1 \\
 & + \underbrace{\sum_{k=1}^{N_L} \sum_{ij \in \mathcal{N}^s} |m_i^k - m_j^k| d_{ij}^2}_{\text{smoothness term}} w^2 \\
 & + \underbrace{\sum_{l=3}^M \sum_{k=1}^{N_L} \sum_{ij \in \mathcal{N}^{st}} f_{ij}^k d_{ij}^l}_{\text{flow cost term}} w^l
 \end{aligned}$$

为了给运动建模，我们定义了视频中有向图相关联的像素流。这个图的连通性是建立在连接相邻帧的像素时空关联度 \mathcal{N}^{st} 上的。流变量 f_{ij}^k 被分配给每一条边

$ij \in \mathcal{N}^{st}$ ，代表这个幼虫 k 在 t 时刻处于像素 i 和 $t+1$ 时刻处于像素 j 在之间的色块流。这些流总是在相邻帧间具有时间相干性的问题的潜变量。

这个能量值是 w^l ， $l = 1, \dots, M$ 的加权和。这些值被分成四类：公式域 (1) 中从下到上，根据取决于第 l^{th} 张图的参数 d_{ij}^l 块之间流的花费是不同的。这个花费 d_{ij}^l 不仅仅可虑到空间距离，表现为各种力度的欧几里距离，而且在局部外观上也是不相同的，等等。（参见3.4段）。这个平滑因子在毗邻的像素块有相似的颜色时支持空间平滑方案。对每一帧，这个因子靠一对像素的空间距离 \mathcal{N}^s 定义。数据真实度证实这些像素的颜色和接近于灰度值 m_i^\dagger 。最终，流保持因子在相邻的时间步上执行近似像素的当前一致性。在传统的流问题中，这些因子是强加的线性约束，而我们将这些因子包含在对象波动强度的解释中。尽管公式没有要求，我们通过给数据精度和流量守恒因子相同的权值 w^1 来减少参数的数量。

在边界条件被观测到的块变量 $\dot{\mathbf{m}}$ ，给定时，公式域 (1) 的优化允许估计潜在的块和流， $\dot{\mathbf{m}}$ 和 \mathbf{f} 。能量优化问题

$$\underset{\mathbf{f}, \dot{\mathbf{m}} \in \mathbb{R}_+^M}{\operatorname{argmin}} E(\mathbf{f}, \dot{\mathbf{m}}; \dot{\mathbf{m}}, \mathbf{w}) \quad (2)$$

在用辅助变量替换掉绝对值的时候权值 \mathbf{w} ，流 \mathbf{f} 和块数 \mathbf{m} 上都是线性的。这是一种标准的技术，参见[4]。因此，问题 2 是可以线性规划²来有效的解决一个凸问题。

3.2. 推断：身份解释

问题 2 要求块数 $\dot{\mathbf{m}}$ 联合所有的孤立（非重叠）像素作为边界条件。在训练集里面，这些边界条件是观测到的；在预测的时候，它们只是可以被观测到的（而不是观测到的）。为了辨别并追踪所有的目标，我们寻找并识别那些根据我们的能量函数似乎是最

²The optimization of Eq. (1) is implemented using ILOG Cplex.

符号	定义
t	时间序号 $t = \{1, \dots, T\}$
i, j	像素标号。在视频的时空量里面共有 N_p 个像素： $i, j \in \{1, \dots, N_p\}$
\mathcal{N}^s	同一帧中相关像素的空间距离
\mathcal{N}^{st}	相邻帧中相关像素的时空距离
l	特征序号， $l \in \{1, \dots, M\}$
k	块中颜色序号， $k \in \{1, \dots, N_L\}$
f_{ij}^k	幼虫/颜色 k 和像素 $i, j \in \mathcal{N}^{st}$ 相联系的流
d_{ij}^l	像素 $i, j \in \mathcal{N}^{st}$ 从第 l^{th} 个特称计算出的花费
w^l	和第 l^{th} 个特征相联系的权值
m_i^k	像素 i 中和幼虫/颜色 k 相联系的块。块数可能是： \dot{m}_i^k 可观测的变量 m_i^k 潜在的变量
m_i^\dagger	在像素 i 中测到的灰度值

表 1. 注释。

可信的边界条件。有效的边界条件必须满足两个约束：第一点，每一个独立的对象中的所有像素必须被标记成相同颜色的块；第二点，每种颜色在每一帧中只被用来标记一个目标。[2, 8] 中那样的方法把整个视频分成了没有遮挡的部分（易于追踪的）和遮挡事件（难以追踪的）。在遮挡事件中，这两个约束条件保证了每一个处于遮挡区域的单独个体都能和前一个目标留下的区域相匹配。如果我们定义 \mathcal{I} 为有效解释的空间，为了找出所有目标的能量最低的身份分配，我们解决了下面的最优化问题：

$$\underset{\substack{\dot{\mathbf{m}} \in \mathcal{I} \\ \mathbf{f}, \dot{\mathbf{m}} \in \mathbb{R}_+^M}}{\operatorname{argmin}} E(\mathbf{f}, \dot{\mathbf{m}}, \dot{\mathbf{m}}; \mathbf{w}) \quad (3)$$

相比于明确地合并 \mathcal{I} 的约束条件，我们探索了这个问题的结构：我们简单反复地解决了 $N_L^c!$ 个关于 N_L^c 个聚成簇的对象可能识别出的解释。注意到这点，在大多数情况下， N_L^c 在我们的数据中只有两三个。

总的来说，当公式域 (2) 中的函数是凸函数时，公式域 (3) 中的附加约束导致了和局部最优组合数有关的非凸的优化问题。我们探测了直到三个对象的相互遮挡。对于四个或者更多相遮挡的对象，我们求助于在4.2段中的近似方法。

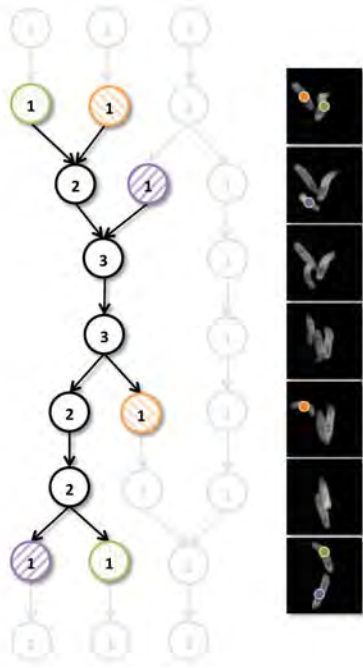


图 2. 与时空上遮挡相一致的样例的图解。在左边，这幅计数图是从[8]中的算法获取到的：每一个深色的线连起来中都标记上了里面包含的目标条数。粗线表示子图中三条幼的圈虫发生遮挡的地方。在右边，是和子图描述相一致的真实区域。既然在每一帧里，只有目标和子图的结点是相一致的。在左边和右边：在训练阶段这些颜色区分的出入遮挡的个体被人类专家标记出来了。人类专家用更高的瞬时清晰度展示这个序列并且可以提供可信的解释。

3.3. 用部分注释学习参数

学习确定了公式域 (3) 中的权重 \mathbf{w} 所以正确的解释会比其他所有的错误解释获得到更低的能量值。我们通过结构风险最小化原则[25, 28, 17].解决了这个能量参数化问题。我们的训练样例, $n = 1, \dots, N$ 对于可被观测到的色块变量是仅在重叠区域前后按照用户注释的孤立幼虫身份标定的, 像图.2中描述的那样。可被论证的, 这个注释需要用户最小的点击力。

一个训练样例 $(\hat{\mathbf{m}}_n, \mathbf{m}_n, \mathbf{f}_n, \mathbf{d}_n)$ 是由测到的变量和潜变量以及图像区域 \mathbf{d}_n 的特征组和而成的。这个学习问题所缺少的函数是对正确解释的能量和所有错误解释能量的最小值间的正偏差的修正。如果我们

定义 $\mathcal{I}_n^- = \mathcal{I}_n \setminus \hat{\mathbf{m}}_n$ 作为一组错的但有效用的边界条

件的话, 和[28]相似的, 这个缺少的函数可以这样写出来:

$$L(\mathbf{w}, \hat{\mathbf{m}}_n) = \min_{\mathbf{f}, \hat{\mathbf{m}} \in \mathbb{R}_+^M} E(\mathbf{f}, \hat{\mathbf{m}}; \hat{\mathbf{m}}_n, \mathbf{w}) - \min_{\substack{\hat{\mathbf{m}} \in \mathcal{I}_n^- \\ \mathbf{f}, \hat{\mathbf{m}} \in \mathbb{R}_+^M}} [E(\mathbf{f}, \hat{\mathbf{m}}, \hat{\mathbf{m}}; \mathbf{w}) - \Delta(\hat{\mathbf{m}}_n, \hat{\mathbf{m}})] \quad (4)$$

这里, $\Delta(\hat{\mathbf{m}}_n, \hat{\mathbf{m}})$ 是这个函数中仅仅取决于可测量的部分。在这个研究里, 我们用的是可被预测到的解释和训练集中黄金标准身份变换的数量。注意到这点, 我们关于错误但有效的推论集

\mathcal{I}_n^- (在3.2段中) 造成了这个不是局部变量里附加的最优化问题一。这个模型是用寻找使正规结构化风险最小的最佳权值来训练的:

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|_2^2 + \frac{C}{N} \sum_n |L(\mathbf{w}, \hat{\mathbf{m}}_n)|_+ \quad (5)$$

$$\text{s.t. } w^1, \dots, w^M \geq 0$$

这里, $|\cdot|_+ := \max(0, \cdot)$ 是一个确保从损失中获得积极贡献而且可以保证公式域的 (1) 中凸性的转移函数。

公式域(4) 中的遗失是非凸的但是是(DC) [24]中两个凸函数的替代。根据 [28], 我们用CCCP过程 [29]找到了关于公式域(5)中问题的最小值。简要地来讲, 这个迭代计划将估计到的潜变量最可能的状态和将所有潜变量视为可测的相替换来解决SVM结构问题。我们基于SVM³结构的n-slack[12]公式实施了CCCP。合起来, 这就允许我们从非常弱的注释和通用特征中学习出模型的参数。

3.4. 特征

多重表达的特征 d_{ij}^l , 确保公式域 (1) 中的模型可以区分序列中可能的推断。这和先前的这些人工参数化函数的工作[9, 30, 23, 11, 7]形成了对照。所有的特征都在表2中做出的概括。除了像从两个像素的空间位置来计算的欧氏距离的平方这类的传统特征外, 我们提出了确定数据中时空位置的一系列新的特征。特别的, 这些特征都是从强度剖面 $\Phi_j(s)$, $s \in [0, 1]$ 中导出来的。只是从第 t 帧中的第 i 个像素到第 $t+1$ 帧中的第 j 个像素计算出来的, 根据两个像素的空间位置联系来参数化 s 。像表 2 中的那样,

³Our implementation builds on the open-source library [Pystruct](#).

我们收集了三组数据：第一组是特征乘上平滑因子和数据保真项，第二组是纯粹的空间花费，第三组包括了由 $\Phi_{ij}(s)$ 区分的特征。这些特征的学习权重在图3中。

特征	
d_{ii}^1	$\exp(-m_i^\dagger/255)$
d_{ij}^2	constant
d_{ij}^3	像素 i, j 间的空间欧氏距离
d_{ij}^4	Second power of d_{ij}^3
d_{ij}^5	Fourth power of d_{ij}^3
d_{ij}^6	$ m_i^\dagger - m_j^\dagger $
d_{ij}^7	$\int_0^1 \Phi_{ij}(s) ds$
d_{ij}^8	$\max_{s \in [0,1]} \Phi_{ij}(s) - \min_{s \in [0,1]} \Phi_{ij}(s)$
d_{ij}^9	$\langle (\Phi_{ij}(s) - \langle \Phi_{ij}(s) \rangle)^2 \rangle$
d_{ij}^{10}	$\int_0^1 I_{\{\Phi_{ij}(s) \leq \text{background}\}} ds$

表 2. 用到的特征的定义。 I 是一个指示函数并且背景的入口参数被设定为50。

4. 实验

在我们的试验中，我们比较了最近[6, 16, 9, 22]这些工作中提出的方法。首先,我们用已经建好的可以追踪多个果蝇成虫的软件Ctrax [6]。Ctrax用基于匀速模型的逐个画面的追踪并且自动合并通过多点修正假设探测到的重合。然后,我们把[9]和[16]中的想法结合在一个我们称之为L-BM的方法中,以便平衡[9]中用过的两端的匹配。简要地来讲,在[9]中我们计算孤立幼虫进入重合区域和那些重合区域中存在的这两边相匹配的最小花费。此外,花费是三个特征的线性组合:孤立幼虫中心间的欧氏距离,在尺寸和平均强度上是不同的;如同在[16]中,最佳权重是通过SVM结构学习得到的。第三点,我们和持续跟踪(CT)[22]对比。像我们提议的那样,CT先是检测发生遮挡的簇,然后再处理掉这些遮挡。与我们的工作相对照,CT适合于对每一时刻的每一遮挡域构建混合高斯模型,然后通过数据关联来追踪。为了公平地和我们的方法进行比较,我们用[8]中的结果初始化了两种算法,并且通过格点搜索[22]对参数进行了优化。

4.1. 数据集及评价标准

我们用一组由33个高分辨幼虫追踪视频组成的具有挑战性的数据测试了我们的方法。

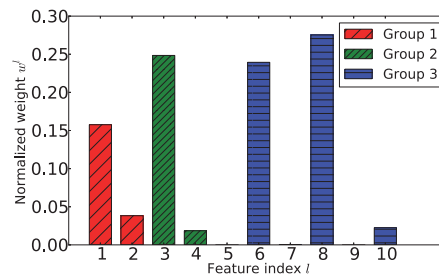


图 3. 参数 w 是从数据中学习得到的而且归一化到和为1。

每一个视频长达5分钟,时间分辨率是每秒3.3帧(总共1000帧)而且平均包含20只幼虫。空间分辨率是135.3

微米每像素且每帧有1400*1400个像素。关于预处理和计算图的构建我们根据的是[8]中的指南。那个计算算法在这个数据集上得到了99.9%的精度,包括了对于独立幼虫近乎完美地追踪。因此我们的评估主要关注的是动物之间互相遮挡重叠的分割上。我们从计算图中提取出了有两个或两个以上的聚簇的子图。这些区域很稀疏并且和两个幼虫的遮挡很符合。以为人类专家对每只幼虫进入和离开每个区域进行了手工地标记以便做出一个黄金标准,正如图2中补充的那样。和[13]很相似,性能的测试是计算算法的输出和黄金标准不一样的特征的个数。错误的检测和跟丢的情况都被记到了错误中。这个测试标准后来用进入到区域的目标总数进行了归一化:1478个区域有两只幼虫,96个区域有3只幼虫,28个区域有4只或者更多的幼虫。

4.2. 实现和实验细节

为了减少我们提出的这个方法的计算量,用到了三个近似值。第1个,为了减少优化中变量的数量我们为每一个相碰选择了时间上呈线性的多达15个子帧。第二个,相邻的时空距离 \mathcal{N}^{st} 的阈值适应性的被选做每个计数器的最小距离以便每一个前景像素可以被至少一条边连起来(平均有10个像素)。空间上相邻的距离 \mathcal{N}^s 直径为1个像素。第三个,当重叠区域有四只或者更多的幼虫时,L-BM只能得到一开始6个解释的最小匹配代价。在这些和我们的主方法之下,我们选择了获得能量最低的那个解释。

我们主要的提议和L-BM一样是仅仅用25组两只幼虫相遇的例子(所有幼虫相遇的3%)训练的,而确是在

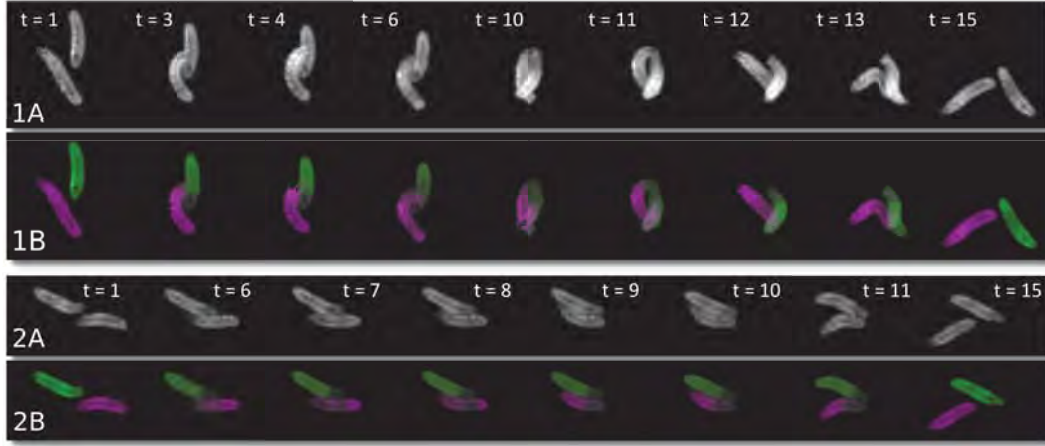


图 4. 测试数据集中的两个例子。原始数据 (1A, 2A) 和我们的算法推出来的最低能量解释 (1B, 2B)。在第1帧和第15帧中目标是分开的而且块的值是边界条件设定的。在中间的这些帧里，两种颜色之和与原始数据的灰度特别的接近而且受到公式域 (1) 中平滑因子的影响。

Method	Total	$N_L^c = 2$	$N_L^c = 3$	$N_L^c \geq 4$
Random Guess	51.5%	50.0%	66.7%	76.1%
Ctrax [6]	13.2%	11.5%	26.3%	57.1%
L-BM	11.8%	10.5%	23.1%	42.8%
CT [22]	9.4%	6.7%	34.7%	64.3%
This proposal	5.3%	4.2%	14.7%	32.1%

表 3. 数据集中的身份分配错误: 分为二, 三, 四条和更多幼虫的遭遇, 和这几种类型的平均权值。

整个数据集上进行测试的。大部分相遇的时间都非常地短, 所以采集到了很难的样例, 我们执行的第一轮的推断后又随机的找了15和10个困难样例的混合来训练这个模型。为了评估执行中随机选择的训练集的可变性, 这个有很多训练样例数字算法的学习曲线被收录在了补充材料中。训练我们的模型差不多花了8个小时, 而在推断中在主频为2.4GHz的Intel的高性能4和机器上所需时间的中值是每个视频1.3个小时。其中90%的时间都花在了3条及更多的幼虫相遇上。更多关于运行时间的细节参见补充材料。

4.3. 结果和讨论

图 4 和图 5 定性的说明了我们的结果。对于测试集中的两个挑战, 图 4 展示了关于潜在块的最低能量解释的推论情形。图 5 比较了三条幼虫相遇的六种可能的解释的潜在变量的情形。由于空间的限制, 更深一步的高时间分辨率的结果和重建的追踪视屏请参见补充材

料。

我们的方法和其他方法的定量对照在表3中。L-BM和Ctrax对这个数据集有相似的表现。然而L-BM方法在错误检测和从训练样例中学习参数中有较少错误的倾向。对于两到三只幼虫重合的情况, 我们的方法在没有用任何外貌特征的情况下一直有着更加精确的结果。在四只或者更多幼虫的情况下, L-BM习惯于在多个候选解释中检索, 无论多么得接近真实满意的表现, 我们的方法都可以结果。对于所有的方法, 性能都随着幼虫的数量的增多情形变得混乱, 而且解释的数量也随着增多。特别地, [22]假定了大致的高斯形状, 为我们的方法对于两条幼虫相遮挡的情况提供了一种计算量较少的选择, 当对于多于两个的目标发生缠绕的情况又有困难, 定性和用数字描述都在表 3 中。

在这个数据集中, 我们这种方法的精度胜过了其他的方法。即便如此, 在一些应用中对于更多个体的结果的出错率还是很高, 而且在这些设置中计算时间还需进一步的提高。

总之, 我们提出并且评估了一种新的追踪算法, 这种算法特别适于具有挑战性的多个难以区分的半透明目标的身份分辨问题。这允许是可形变的和复杂运动, 比如从另一只身上爬过。总的来说, 公式域 (1) 表达了一个“最小作用原理”, 在适当的参数化之后, 一

个情形可以靠最小转变或者大量不同颜色的流动被解读出来。这个结构化的学习框架允许调整这些运输过程的花费，以便在给定训练集的情况下做出真正花费最少的解决方案。只在单个幼虫进出遮挡团的时候指定幼虫的身份就可以很方便的编译好一个数据集。我们解决了紧密联合的潜变量的学习问题，而且在挑战性的群居幼虫序列上获得了令人鼓舞的成绩。虽然我们这个方法对于大型群体的复杂度还没有被衡量，但它打开了动物社会互动研究中的一扇新的大门。

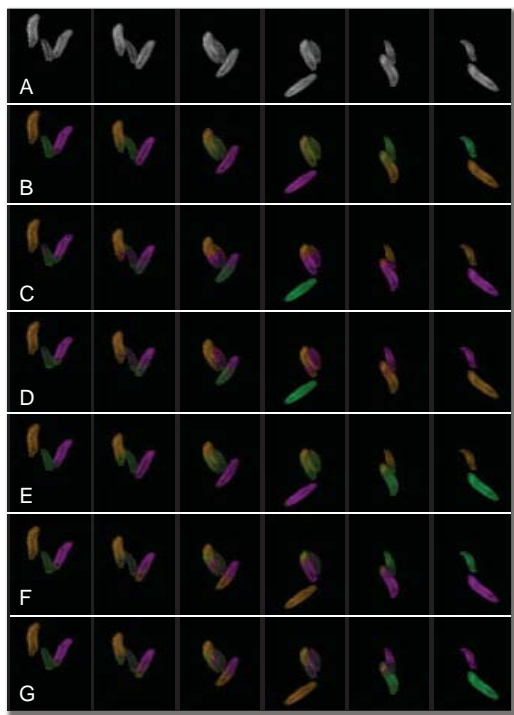


图5 从三条幼虫相遇里面选出来的子帧。从上面可以看到：第一行原始数据 (A) 和按照能量递增顺序 (从B到G) 排列的6中可能的解释。我们的方法正确的将能量最低的解释排在了第二行 (B)。

参考文献

- [1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows: theory, algorithms, and applications*. Prentice Hall, 1993. 3
- [2] B. Bennett, D. R. Magee, A. G. Cohn, and D. C. Hogg. Using spatio-temporal continuity constraints to enhance visual tracking of moving objects. In *ECAI*, 2004. 2, 4
- [3] R. Bise, K. Li, S. Eom, and T. Kanade. Reliably tracking partially overlapping neural stem cells in dic microscopy image sequences. In *MICCAI-OPTIMHSE*, 2009. 2
- [4] J. Bisschop. *AIMMS-Optimization modeling*. Paragon Decision Technology, 2006. 4
- [5] K. Branson and S. Belongie. Tracking multiple mouse contours without too many samples. In *CVPR*, 2005. 1, 2
- [6] K. Branson, A. A. Robic, J. Bender, P. Perona, and M. H. Dickinson. High-throughput ethomics in large groups of drosophila. *Nature methods*, 2009. 1, 2, 6, 7
- [7] R. T. Collins. Multitarget data association with higher-order motion models. In *CVPR*, 2012. 1, 2, 5
- [8] L. Fiaschi, G. Konstantin, B. Afonso, M. Zlati, and F. A. Hamprecht. Keeping count: leveraging temporal context to count heavily overlapping objects. In *ISBI*, 2013. 2, 4, 5, 6
- [9] J. Henriques, R. Caseiro, and J. Batista. Globally optimal solution to multi-object tracking with merged measurements. In *ICCV*, 2011. 1, 2, 5, 6
- [10] S. J. Husson, W. S. Costa, C. Schmitt, A. Gottschalk, et al. Keeping track of worm trackers. *WormBook: the online review of C. elegans biology*, 2012. 2
- [11] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In *CVPR*, 2007. 1, 2, 5
- [12] T. Joachims, T. Finley, and C.-N. J. Yu. Cutting-plane training of structural SVMs. *Machine Learning*, 77(1):27–59, 2009. 5
- [13] B. Keni and S. Rainer. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008. 6
- [14] Z. Khan, T. Balch, and F. Dellaert. MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements. *TPAMI*, 2006. 1, 2
- [15] K. Li, M. Chen, T. Kanade, E. D. Miller, L. E. Weiss, and P. G. Campbell. Cell population tracking and lineage construction with spatiotemporal context. *Medical image analysis*, 12(5):546, 2008. 2
- [16] X. Lou and F. A. Hamprecht. Structured learning for cell tracking. In *NIPS*, 2011. 2, 3, 6
- [17] X. Lou and F. A. Hamprecht. Structured learning from partial annotations. In *ICML*, 2012. 2, 3, 5
- [18] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: multitarget detection and tracking. In *ECCV*, 2004. 1, 2
- [19] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. Locally orderless tracking. In *CVPR*, 2012. 3
- [20] X. Ren and J. Malik. Tracking as repeated figure/ground segmentation. In *CVPR*, 2007. 3
- [21] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. In *ICCV*, 2000. 3
- [22] M. Schiegg, P. Hanslovsky, B. X. Kausler, L. Hufnagel, and F. A. Hamprecht. Conservation Tracking. In *ICCV*, 2013. 2, 6, 7
- [23] B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua. Multi-commodity network flow for tracking multiple people. *TPAMI*, 2013. 1, 2, 5
- [24] P. D. Tao and L. T. H. An. Convex analysis approach to DC programming: theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997. 5
- [25] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6(2):1453, 2006. 5
- [26] C. Wählby, T. Riklin-Raviv, V. Ljosa, A. L. Conery, P. Golland, F. M. Ausubel, and A. E. Carpenter. Resolving clustered worms via probabilistic shape models. In *ISBI*, 2010. 2
- [27] F. Wang and L. J. Guibas. Supervised earth mover's distance learning and its computer vision applications. In *CVPR*, 2012. 3
- [28] C.-N. J. Yu and T. Joachims. Learning structural SVMs with latent variables. In *ICML*, 2009. 1, 2, 5
- [29] A. L. Yuille and A. Rangarajan. The concave-convex procedure. *Neural Computation*, 15(4):915–936, 2003. 5
- [30] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR*, 2008. 1, 2, 5