

指导教师： 杨涛

提交时间： 2016/3/20

# CVPR2015 Paper Translation

No: 01

姓名： 张洛铭

学号： 2013302588

班号： 10011305



## JOTS: 联合在线跟踪与分割

### 抽象

我们提出了一种新的联合在线跟踪与分割 (JOTS) 算法的多部分跟踪和分割集成到一个统一的能源优化框架来处理视频分割任务。多部分分割, 根据所估计的零件模型构成与正规化一个像素级标签分配的任务, 并且跟踪被配制为根据像素的标签, 而这又是用于改进该模型估计部的模型。多部分跟踪和分割进行了反复, 以尽量减少对 SegTrack 和 SegTrack V2 数据库的 RANSAC 式 approach. Extensive 实验所提出的目标函数表明, 该算法对国家的最先进的方法进行顺利

### 1. 简介

在视频分割目标的最新进展, 提取主要利用离线方法从准确的边界背景的目标对象 [24, 13, 21, 16, 26, 19, 27, 20]。尽管许多证实的成功, 现有的方法是用于意味着在线处理应用不太有效。例子比比皆是, 包括视频监控, 动作识别和人机交互, 仅举几例。近来, 提出了一些在线视频分割方法, 例如, [7] 和 SPT [18]。在没有强大的本地制约全局对象的外观造型 [7] 和 SPT [18] 目标无关的建议生成步骤可能会导致不准确的分割结果, 特别是在复杂的背景或大型动作场面。

图像分割旨在分区基础上在空间域中的某些特征 (例如, 颜色, 强度, 或纹理) 个像素, 同时跟踪打算分割的基础上在时间域中的一致性属性像素。显然, 每个任务有利于其他特别为网上视频分割, 并且这两个模块应该考虑在相同的框架。因此, 在这项

工作中, 我们提出了一个联合在线跟踪与分割 (JOTS) 算法, 制定了视频分割的任务, 因为在单田间能量函数网上多部分跟踪和分割。联机多部分跟踪提供了用于分割有效顺序运动和结构的限制, 而多部分分割生成准确的本地外观和位置信息, 以促进跟踪。跟踪和分割阶段通过 RANSAC 式的方法来验证各模块的精确的性能所产生的结果迭代优化。所提出的算法的主要步骤显示于图 1。如图 1 (c) 一种示例中, 目标多部分模型中的前两个迭代学习失败, 以适应当前的像素井 (某些型号的中心位置是不位于具有相同标签的像素) 的质心。作为多零件模型拟合较好的像素通过跟踪, 标签误差减小。同时, 作为标记误差减小, 实现更精确的多部分跟踪。

这项工作的贡献总结如下。首先, 一个新的联合在线跟踪和分割算法在线视频分割, 其中该多部分跟踪和分割都集成在一个统一的能量目标函数来实现更好的性能。其次, 提出了能量函数的最小化通过与  $\alpha$ -扩展算法 RANSAC 式的方法是有效解决。两个基准数据集, 即 SegTrack 和 SegTrackv2 三, 广泛的实验, 对国家的最先进的进行了证明我们的方法的有效性

### 2. 相关工作和问题背景

**分割。** 视频分割备受关注, 因为它在视觉问题的重要性。许多算法已经提出了用过去和未来批处理 [24, 13, 21, 16, 26, 19, 27, 20] 的图像序列的帧来解决这个问题。虽然这些方法产生令人鼓舞的结果, 他们并不适用于网上视觉任务, 如视频监控, 动作识别和人机交互。

在 [7] 中, 在线视频分割的综合概

率模型被提出，它结合了动态隐形状，拓扑形状的限制，自适应外观模型，和分层流动。因为只有没有本地约束像素级信息被用于在前景和背景类区分，它很可能包括误报特别是对于具有杂乱背景的视频。Li 等人。[18]通过关联的连续帧中物体的提案池提出了一种无监督的视频分割方法。然而，目标无关的建议可能导致不准确的分割结果。

**跟踪。**几个跟踪逐分割方法[9, 12, 11, 28]已经开发了在文献中。在[9]中，水平集制剂呈现准确地提取对象的边界进行跟踪。Godec 等。[12]延伸的霍夫森林分类到在线设定和集成基于投票检测和 GrabCut 分割[23]用于跟踪的方法。在[11]，Duffner 和加西亚提出基于霍夫一个快速自适应方法与基于像素的描述符和分割（类似于[2]），以处理非刚性变形为对象跟踪变换。Zhong 等。[28]与跟踪，以减轻偏最小二乘回归分析，基于结构的标签信息的漂流问题整合分割。前述方法集中在使用分割技术，以帮助跟踪，而不是从背景准确提取目标。

除了跟踪逐分割方法，是根据部分跟踪方法[22, 25, 6, 15]已经在近年提出。然而，根据某些图片属性这些方法主要生成区域，而不是集中在感兴趣的对象，因此，边界是不准确（即，某些区域包含前景和背景像素）。一种方法是段中的每一帧划分成超像素和使用条件随机场算法来分离前景和背景区域[22]。跟踪过程是基于两个类颜色分布的匹配。Wang 等人。[25]提出了一种基于在其中属于前景超像素的概率用于目标从背景中分离超像素一个判别外观模型。在[6]，Cai 等设计一个动态图形为基础的方法来解释非刚性运动。Hong 等。[15]提出了一种层次化的外观表示模型的基础上，它利用多个量化等级，包括像素，超像素共享信息的图形模式追踪，和边框。上述方法集中在跟

踪非刚性运动目标与分割的帮助。作为基于超像素粗中级分割使用，而不是从低级别的像素的细节，所产生的对象的边界是不准确。

最近，基于超像素的时间相关的跟踪方法[8]显影。与此相反，所提出的 JOTS 算法着重于从像素级的背景准确分割目标对象。

### 3. 问题描述

给出简单的用户注释随后在第一帧中的交互式分割方法[14]，我们第一段从背景的目标对象，然后使用在 SLIC 算法[1]，以产生初始虚拟模型，如在图 1 中描绘(一)(另见第 5.3 节)。

设  $\mathcal{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_k\}$  是一个目标对象的带标签集第  $k$  份  $\{1, \dots, k\}$ ，其中  $\mathcal{M}_i = (A_i, P_i, \Theta_i)$  是第  $i$  个模型， $A_i$  是模型的 HSV 直方图  $P_i$  为模型的中心位置， $\Theta_i$  是位置设置属于模型中的像素。我们还构建一个模型  $\mathcal{M}' = \{\mathcal{M}'_1, \dots, \mathcal{M}'_n\}$  从复杂的背景与标签  $b_0$  描述的像素作为异常值，其中  $n$  是在背景模型部件的数量， $\mathcal{M}'_i = (A'_i, P'_i, \Theta'_i)$  是第  $i$  部分， $A'_i$  是相应的 HSV 直方图， $P'_i$  为各个中心的位置，并且  $\Theta'_i$  位置集属于所述模型的像素。

从以前的工作中，目标对象的基础上给定的目标模型（例如，外观或位置的模型），其段不同，在此工作的视频分割任务是由多部分跟踪和分割在一个框架配制。也就是说，我们优化像素同时标记  $f$  和所述目标多部分模型  $\mathcal{M}$ 。对于图像中的每个像素  $p$ ，我们用标签  $f_p \in \{b_0\} \cup \{1, \dots, k\}$  表示它属于（即，多部分分割），其一部分，而不是仅仅确定给它分配它作为在以前的方法的前景或背景，以及在当前

帧（即，多部分跟踪）同时优化所述目标多部分模式  $\mathcal{M}$ 。获得的每个像素的标签分配，视频分割后任务为每个帧被自然完成。

为了降低计算复杂性，我们只在当前帧，这是由前一多部分的模型  $\mathcal{M}$  的视频分割问题被配制如下确定分配的标签的预测目标位置附近的像素：

$$\{\mathcal{M}^*, f^*\} = \underset{\mathcal{M}, f}{\operatorname{argmin}} E(\mathcal{M}, f | \overline{\mathcal{M}}),$$

其中  $E(\mathcal{M}, f | \overline{\mathcal{M}})$  被从两个多部分跟踪和分割集成到一个统一的能量函数的分段阶段获得。  $\mathcal{M}^*$  和  $f^*$  都是在当前帧中的多部分模型和像素标记结果，分别。为了解决 (1)，我们首先使用动态结构图匹配方法[6]具体地，我们首先使用 SLIC 算法[1]，以产生在当前帧的多个候选部获得预测的目标位置，并使用频谱匹配技术[17]发现在先前的目标模型中的部件和候选部分之间的匹配，如图 1 (b) 中。该粗略估计目标位置是基于部分匹配的选票计算。最后，我们设置在含有  $\eta$  比在先前的分割区域中的目标大小（根据经验设定为 1.8）的粗略估计目标位置为中心的边界框。

## 4. 联合在线跟踪与分割

### 4.1 数据项.

在该分割区域中，最佳的标签  $f^*$  对于每个像素和在当前帧中的最佳多部分模型  $\mathcal{M}^*$  是通过最小化获得的建议的能量  $E(\mathcal{M}, f | \overline{\mathcal{M}})$  来 RANSAC 样式的方法来获得在两个步骤中的溶液：1。像素标签被分配与由  $\alpha$ -扩展算法的当前估计多部分模型[4]；2. 目标部分根据像素的外观的可能性和运动的一致性与当前标记跟踪。这两个步骤进

行迭代，直到达到目标函数的最小能量，使得多部分跟踪便于多部分分割，反之亦然。迭代优化后，我们更新基于最优标签和输出最终分割结果（参见图 1 (d)）的多部分模式。所提出的优化过程在以下部分中描述。

两个方面，即，目标外观和位置，时空连续性提供对在线视频分割任务有效的信息。如果像素  $P$  被标记为李，我们预计像素属于两个方面部分型号  $\mathcal{M}_i$  的小能量。在 (2) 中的数据项被定义为

$$\begin{aligned} D(f, \mathcal{M}) &= \sum_{p \in S} D_p(l_i, \mathcal{M}_i) \\ &= \sum_{p \in S} (\alpha_1 \cdot \Phi_a(\rho_p; \mathcal{A}_i) + \alpha_2 \cdot \Phi_l(\ell_p; \mathcal{P}_i)), \end{aligned} \quad (3)$$

其中  $D_p(l_i, \mathcal{M}_i)$  是标记为立的像素  $p$  的数据的能量， $S$  是在该分割区域中， $\alpha_1$ ,  $\alpha_2$  设定像素中的权重参数，

$\Phi_a(\rho_p; \mathcal{A}_i)$  和  $\Phi_l(\ell_p; \mathcal{P}_i)$  是基于外观和位置，分别能量项。在这种制剂中，

$\Phi_a(\rho_p; \mathcal{A}_i)$  是由属于模型  $\mathcal{M}_i$  的，其被计算为在 HSV 直方图艾的像素的单元条值的像素  $p$  的外观的可能性引起的外观的能量。另外， $\Phi_l(\ell_p; \mathcal{P}_i)$  是通过基于从

中心位置  $\mathcal{P}_i$  的由在垂直和水平两个单高斯模型的产品计算的位移的像素  $p$  的位置的可能性引起的位置的能量。

同样地，属于背景的像素的数据的能量（即，异常）由外观描述以及位置，并限定所有的背景子模型的最小能量，即  $D_p(f_p, \mathcal{M}^*) = \min_j D_p(f_p, \mathcal{M}_j^*)$ 。

### 4.2. 长期平稳

直观地说，如果两个相邻的像素具有相似的外观，相同的标号被分配给它们与小的能量。另一方面，从背景目

标的运动通常是不同的，特别是在对象边界（鼻内，运动的不连续性）。这些外观和运动线索提供区分目标对象和背景像素的有效信息。基于这些因素，在光滑的术语  $V_{p,q}(f_p, f_q)$  在 (2) 被定义为

$$V_{p,q}(f_p, f_q) = \mathbb{I}(f_p \neq f_q) \cdot (\alpha_3 \cdot \Delta_c(p, q) + \alpha_4 \cdot \Delta_f(p, q))$$

在那里我将返回  $\mathbb{I}(\cdot)$ ，如果它的参数是正确的，否则为 0，和  $\Delta_c(p, q)$  和  $\Delta_f(p, q)$  是之间的欧氏距离的两个相邻的像素  $p$  和  $q$  中的 RGB 颜色空间和分别光流场 [5]。在上面的配方， $\alpha_3$ ， $\alpha_4$  是权重参数。

### 4.3. 正则项

为了避免过度拟合问题 [10, 3]，我们正规化的数据项。在正则项 (2) 包括三个因素：1. 面积：它鼓励所有具有类似尺寸的型号；2. 简介：它惩罚像素标签不完整的和不规则的模型；3. 复杂性：它惩罚的像素标注的型号数量。

$$H(f, \mathcal{M}) = \sum_{i=1}^k \mathbb{I}(\exists p : f_p = l_i) \cdot (\alpha_5 \cdot H_a(f, \mathcal{M}_i) + \alpha_6 \cdot H_p(f, \mathcal{M}_i) + \alpha_7 \cdot H_c(f, \mathcal{M}_i)),$$

其中  $\mathbb{I}(\cdot)$  返回 1，如果它的参数为 true，否则为 0， $H_a(f, \mathcal{M}_i)$ ， $H_p(f, \mathcal{M}_i)$ ， $H_c(f, \mathcal{M}_i)$  的领域，亲文件，分别零件模型弥的复杂性正规化条款。在这个公式中， $\alpha_5$ ， $\alpha_6$ ， $\alpha_7$  和是相应的权重。

对于所有的背景子模型，我们设定  $H_0$ ，其中  $i=1, \dots, N$ 。这些正规化术语描述如下。

**区。**大面积的模型不处理大型物体变形良好。另一方面，具有小面积的模型是易感于背景噪声。因此，我们定义该区域正规化术语

$$H_a(f, \mathcal{M}_i) = |\Theta_i| - \frac{1}{k} \sum_{j=1}^k |\Theta_j|,$$

$|\Theta_j|$  表示在  $\mathcal{M}_j$  像素的数量。

**简介。**当大的变形发生一些目标部分可以包括，对象范围和中心位置，可能无法精确地估计。为了鼓励新车型，抑制那些不准确的，我们定义亲文件正规化任期

$$H_p(f, \mathcal{M}_i) = \Delta_p(\mathcal{P}_i, \mathcal{C}_i) \cdot \Omega(\forall p \in B_i, \Delta_p(\ell_p, \mathcal{C}_i)), \quad (7)$$

其中  $\mathcal{P}_i$  和  $\mathcal{C}_i$  是零件模型和弥分别标记为  $l_i$ ，像素构成的区域的中心位置。

$\Delta_p(\cdot, \cdot)$ ：

是欧几里德距离函数来计算 2D 图像中的两点之间的距离平面。此外， $\ell_p$  是像素  $p$  的位置，和  $B_i$  是边界像素集由标记为  $l_i$  的像素构成的区域的。  $\Omega(\cdot)$  是方差函数来计算边界像素之间的距离的方差

**复杂。**恒定的复杂性正则项来惩罚与大量目标的标记结果模型，即

$$H_c(f, \mathcal{M}_i) = 1.$$

## 5. 能量最小化

在 (2) 是具有挑战性的目标函数的能量最小化问题涉及两组变量。我们优化  $F$  和  $\mathcal{M}$

可选地以类似于精神 [10, 3] 与多部分跟踪和多部分分割

初始多部分模型  $\mathcal{M}[0]$  是从先前帧的最优模型中获得。显然是不正确的

模型可以包含在  $M[0]$ 。在多部分割阶段，像素标记  $F[0]$  通过与正则化项，并在  $M$  中一小部分的可靠的模型  $[0]$  被选择的  $\alpha$ -扩张算法来计算。在多部的跟踪阶段，所选择的模型由用能量函数 (9) 重新估算 HSV 直方图和位置的模型改进。接下来，我们添加基于当前标签的一些假设的部分模式拓展多部分型号  $M[1]$ 。重复这两个阶段以产生一系列标号的  $F[0]$  中， $f[1]$  中， $f[2]$ ， $\dots$  而且多部分模型组  $M[0]$ ， $M[1]$ ， $M[2]$ ， $\dots$ ，直到在 (2) 不再减小的能量 (参见图 1 的 (c))。因此，能量被最小化，以获得最佳的标记  $F^*$  和多部分模型  $M^*$

能量函数  $E(M, F)$  是非负与天然下界 0 同时，能量是非超过迭代增加，以确保这个优化过程的收敛。我们提出在图 2 中的一些例子，以显示能量如何迭代减小。同时，作为整体的能量降低时，相应的交叉点，过联合重叠分数的 JOTS 算法的增加，这表明我们的能量最小化的方法的有效性。多部分跟踪和分割模块下进行了描述。

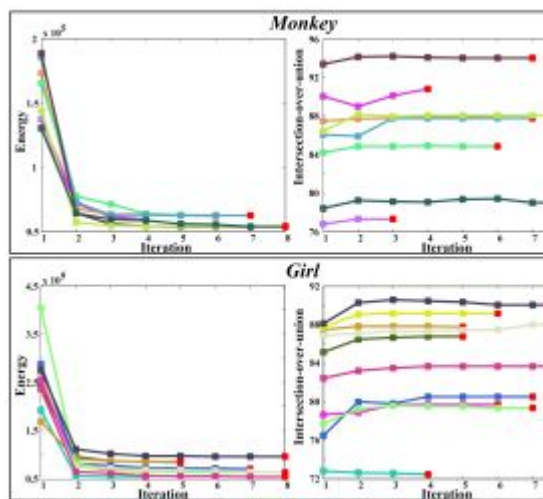


图 2. 左起：能源曲线和分割结果的基础上在猴子和女孩序列的几个样本帧的迭代优化过程中的交叉点，在工会的重叠度。不同的帧是由不同颜色的

说明。每条曲线上的红色方块表示各反复处理的结束。

### 5.1. 多声部分割

从背景段的多个靶部分，我们分配在与多个标签  $\{L_0, L_1$  的分割区域中的像素  $p$ 。  $\dots$ ， $L_K\}$ ，而不是简单地其列为在以前的方法的前景或背景。像素标注问题转化为两两马尔科夫随机场的能量最小化，

$$f^* = \operatorname{argmin}_f D(f, \mathcal{M}) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(f_p, f_q) + H(f, \mathcal{M}), \quad (8)$$

其中  $D(F, M)$  是一个基于标签  $f$  和多零件模型男，副总裁， $Q(FP, FQ)$  中的数据项是描述长期平稳相邻像素之间的相互作用， $N$  是像素间的 4-邻里关系在  $S$  的优化问题可以通过  $\alpha$ -扩展算法 [4] 与图由于能量函数有效地切割要解决保持子模块。

### 5.2. 多声部跟踪

一旦在分割区域的像素标志  $F$  的计算，我们重新估算多零件模型  $M = \{M_1, \dots, M_K\}$  通过最小化能量  $E(M, F)$ 。给出当前像素标签楼中的平滑项 (2) 被固定，并且所述多部分的跟踪问题被配制成

$$M^* = \operatorname{argmin}_{\mathcal{M}} D(f, \mathcal{M}) + H(f, \mathcal{M}), \quad (9)$$

它是具有挑战性的解决 (9)，因为正则项是难以相对于所述多部分模型最小化  $M$  [3]，我们忽略先正则项，并专注于减少 (9) 使用最大似然估计 (MLE) 方法，以获得最佳的模型  $M^*$  的第一项类似。也就是说，对于在标记  $F$  中的第  $i$  个选择的模型  $M_i$ ，我们估计 HSV 直方图  $A_i$ ，中心位置  $P_i$ ，并基于标记为李当前观察象素设置  $\Theta_i$  像素位置。然后，如果 (2) 的总能量减少，我们使用所估计的模型 ( $A_i, P_i$ ，

$\Theta$  bi) 取代 $\Theta$ ; 否则, 零件模型弥被保留。

这种优化方案是由两个因素动机: 1. 简化最小化 (下降的正则化项) 被有效地由最大似然估计方法来解决, 同时保持的总能量不增加; 2. 最小化的简化应该对完整能量最小化方案微不足道的影响。即, 如果当前的解决方案是靠近好最小值, 正则项  $\partial \partial MH(F, M)$  的梯度将是小的, 因为该解决方案已服从在第 4.3 节中讨论的约束。否则, 一个大的梯度  $\partial \partial MH(F, M)$  表示存在另一种模式是更合理的约束。该模型将通过下面的步骤来优化虚拟模型 (在下节中描述) 被拾起。因此, 我们暂且能量最小化到后来的多部分分割优化阶段的困难的方面。

### 5.3. 扩展·虚拟零件模型

**生成初始假设模型。**在第一帧, 在最初的目标区域由 SLICE 算法生成多个部分 [1]。如果所产生的部分和用户注解目标区域之间的重叠率低于阈值  $\theta$  1 大 (例如, 在此工作 0.5), 我们将其添加到生成的初始部分的模型和否则考虑其作为背景的一部分。

**炼油假设模型。**要获得从分割现有的更好的零件模型, 我们用两个标准合并和拆分区: 1. 只有邻近的具有类似外观的小区域是随机选择产生新的车型。的区域被认为是小, 如果像素的数量小于所述平均数量的当前多部分模型的像素。2. 面积比目前使用的部分模型的平均面积的两倍大一个标记区域被 SLICE 算法分成多个的。

## 6. 实验

我们评估两个视频分割基准数据库, 即 SigTrack [24] 和 SegTrack V2 [18] 数据库的算法。如 [18] 所讨论

的, 在不同大小的物体的像素的错误有很大的不同。此外, 像素误差度量是手动注释的错误敏感。为公平和全面比较, 对每帧的平均像素误差的结果报告于原始 SegTrack 数据库并将结果

对定量评价几个国家的最先进的方法 [16, 13, 27, 18, 24, 12, 7, 25, 6] 在表 1 和表 2 前两名的执行方法都用红色显示和蓝色, 分别。一些分割结果示于图 3 中。

### 6.1. 实施细则

在每个实验中, 用简单的用户注释后跟分割方法 [14] 和在 SLIC 算法生成的初始假设模型 [1] 为我们的方法 (参见图 1(a))。所有实验都具有 2.9 GHz 的英特尔 i7 处理器和 16 GB 内存的计算机上进行。所提出的 JOTS 算法的运行时间复杂性依赖于目标对象的大小。例如, 在 MATLAB 无代码优化实现的, 它需要每帧约 20 秒到段中含有  $320 \times 240$  像素的图像的 Monkeydog 猴序列的靶。建议 JOTS 算法的源代码将被公布于众。

对于每一个序列, 超像素为 SLIC 算法 [1] 在初始化的数目根据目标的大小设定。根据经验, 在 JOTS 算法执行以及当两个目标和背景各部分模型包含大约 50 至 200 像素。在 JOTS 算法的所有其他参数固定在所有实验中。我们使用 6 箱的 HSV 直方图的每个信道来描述一个目标对象。对于预设重量的参数, 我们采取了以下默认值:  $\alpha_1=2.0$ ,  $\alpha_2=1.2$ ;  $\alpha_3=3.0$ ,  $\alpha_4=3.0$ ;  $\alpha_5=20$ ,  $\alpha_6=20$ ,  $\alpha_7=10$ 。它需要 2 至 8 次迭代来解决 (8) (也参见图 2)。

### 6.2. 数据库

SigTrack 数据库。该 SigTrack

数据库[24]包括 6 具有挑战性的视频 (Birdfall, 猎豹, 女孩, Monkeydog, 降落伞, 和企鹅) 与像素级的人注解的分割结果的主要前景物体。它包括多个交互的对象 (猎豹和企鹅), 突然运动 (Monkeydog 和 Birdfall), 复杂的变形 (女孩和 Monkeydog), 和外观的变化 (降落伞)。

SegTrack V2 数据库。这个数据库是数据库 SegTrack 更多注释对象和 8 个新的视频序列的扩展包括: 天堂, 小轮车, 漂移, 蜂鸟, 猴子, 青蛙, 蠕虫, 和孙中山军人的鸟。有 24 个对象超过 947 批注框在这个数据库中包括不同挑战性的因素视频分割, 包括多个交互的对象 (猎豹, 漂移和企鹅), 外观变化 (天堂鸟和漂移), 遮挡 (猎豹, 小轮车 14 的序列和漂移), 和复杂的变形 (BMX-人, 蜂鸟, 青蛙, 孙中山军人, 猴, 和蠕虫)。

### 6.3. 定量比较

表 1 显示了该算法和国家的最先进的视频分割方法的定量结果 [16, 27, 18, 24, 12, 7, 25, 6]。总体而言, 建议 JOTS 算法对毫不逊色使用像素误差度量 SegTrack 数据库上大多数在线和离线方式进行。此外, 表 2 表明 JOTS 算法以及对其他方法进行 [18, 16, 13, 12, 25, 6] 使用的交点-过联合重叠度量 SegTrack v2 的数据库 1。详细分析和定量评价的讨论接下来介绍。

多个交互的对象。在企鹅和猎豹序列的目标有相似的外观, 邻近的对象。对于离线方法 SPT+ CSI [18] 和 [16, 13, 27, 24], 都在双方的过去和未来帧可用的对象建议用于分割。不可避免地, 不准确的对象的建议, 并在这些方法不对关联容易导致较大的分割误差, 特别是在正确的目标是由多个相邻/相互作用具有类似外观的物体所包围, 如表 2 所示同样的问题也存

在该例网上的方法 SPT [18], 它依赖于对象的建议。与此相反, JOTS 算法在这些序列表现良好。这可以归因于该 JOTS 算法利用两个目标部分及其相邻的背景, 这有助于区分这些区时它们是相似的时间一致性。

突然的运动。在 Birdfall 和 MonkeydogMonkey 序列的目标, 表现出快速, 突然的运动。在 JOTS 算法高斯选址模型可能无法从目标的前一帧中心位置处理大排量。在这种情况下, 目标的粗中心位置由光谱匹配法的分割前的预测 (参见图 1 (b)), 这使得 JOTS 算法来处理的突然运动的挑战。与此相反, 该方法 [7] 仅由光流, 这使得它不能处理在保守帧目标的大的位移, 并且可以通过在复杂场景的背景噪声, 例如, 对象可以容易受跟踪目标在 Birdfall 序列。

复杂的变形。它是具有挑战性的段中的女孩, 猴狗, 猴, 蜂鸟, 青蛙, 蠕虫, 孙中山军人, 猴和 BMX 人称序列由于大变形的非刚性对象。尽管这些序列包含复杂对象的变形或杂乱背景的挑战, JOTS 算法基于部分表示是能够有效处理此类案件。值得一提的是, 在线超级基于像素的跟踪方法, 例如, [25] 和 [6], 不反对 JOTS 算法这些序列表现良好 (见表 1 和表 2)。该超像素在通过这些方法每一帧独立生成。因此, 一些超像素的同时包含前景和背景像素, 并分割结果是不准确。另外, 对 HT 方法 [12] 也不会很好, 主要执行由于无需在分割有效局部约束 (即, 从前景和背景像素容易混淆仅基于全局外观信息) 的基于像素的表示。随着像素级的分割和部分级跟踪的统一迭代优化配方相结合, 既表示相互促进对我们的方法准确的分割和跟踪结果

外观变化。大的外观变化的挑战发生在降落伞, 天堂鸟和漂移序列。所提出的 JOTS 算法针对其他视频分割方法 [13, 18, 16] 在这些序列表现良



好（见表 2）。该算法[13]直接聚合从前景和背景都超像素，而不考虑所述目标对象具体地说，因而分割结果是不准确。Incontrast, 所述 JOTS 算法在一个统一的能量最小化制剂集成目标部件状态估计（从多部分跟踪）和像素标记（从多部分分割），使得更精确的分割结果所用的 RANSAC 式迭代优化来产生处理（见图 2）。网上的方法 SPT [18]和离线方法 SPT + CSI [18]和[16]处理通过精炼和连续帧中的对象的提议相关联的目标的外观的变化。与此相反，JOTS 算法以及在通过优化过程中选择好的零件模型（如在第 5.3 节中讨论）在线处理的目标外观变化进行。

闭塞。如表 2 所示，当目标对象被遮挡提议 JOTS 算法表现良好（例如，在猎豹，漂移和企鹅序列），其可以由 RANSAC 式优化方法，用于选择一小部分的可靠对象模型来解释以产生更精确的结果（如在第 5 部分讨论）。但是，当目标对象中的 BMX-自行车序列的第一帧经历重闭塞，无监督方法 SPT[18]，SPT + CSI[18]和[16, 13]针对我们的方法表现良好。因为闭塞部件被包含在对象的建议，所有这些方法都能够在目标的未被遮挡和闭塞部件在 BMX-自行车序列相关联以实现更好的性能。然而，我们的方法考虑连续帧来分割对象之间仅时空一致性线上。因此，难以对所提出的 JOTS 算法。图 4. 基于路口环比工会重叠度在 SegTrack 数据库中的零件模型大小的敏感性分析。两个红色虚线之间的区域为初始超像素的数量的建议值范围。识别单独的闭塞部件，因为它们的信息不是在第一帧可用。总的来说，我们的方法以及在咬合时发生了挑战这两个数据库执行。

#### 6.4. 讨论

零件模型大小的敏感性分析。我

们研究初始超像素的数目的在 - 注量在建议 JOTS 算法，这样就解决初始部分模型的大小。如在图 4 中，该 JOTS 算法相对健壮到初始超像素数的小扰动。具体地说，我们注意到 JOTS 算法的性能提高到达到稳定状态的硬质的目标对象，例如，降落伞，企鹅，和 Birdfall 序列超像素的数量增加。而对于非刚性的物体，例如，CheetahDeer，女孩，和 Monkeydog 猴序列中，JOTS 算法实现相对较好的结果仅在该区域之间的两个红色虚线。显然，刚性目标的几何结构是相对稳定的。因此，当超像素的数目大于某一值时，JOTS 算法能够获得目标的足够的本地信息，以产生期望 results. Meanwhile, 对于非刚性目标对象，它们的几何结构显著变化，这削弱的辨别力零件模型的位置方面。因此，如果超像素的数目过少，则零件模型不太判别以输出令人满意的结果。另一方面，如果过大，则部分车型将小并会被删除目标的运动，这也导致了糟糕的表现时的噪音。

正则项的有效性。为了证明在所提出的正则化项的有效性 JOTS 算法，我们与两个基线法，即比较一下，“JOTS-COM”和“W / 0 转正”的 SegTrack 数据库使用的交集，在工会的重叠度。了“JOTS 泡沫”的方法只使用了复杂性正则项中的正则化项 (5)，即，

$$h_f(\mathcal{M}) = \sum_{i=1}^k \mathbb{I}(\exists p : f_p = i) \cdot (\alpha_T \cdot h_f^3(\mathcal{M}_i)),$$

和“W / 0 型正”的方法不使用正则项中的能量目标 (2)，即，高频 (M) = 0。如图 5 所示，JOTS 算法的性能比在所有序列都基线方法更好，这表明并设计正则化项的有效性。(6) 正则项不仅考虑的车型数量（即，复杂性正则项）的限制删除一些无用的机型，同时也是模型的尺寸和规律性的约束条件（即，面积和资料正则项），以改善性能（如第 4.3 讨论）。

多部分的跟踪和分割的集成。为

了证明多部分的跟踪和分割的整合在一个统一的目标函数的有效性，我们构造两个基本方法，即“W / O 重复”和“W / O 求精”。的“W / O 型的迭代”方法指示多部分跟踪和分割不迭代优化，“W / O 型精”的方式表示该模型精制工序（参见第 5.3 节），不包含在该迭代过程。图 5 示出了 JOTS 算法优于在所有序列都基线方法。如果没有迭代的过程中，多部分跟踪和分割模块不互相帮助，因此结果不准确。未经精制推测模型，多好的模型不能被添加，从而降低了结果的准确性。

## 7. 结论

在本文中，基于多零件模型的联合在线跟踪和分割算法在线视频分割。两个多部分分割为像素标记和跟踪作为零件模型估计过程都集成在一个统一的能量最小化制剂，其有效地通过与  $\alpha$ -扩展算法一个 RANSAC 样式的方法解决。此外，多个约束集成正规化像素标签和零件模型估计。两个基准数据库广泛的实验结果表明，该算法对国家的最先进的方法进行视频分割的有效性。

## 感谢

龙吟文，震雷斯坦李正东是由中国项目国家自然科学基金项目（No. 61203267，No. 61375037，No. 61473291），国家科技支撑计划项目（No. 2013BAK02B01），中国社科院支持科学项目（编号 KGZD-EW-102-2），和正宗 R&D 经费。杜大卫是由中国项目国家自然科学基金项目（No. 61472388）的支持。MingHsuan 杨在由国家科学基金会职业格兰特（No. 1149783）和美国国家科学基金会资助的 IIS（No. 1152576）的部分支持。