

指导教师： 杨涛

提交时间：2016/3/15

CVPR2015 Paper

Translation

No: 01

姓名： 段文章

学号： 2013302498

班号： 10011302



基于测量突出特征的视频对象分割

Wenguan Wang

jianbing Shen

Fatih Porikli

摘要

我们引入一个无监督, 基于测地距离的, 突出的视频对象分割方法。与传统的方法不同, 我们的方法结合了突出作为对象之前通过强劲的计算测量。我们考虑两个区别的视觉特点: 运动空间边缘和时间界限指标的前景对象的位置。首先, 我们从这些指标生成时空显著地图使用测地距离。建筑的观察前景区域包围区域高时空边缘值, 测地线距离提供了一个初始估计前景和背景。然后, 通过测地生产高质量的显著结果距离在后续帧背景区域。通过结果显著地图, 我们为前景和背景构建全局外观模型。通过实施运动连续性, 我们建立每一帧的动态选址模型。最后, 时空显著地图, 外观模型和动态位置模型组合成一个能量最小化框架实现时空上都一致的对象分割。在基准数据集视频演示中, 广泛的定量和定性实验显出了的优越性, 该方法在最先进的算法。

一. 引言

针对无人监督的视频对象分割方法

从整个视频自动提取对象。这种分割显示许多特定的视觉中获益任务和应用程序, 比如视频摘要, 压缩和人机交互等一些。外观和运动线索通常的信息采用视频分割方法。一些方法在 [6, 17, 12] 通过多个帧的点轨迹来分析利用可得到的运动信息。[6] 为基于轨迹提供了一个视频分割的框架通过对轨迹之间的关系建立关联矩阵。[17] 分组像素与连贯的运动通过远程计算运动向量从过去和未来的帧。[12] 发现不连续相邻空间轨迹之间的嵌入密度。随着工作的 [15] 指出, 这些基于轨迹的技术与跟踪相关的挑战 (漂移, 闭塞和初始化) 和集群 (模型选择和计算复杂度) 和缺乏之前信息对于一个成功的对象分割。一些努力 [5, 26, 30] 提出有效的自下而上的优化框架最后分割采用外观和运动线索。最近, 几个方法 (15、19、32) 探讨了前景对象应该是什么样子的视频数据概念。建议这些方法 [11 8] 在每一帧和转换视频的任务对象分割成一个对象区域选择问题中产生相当大的对象。在这一选择过程, 运动和外观结合测量的对象的信息。更具体地说, 一个聚类过程就是寻找对象。[15] 一个约束最大重量派

系技术选择模型马英九提出的过程 [19], 和分层提出了基于有向无环图的框架 [32]。然而, 这些提案的基础技术计算复杂度高, 他们的依赖性大量的建议导致了困难和选择过程的复杂性。我们的目标是段的前景对象从背景帧的视频序列任何用户注释和语义之前。我们的方法是基于提出的视觉特点检测技术包含一些视觉线索如运动边界, 边缘和颜色。对象和背景估计由我们的方法生成提供一致的和可靠的先知先觉为更高层次的对象分割任务。这个主题是少了, 主要是由于只有少数特别的方法为视频设计的特点。

这些特点的方法

[14, 20, 28, 26, 13, 21], 然而, 通常建立他们系统的简单组合现有的图像特点模型与运动线索。此外, 这些方法的性能不够好指导分割。我们的方法正确估计对象的位置和背景地图并获得统一的特点。在另一方面, 我们的视频对象分割算法基于测地距离, 这已经被证明交互式图像分割是有效的与用户刷 [3, 25, , 2, 10]。然而, 在许多视觉应用程序, 比如处理大量的视频数据, 它通常是乏味的和不切实际的用户处理手动的视频帧。在本文中, 我们试图引进测地距离完全自动分割框架, 它

与以往的方法是不同的是 [3, 25, 2, 10] 需要仔细的用户帮助。

二. 我们的方法

图 1 展示了我们的方法的概述。首先, 输入框架分割像素。对于每个像素, 两种类型的边缘提取: 空间静态边缘在同一帧边界边和运动估计从相邻的帧。测地距离, 定义为两个像素之间的最短路径图像, 然后采用内部框架图计算每个像素对象的概率。基于对象的观察区域包围的地区高时空边缘值, 对象的概率计算测地线距离最短的框架边界。使用自适应阈值获得初始框架的标记为背景和前景地区。下一个框架构造图生产时空特点映射的计算测地线距离估计背景区域的两个相邻帧。最后, 实现精估计的前景, 全球前景和外观模型背景建立了卓越的结果。动态位置模型估计每一帧的运动信息从一些后续帧中提取。时空显著地图、全球外观模型和动态选址模型组合成最终的分割的能量函数。我们将在网上公开源代码 1。

2.1 使用时空边缘对对象的评

估。边在预测对象的边界提供良好的指导, 同时非常有效。

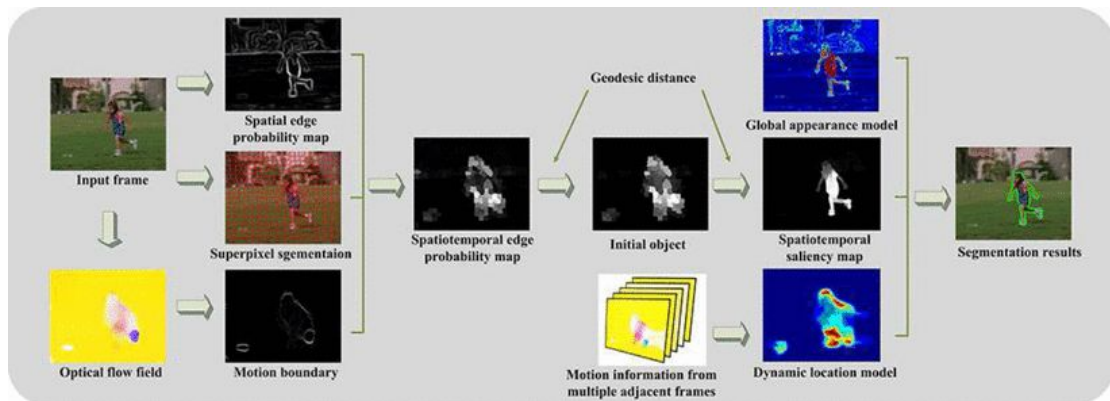


Figure 1

运动信息还提供了一个简化的但是非常有效的指标对象, 突然改变的像素从邻居经常获得更多的关注。如图 1 所示, 静态边缘的位置对单帧和两个连续帧的光流场估计为检测对象可以提供有用的信息。我们的方法基于这两个区别的功能启动对象的位置。给

定一个输入视频序列 $F =$

$\{ F_1, F_2, \dots \}$, 我们计算一个边缘

概率地图 $E_c^k(x_i^k)$ 对应于 F 帧像素

x_i^k 使用 [16]。对后续帧之间的光流是通过大位移运动估计算法 [7]。我们每一帧都使用 SLIC [1]。

$$E^k = \hat{E}_c^k \cdot \hat{E}_o^k.$$

- (1) $\mathcal{G}^k = \{V^k, \mathcal{E}^k\}$ 用像素 Y^k 和点 V^k 对节点之间的链接, 因为边缘 \mathcal{E}^k 。边缘 e_{mn}^k 的权值 W_{mn}^k 与相邻像素 Y_m^k 和 Y_n^k 的定

设计背后的直觉是, 如果前景对象同于背景的运动模式, 大型大小的梯度光流应该对象边界。此外, 静态边缘地图给老师根据空间对象的边界信息。当空间和时间不连续边缘运动通过 (1) 融合在一起, 输出时空边缘地图能够暗示前景对象的位置。这种现象可以从图 1 很容易观察到, 要么高时空对象区域边缘值, 或者是这些高概率出现的区域包围。根据这个论点, 我们选择使用的测地距离视觉显著区域从背景和测量他们的可能前景。

内部帧图像的构造 对于帧 F^k 我们构建一个无向加权图

义如下

$$e_{mn}^k = \|E^k(Y_m^k) - E^k(Y_n^k)\|, \quad (2)$$

这里 $E^k(Y_m^k)$ 和 $E^k(Y_n^k)$ 对应于时空的边界像素的概率分别 Y_m^k 和 Y_n^k 。

基于图结构, 我们推导出 $|V^k| \times |V^k|$ 是矩阵 W^k 。这里 $|V^k|$ 是 V^k 中节点的数量。 W^k 中元素 (m, n) 是 $W^k(m, n) = C_{mn}^k$ 。对于每个像素 Y_n^k 前景的概率 P_n^k 计算最短的测地线距离图像边界用

$$P_n^k = \min_{T \in T^k} d_{geo}(Y_n^k, T, G^k), \quad (3)$$

这里 T^k 表明了沿着帧 F^k 四个边界的像素, 在图像 g^k 中任意属于 v^k 的两像素 v_1, v_2 的测量距离 $d_{geo}(v_1, v_2, g^k)$ 被定义为积累的边的权值沿着最短路径图的公式

$$d_{geo}(v_1, v_2, G^k) = \min_{C_{v_1, v_2}} \sum_{p=0,1} |W^k \cdot \dot{C}_{v_1, v_2}(p)|, \quad (4)$$

这里 $C_{v_1, v_2}(p)$ 是连接两点 v_1, v_2 的距离。如果一个像素超出所需的对象, 其前景概率很小, 因为可能存在一个路径图像边界, 不通过时空的边缘值高的地区。然而, 如果一个像素是在对象, 这个像素的概率大的区域, 周围是边缘, 这就增加了测地距离图像边界。我们对于每一帧正常前景对象的概率

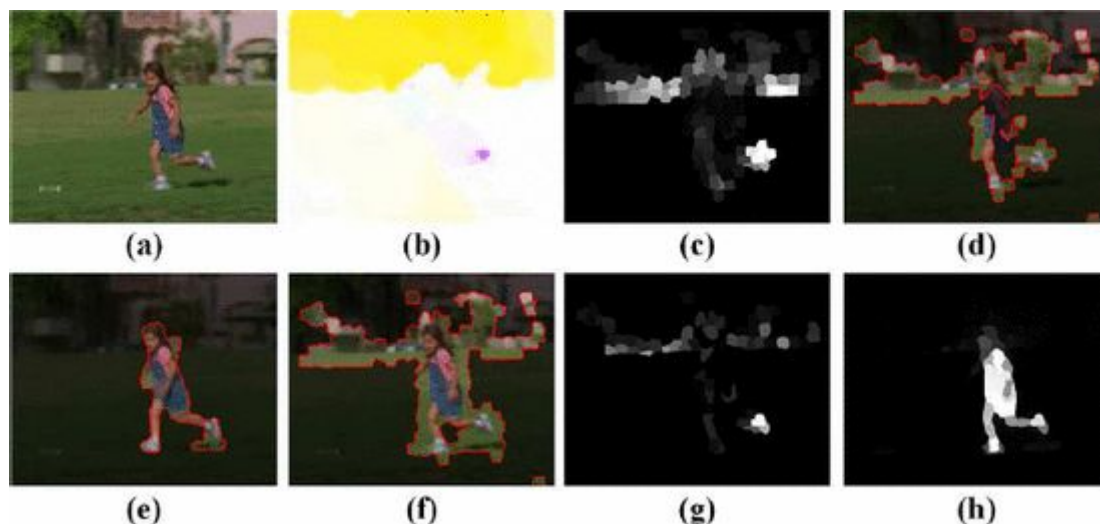
P_n^k 取值都是从零到一。对象的概率地图框架是由 b 表示我们的图很稀疏, 所有像素的最短路径是由约翰逊算法有效地计算。

2.2 时空的特点

在地图 P^k 获得的前景概率公式可以找到前景对象但不是很精确。特别是, 由于过度分割, 概率附近区域的背景对象边界是不必要的增加。此外, 错误的结果可能来自光流估计的不准确。所幸的是, 前景和背景是视觉上不同的(显著)的定义和对象是暂时连续相邻帧之间。这里提供一个方法, 利用这些信息来获取时空显著结果和处理对相邻帧之间。

内部帧的图像结构

有两种类型的边缘: 帧内边缘链接所有空间相邻像素和帧边缘连接暂时相邻像素。像素空间连接是否在同一帧相邻, 暂时参考像素的相邻像素属于不同的框架, 但沿着时间轴有重叠。我们分配的重量之间的欧几里得距离的平均颜色 CIE-Lab 颜色空间。对于每一帧, 用于自适应阈值分解框架为背景区域 B^k 和 U^k 通过对象的概率地图 P^k 对象区域。此外, 前一帧的背景信息提供了宝贵的之前, 可以消除由于不



准确的光流估计的工件。因此,我们定

义 k 的背景区域 \mathbf{B}^k 框架

为:

$$\begin{aligned} \mathbf{B}^k &= \{Y_n^k | P_n^k \leq \sigma^k\} \\ &\cup \{Y_n^k | Y_n^k \text{ is temporally connected to } \mathbf{B}^{k-1}\}, \quad (5) \\ \mathbf{U}^k &= \mathbf{Y}^k - \mathbf{B}^k, \end{aligned}$$

基于图像 g^k , 我们在帧 F^k (F^{k+1}) 获得了像素 Y_n^k (Y_n^{k+1}) 的卓越的价值 S_n^k

(P_n^{k+1}) 如下

$$S_n^k = \min_{B \in \mathbf{B}^k \cup \mathbf{B}^{k+1}} d_{geo}(Y_n^k, B, G^k). \quad (6)$$

(6) 的关系的主要原理是, 凸起一个像素的值是衡量其最短路径背景区域的颜色空间, 既考虑时空背景信息。图 2 说明了这个过程。后获得时空特点映射 S^k 和 S^{k+1} 框架 F^k 和 F^{k+1} , 我们一直

执行这个过程下两个相邻帧 F^{k+1} 和

F^{k+2} 在年底前视频序列。

2.3 时空对象分割

我们制定视频对象分割两个标签(前景和背景)作为一个像素标签问题。每

一个像素 $x_x^k \in X^k$ 可以有一个从零到一

的标签 l_i^k , 这里 0 对应于背景和 1 对

应于前景。在所有的帧中像素的一个

标签代表视频的一个分割。类似于其他细分工作[15], [27], 我们定义了所

有像素的标签能量函数公式:

$$\begin{aligned} \mathcal{F}(\mathbf{L}) &= \sum_{k,i} \mathcal{U}_i^k(l_i^k) + \lambda_1 \sum_{k,i} \mathcal{A}_i^k(l_i^k) + \lambda_2 \sum_{k,i} \mathcal{L}_i^k(l_i^k) \\ &\quad + \lambda_3 \sum_{(i,j) \in \mathbf{N}_s} \mathcal{V}_{ij}^k(l_i^k, l_j^k) + \lambda_4 \sum_{(i,j) \in \mathbf{N}_t} \mathcal{W}_{ij}^k(l_i^k, l_j^{k+1}), \quad (7) \end{aligned}$$

在空间像素邻域公式由在一个帧的八

个空间相邻像素, 颞像素邻域公式由

前方后方九个邻居在相邻帧, 和公式

索引的像素。这个能量函数包含三个

一元 \mathbf{U}^k \mathbf{A}^k 和 \mathbf{L}^k 还有连个成对的方面

\mathbf{V}^k 和 \mathbf{W}^k 这取决于时空上的标签相邻

的像素。标量参数 λ 在各种条件下的

权重。在我们的试验中, 我们设定

$\lambda_1 = \lambda_2 = 0.5$, $\lambda_3 = \lambda_4 = 4$. U^k 的目的是评估可能的一个像素是前景或背景如何根据之前的时间显著地图计算步骤。

一元外观项 A^k 鼓励像素标记为有类似的颜色为前景像素高凸起。第三一元项 L^k 定义标记像素位置的先验估计从动态位置模型。两两项鼓励时空平滑 V^k 和 W^k , 分别详细描述了所有条款。

卓越术语 U^k

一元凸项 U^k 是基于我们的特点检测结果, 惩罚标签分配显著较低的像素值的前景。这个术语 U^k 具有以下形式

$$U^k(l_i^k) = \begin{cases} -\log(1 - S^k(x_i^k)) & \text{if } l_i^k = 0; \\ -\log(S^k(x_i^k)) & \text{if } l_i^k = 1. \end{cases} \quad (8)$$

外表术语 A^k

前景和背景的外观模型, 两个加权颜色直方图计算在 RGB 颜色空间中, 应该用 H_f 和 H_b 。每个颜色通道均匀量化成 10 箱, 共计 10^3 箱。堆放到每个像素根据其颜色值和加权直方图的卓越价值, 在像素的权重 x 是 $S^{k(x)}$ 和 $1 - S^{k(x)}$ 。然后我们建立前景和背景的全球外观模型规范 H_f 和 H_b 。更特别, 属

于两类像素的像素采样形成 H_f 和 H_b :

一个凸起的像素值大于平均值的自适应阈值定义为时空显著地图, 和一个像素空间连接到前像素。我们这些像素表示为 X_s 。该策略充分利用时空信息的显著结果, 能够消除不良影响的一些背景与前景颜色相似区域, 从而提供更精确的背景估计。让 $c(X_i^k)$ 表示像素 X_i^k 的 RGB 颜色值像素的直方

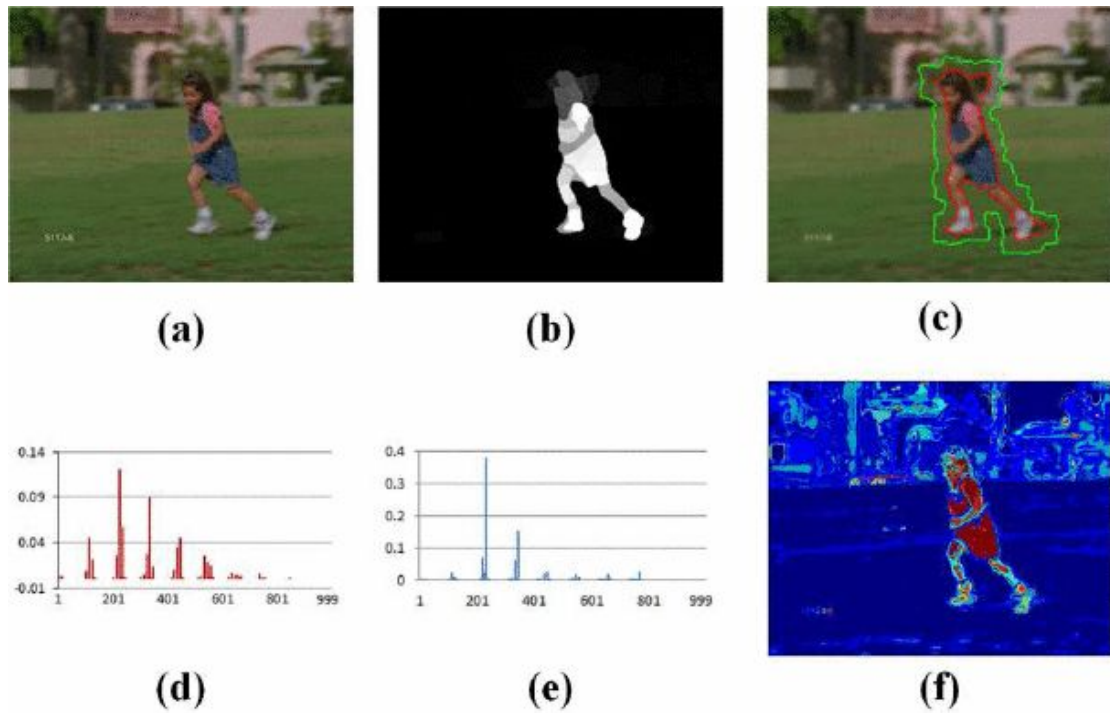
图, 一元外观项 A^k 的定义

是:

$$A^k(l_i^k) = \begin{cases} -\log\left(\frac{H_b(c(x_i^k))}{H_f(c(x_i^k)) + H_b(c(x_i^k))}\right) & \text{if } l_i^k = 0; \\ -\log\left(\frac{H_f(c(x_i^k))}{H_f(c(x_i^k)) + H_b(c(x_i^k))}\right) & \text{if } l_i^k = 1. \end{cases} \quad (9)$$

位置术语 L^k

甚至高于努力使尽可能准确的偿还的外观模型, 估计场景时仍然可以扭曲的复杂或背景区域与前景分享相似的外观。物体运动连续性之间的一些后续的框架, 提供了一个宝贵的前定位区域可能包含的对象。因此, 我们设计一个方法来估计前景对象对运动信息的位置从一个小数量的相邻帧。k 框架, 我们在一些后续帧积累前方后方 t 帧的梯度光流大, 收益率(殖利率)对象的轨迹:



图三

$$E_t^k = \sum_{i=k-t}^{k+t} E_o^i = \sum_{i=k-t}^{k+t} \|\nabla V^i\|. \quad (10)$$

有一个更大的 t 在一定框架、远程运动信息将考虑忽略一些不可靠的光流估计从小型的帧数。然而,这可能使 E_t^k 丧失辨别能力对象因为太多的运动信息是不必要的。 t 一样小 0 时,只考虑当前帧的运动信息可能正是对象位置但有时会失败,因为光流估计不准确。在我们的实验中,我们设置 $t=5$ 。然后我们使用 2.1 节中所述基于内部帧图施工方法来计算每一帧的动态选址模型。最后,我们可以得到位置之前像素

x_i^k , 一元位 L^k 的定义

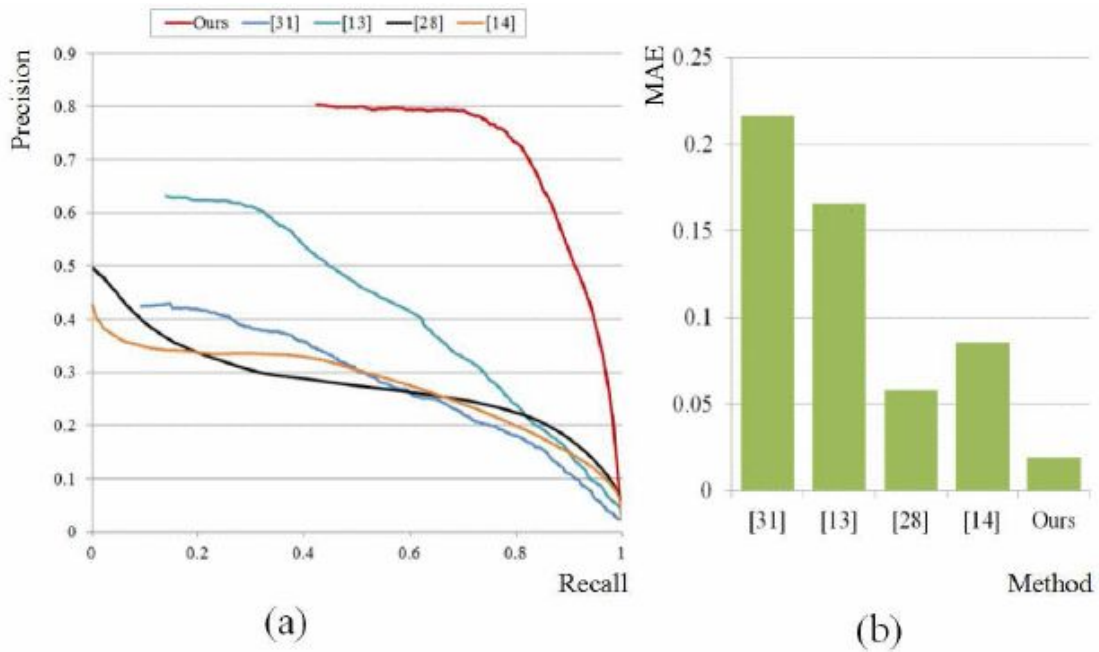
是:

$$L^k(l_i^k) = \begin{cases} -\log(1 - L^k(x_i^k)) & \text{if } l_i^k = 0; \\ -\log(L^k(x_i^k)) & \text{if } l_i^k = 1. \end{cases} \quad (11)$$

成对术语 V^k W^k

V^k W^k 组成的一致性,约束分割时空上的标签是一致的。这两个术语遵循 [27] 中定义的传统形式,有利于分配相同的标签相邻像素有相似的颜色。在定义完整的能量函数 $F(l)$,我们可以使用图像切割来计算最优二进制标记,从而得到最终的分割结果。

三. 实验结果



图四

我们的观点是自动检测视频序列和部分前景对象。在本节中,我们在视频特点检测上第一个测试我们的方法。即使它不是我们算法的最终目标,我们仍然通过比较我们的时空特点结果与技术发展水平显著方法

[31], [13], [28], [14]跟踪数据集[29]评估我们的方法的有效性。然后我们比较我们的分割结果和9替代方法跟踪[29],跟踪v2[18]和Youtube数据集。在该算法中,我们利用时空的边缘信息来计算前显著地图视频。作为我们的方法这是重要的一步,我们通过其他卓越的方法评估结果。使用代码获得相应的作者,我们的结果与五个备选方法[31], [13], [28], [14]比较我们的时空特点。第一个方法旨在图

像特点检测视频特点而设计的后三个检测。评估我们的方法的性能,我们测试我们的结果基于两个广泛使用的标准,包括公关(精密召回)曲线和MAE(平均绝对误差)。我们第一次使用精密召回分析评估我们的方法。精确的定义是正确的百分比显著像素分配,尽管召回措施的比例显著像素检测。情节精确召回治愈,我们从每个方法生成二进制显著地图使用一个固定的阈值。公关曲线所获得的不同阈值从0到255。

对于一个更加平衡的比较,我们遵循Perazzi.[24]评价之间的平均绝对误差(MAE)连续凸起映射和二进制地面实况对所有图像或帧像素。MAE定义为 $MAE = |S - G| / N$,这里N是图像或帧像素的个数。



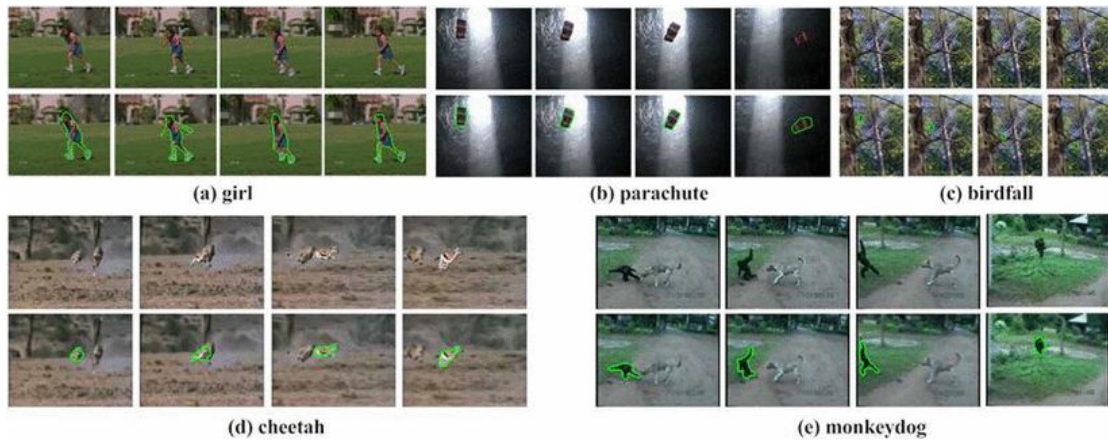
图五

method	Ours	[32]	[23]	[15]	[6]	[19]	[4]	[22]	[29]	[9]
birdfall	209	155	189	288	217	468	468	606	252	454
cheetah	796	633	806	905	890	1175	1968	11210	1142	1217
girl	1040	1488	1698	1785	3859	5683	7595	26409	1304	1755
monkeydog	562	365	472	521	284	1434	1434	12662	563	683
parachute	207	220	221	201	855	1595	1113	40251	235	502
Avg.	427	452	542	592	868	1727	1911	19079	594	791
supervised	N	N	N	N	N	N	N	N	Y	Y

表一

MAE 估计之间的近似程度显著图和地面真理, 这是 $[0, 1]$ 规范化。MAE 提供了一个更好地估计显著图和地面之间的不同的真理。结果精度召回曲线见图 4(a), 它提供了一种可靠的比较多好各种各样的显著地图强调突出地区图像。结果表明, 该算法明显优于其他方法。阈值是接近 255 时, 召回的值 [31], [13], [28], [14] 非常小, 甚至召回 [28] 和 [14] 的值降低为 0, 因为他们

的卓越地图不响应对象的关注。最低召回该方法的值不降至零, 因为对应的凸起地图能够有效地检测出突出的地区, 并有很强的反应。此外, 我们卓越的方法达到最佳性能精确率高于 0.8, 这表明我们的卓越地图在响应显著区域更精确。MAE 结果呈现在图 4(c)。相比于其他方法的最好的结果 [29]。我们 MAE 显著地图成功地减少 75%。



图六

图 5 给出了不同的视觉比较方法, 在亮像素显示显著的概率更高。图像显著方法的性能[31]并不好, 有些凸起地图甚至不能正确检测前景对象。运动信息的缺乏限制了他们的能力, 准确定位对象, 特别是在前景和背景颜色相似。在大多数情况下, 凸起方法 [13]、[28], [14] 视频能够准确定位凸物体, 它执行比图像特点检测的方法 [31]。因为这些时空利用运动信息的方法。然而, 一些显著地图使用 [13], [28], [14] 中产生低分辨率和倾向于相对较低的概率分配给对象内部的像素。这是因为光流估计有时是不正确的。基于之前的分析, 我们可以得出两个重要结论: (1) 运动信息检测前景对象给予有效的指导; (2) 使方法过分依赖于运动的信息并不是一个很好的选择。综合利用空间和时间空间的各种特性(等颜色、边缘、运动)应该产生更满意的结果。总的来说, 我们的

模型能够更好地估计显著地图在像素级和在杂乱的背景物体的轮廓。时空上我们的框架产生相干视频对象分割结果完全无监督的方式, 和我们比较 9 个方法是最密切相关的作品发表在最近几年。平均每帧像素出错率[29]介绍进行评估, 这是像素的数量根据地地面真理更进一步的细分。平均每帧像素出错率相比, 这些方法 [32], [23], [15], [6], [19], [4], [22], [29], [9] 为每个视频跟踪数据集 [29] 总结在表 1 所示。方法在 [32], [23], [15], [6], [19], [4], [22], 我们的方法不受监督。他们在视频自动检测对象以及部分对象。方法 [29] 和 [9] 的监督, 需要一个初始的第一帧的注释。结果显示, 我们的方法在测试视频中具有最低的平均每帧分割错误。图 6 显示了定性结果的视频跟踪数据集 [29]。它可以观察到, 我们的方法有能力段大变形形状的对象(女孩), 前景/背景颜色重叠(降落伞)和摄像

dataset	Ours	[32]	[23]	[15]	[6]
SegTrack v2	4766	25289	5859	23161	16074
Youtube	2208	11148	3461	20115	16858

表二

机运动(猴狗),同时也生产准确分割,即使对象是非常小的(鸟下降),或前景与快速运动模式(猎豹)。我们进一步开展实验跟踪 v2 数据集[18]和 12 组随机选择来自 Youtube 的视频对象和我们与[32]的方法相比,[23],[15],[6]。平均每帧的像素出错率见表 2。看到,我们的方法显著优于所有其他正轨 v2[18]和 Youtube 数据集。

四. 结论

我们提出了一种无监督的方法,包含了测地距离凸起授权的视频对象分割。提出了时空的边缘地图显示能够表明前景和背景的位置。我们的方法综合时空边缘地图和测地线距离获得准确的时空特点作为前对象分割结果。我们生产时空显著地图通过计算测地线距离估计背景内部帧图上每一对相邻帧。最后,我们计算分割结果通过结合显著,全球外观模型和位置模型到图像切割能量最小化。许多性能结果表明,我们的方法产量明显高于最先进的方法。

脚注

这部分工作是支持中国国家基础 Research 计划(973 计划)(2013 号 cb328805), 中国国家自然科学基金(61272359), 澳大利亚研究理事会的发现项目资助计划(项目 DP150104645), 和新世纪优秀人才计划的大学(ncet - 11 - 0789)。专业的资金联合建筑项目北京市教育委员会。

参考

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk "[Slic superpixels compared to state-of-the-art superpixel methods](#)" *IEEE TPAMI*, vol. 34, no. 11, 2012
2. C. Antonio, S. Toby and B. Andrew "Geos: geodesic image segmentation" *ECCV*, 2008
3. X. Bai and G. Sapiro "[A geodesic framework for fast interactive image and video segmentation and matting](#)" *ICCV*, 2007
4. O. Barnich and M. Van Droogenbroeck "Vibe: a universal background subtraction algorithm for video sequences"

- IEEE TIP, vol. 20, no. 6, 2011
5. W. Brendel and S. Todorovic "Video object segmentation by tracking regions" CVPR, 2009
6. T. Brox and J. Malik
- "Object segmentation by long term analysis of point trajectories"
- ECCV, 2010*
7. T. Brox and J. Malik
- "Large displacement optical flow: descriptor matching in variational motion estimation"
- IEEE T-PAMI, vol. 33, no. 3, 2011*
8. J. Carreira and C. Sminchisescu
- "Constrained parametric min-cuts for automatic object segmentation"
- CVPR, 2010*
9. P. Chockalingam, N. Pradeep and S. Birchfield
- "Adaptive fragments-based tracking of non-rigid objects using level sets"
- CVPR, 2009*
10. A. Criminisi, T. Sharp, C. Rother and P. Perez
- "Geodesic image and video editing"
- ACM TOG, vol. 29, no. 5, 2010*
11. I. Endres and D. Hoiem
- "Category independent object proposals"
- ECCV, 2010*
12. K. Fragkiadaki, G. Zhang and J. Shi
- "Video segmentation by tracing discontinuities in a trajectory embedding"
- CVPR, 2012
13. H. Fu, X. Cao and Z. Tu
- "Cluster-based co-saliency detection"
- IEEE TIP, vol. 22, no. 10, 2013
14. C. Guo, Q. Ma and L. Zhang
- "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform"
- CVPR, 2008
15. Y. J. Lee, J. Kim and K. Grauman
- "Key-segments for video object segmentation"
- ICCV, 2011
16. M. Leordeanu, R. Sukthankar and C. Sminchisescu
- "Efficient closed-form solution to generalized boundary detection"
- ECCV, 2012
17. J. Lezama, K. Alahari, J. Sivic and I. Laptev
- "Track to the future: Spatio-temporal video segmentation with long-range motion cues"

CVP R, 2011

18. F. Li, T. Kim, A. Humayun, D. Tsai and J. Rehg

"Video segmentation by tracking many figure-ground segments"

ICCV, 2013

19. T. Ma and L. J. Latecki

"Maximum weight cliques with mutex constraints for video object segmentation"

CVPR, 2012

20. V. Mahadevan and N. Vasconcelos

"Spatiotemporal saliency in dynamic scenes"

IEEE TPAMI, vol. 32, no. 1, 2010

21. S. Mathe and C. Sminchisescu

"Dynamic eye movement datasets and learnt saliency models for visual action recognition"

ECCV, 2012

22. P. Ochs and T. Brox

"Higher order motion models and spectral clustering"

CVPR, 2012

23. A. Papazoglou and V. Ferrari

"Fast object segmentation in unconstrained video"

ICCV, 2013

24. F. Perazzi, P. Krahenbuhl, Y. Pritch and A. Hornung

"Saliency filters: Contrast based filtering for salient region detection"

CVPR, 2012

25. B. Price, B. Morse and S. Cohen

"Geodesic graph cut for interactive image segmentation"

CVPR, 2010

26. E. Rahtu, J. Kannala, M. Salo and J. Heikkila

"Segmenting salient objects from images and videos"

ECCV, 2010

27. C. Rother, V. Kolmogorov and A. Blake

"Grabcut: Interactive foreground extraction using iterated graph cuts"

ACM TOG, vol. 23, no. 3, 2004

28. H. J. Seo and P. Milanfar

"Static and space-time visual saliency detection by self-resemblance"

Journal of vision, vol. 9, no. 12, 2009

29. D. Tsai, M. Flagg, A. Nakazawa and J. M. Rehg

"Motion coherent tracking using multi-label mrf optimization"

IJCV, 2012

30. C. Xu, C. Xiong and J. J. Corso

"Streaming hierarchical video
segmentation"

ECCV, 2012

31. C. Yang, L. Zhang, H. Lu, X. Ruan
and M.-H. Yang

"Saliency detection via graph-based
manifold ranking"

CVPR, 2013

32. D. Zhang, O. Javed and M. Shah

"Video object segmentation through
spatially accurate and temporally dense
extraction of primary object regions"

CVPR, 2013

