

指导教师： 杨涛

提交时间： 2016/3/20

CVPR2015 Paper Translation

No: 01

姓名： 白成成

学号： 2013302470

班号： 10011301



使用自适应聚集决策树和动态多级外观模型的单目标跟踪

Jingjing Xiao

Rustam Stolkin

Ale's

Leonardis

School of EESE

School of Mechanical Engineering

School of Computer Science

University of Birmingham

University of Birmingham

University of Birmingham

摘要

本文介绍了在视频序列中任意对象的单目标跟踪的方法。目标三个不同的粒度级别（像素级，基于部件级和边界框级）上建模，这种级别的跨约束，确保鲁棒模型再学习。其主要贡献是一种自适应聚集决策树方法，动态选择的必要充分代表在每一帧每个目标的部分特征的最小组合，从而提供鲁棒性的计算效率。自适应聚集决策树在跟踪算法的两个独立的部分实现的：首先，以使在连续帧之间的基于部件级鲁棒匹配；其次要选择学习目标的新部件的最佳超像素。我们测试使用两种不同的跟踪基准

（VOT2013-2014 和 CVPR2013 跟踪挑战），基于两种不同的测试方法跟踪，并显示它是显著比这两个最好的国家的最先进的方法更稳健跟踪的挑战，同时也提供有竞争力的跟踪精度。

1. 介绍

经过几十年的视觉跟踪研究，即使是最先进的跟踪在复杂的场景下仍旧倾向于失败，包括杂波和伪装在一个或多个功能的方式，快速的和不稳定的目标运动，阻塞故障，并改变它们的形状和外观上的目标 时间。这些有问题的跟踪条件主要导致跟踪算法的三个基本部件故障：1) 目标对象的视觉外观的模型；2) 的机制在每帧匹配模型部分图像区域；3) 机制不断重新学习或更新的改变它们随着时间的推移外观目标的模型。

本文提出了一种目标跟踪算法，通过解决这三个基本领域实现先进的鲁棒性。我们提出了一个灵活的目标表示可以自适应地利用任意数量的不同的图像特征。目标是在三个不同的粒度级别，包括单个像素的水平，水平的部分和编码有关目标作为一个整体的整体信息的边界框水平参照。跨约束这些更新，过程中的不同级别之间启用连续目标模型再学习健壮和稳定。

本文的主要贡献是一种自适应聚集决策树方法来动态选择的必要充分代表在每一帧每个目标的部分特征的最小组合，从而提供鲁棒性不牺牲计算效率。我们显示如何自适应聚集决策树所用的跟踪算法的两个独立的部分被利用：首先，以使在每帧中的基于部分的水平健壮匹配；其次要选择学习目标的新部件的最佳超像素。在匹配过程中，自适应聚集决策树用于搜索该组超像素的当前帧中的，以找到在先前帧中的目标部分的最佳匹配。在模型更新，决策树被用来搜索最适合的超像素从中到新的目标零件模型，以取代已被检测为漂流的旧的目标的一部分。

我们已经用最新的基准进行了原则性的评估，和其他最好的追踪进行了比较。结果表明，该方法优于最佳 4 纤夫 VOT 2013 和 2014 年基准集。它也显著优于 7 最佳的可用方法对 CVPR2013 数据集 wrt 鲁棒性，同时也实现了竞争的跟踪精度。

本文剩下的部分组织如下。在第 2 部分，讨论了相关的工作。多层次的

目标模型和它的初始化在第 3 部分被介绍。第 4 部分解释自适应聚集的决策树，并显示它如何用于目标匹配和模型更新每个连续的帧。第 5 部分提出和论述了测试追踪器在 VOT2013、VOT2014 和 CVPR2013 基准的视频数据集上的实验结果。第 6 部分是结论总结。

2. 相关工作

在本节中，我们回顾最近跟踪算法三个主要组成部分：目标表示、匹配机制，以及模型更新机制。

目标代表性的选择是任何跟踪的重要组成部分。两个主要的研究流能够被区分。第一种使用跟踪，例如整体（总）目标模板[22]和[24]，本身从[10]的基本思想衍生物。然而，这种方法具有在处理显著外观变化与目标的变形困难。以后的工作中，[27]表明信息如何上下文可以用来支持任何目标模型的一部分，是最挑剔的当前帧自适应选择，但仍限于其本身没有更新一个简单的整体目标模式。以后的工作中，[18] [19]提出了基于补丁的办法，以提供对靶匹配更多的灵活性。但是，对于本地修补程序运动几何约束的选择仍然是一个悬而未决的问题，而环境的混乱和伪装经常可以分散这种局部修补，使他们漂移。[33]避免了补丁运动复杂几何约束，通过处理的问题，因为前景和背景超像素的分类。然而，由于每个超像素独立地进行分类，该方法仍然容易发生故障与杂乱背景场景。相比之下，我们的方法还利用超像素的，但更健壮的跨约束多目标模型结构中利用它们。最近的工作[30]也为更强大的代表性全局和局部基于补丁目标模型结合在一起。然而，这项工作稠合以均匀的方式，这将导致条件，其中一个或多个特征变得小于其他鉴别下故障多个特征。相比之下，我们的方法，

通过在有利于任何特征或特征组合的自适应选择实现更好的结果是最识别为在每一个新的帧中的每个目标部分。其他最近的工作[21]同时跟踪，学习和分析使用分层和成分与或图表示的目标。然而，该算法使用了一个相对固定的细胞格来量化 AOG 结构，其可包含与小语义节点。相比之下，我们的方法直接链接树节点到目标超像素，它们更可能代表了目标对象的均匀的和有意义的部分。最近的工作中[17]，它进行的基础上树结构目标的外观模型跟踪算法。他们估计，采用了多项关键帧的最优树，并采用歧管半监督对准技术来构建一个树的所有帧。然而，这种脱线跟踪器是不适合的，它连续地改变它们的形状和视觉外观随着时间的目标顺序跟踪。相比之下，我们提出了一个具有高度适应性决策树结构，重新学习是在每一个新的帧每个目标的一部分，这使得与快速变形目标视频的好成绩。来估计所述目标的状态中，算法必须从候选图像区域观测到的目标表示模型相匹配。单个特征是不足以处理大出现的变化，以及最近的作品[25] [29] [28] [21]日益利用的多个特征的组合。一种方法是从所有特征计算的可能性，然后乘以所有值来估计所述目标状态，例如[30]。然而，在这种方案中，一个表现不佳的功能，可能会损坏跟踪性能，即使在其他功能都非常歧视。因此，代替以同等的重要性，其它方法，例如处理所有的功能[8] [32] [29] [28]，尝试有利于在每个时间步骤中的最判别特征（或目标模型的最判别份），以确定和重量。Brasnett 等。 [6]提出了一种方案有利于表现最好的功能权重，并在每个新帧自适应更新这些权重。但是，该方法忽略来自本地背景区域特征的显着性。与此相反，最近的作品[29]和[28]提出了两种不同类型的自适应方法其中两个成功利用用于最佳加权的

贡献从每个跟踪期间在线功能的上下文信息。然而，这些方法只使用一个简单的整体目标模式是不足以应付大的目标变形和外观的变化。[25]提出一种同时使用目标和上下文 SIFT 特征的匹配方法。然而，匹配指数是由最邻近搜索中，当靶经历快速和显著变形和外观的变化可能表现不佳直接获得。与此相反的治疗的所有特征，例如[30]，我们的自适应聚集决策树方法能够自适应地支持最歧视性的特征选择对每个目标部分到每个新帧匹配。然而，这种自适应特征选择还嵌入一个跨约束多层次的目标代表性，使更强大的匹配和模型更新比例如简单的整体目标模型内[29]或[28]。

对于稳健的，一般的跟踪，它不断地更新或重新学习目标模型，以应付外观变化是必不可少的。一个合适的目标模式应该使跟踪器以克服重新学习的过程，可能会破坏目标模型的错误，并支持长期跟踪漂流无[22]。早期的方法，如[24]，在每一帧为以前的模型的一个简单的线性组合与目标区域的当前图像中的最新估计更新模型（有时称为自回归更新过程）。没有为目标部件的精确划定的其他方法，例如更新方法很可能会失败，在连续的更新给定足够长的跟踪时间，由于累积误差和噪声。在 MIL [4]和其他跟踪，如 OB [12]和 SB [13]目标模型的更新是一个不断发展的助推分类，跟踪图像补丁和学习对象进行外观。有趣的是，OB [12]可以被视为非贝叶斯近似简单 ABC 贝叶斯方法[27]，但其使得目标模型的连续再学习。然而，在线升级要求数据是独立同分布的，这是一个条件中最真实的视频序列，其中，数据通常时间相关[25]不满意。更健壮的更新机制是由[30]，其形成一个横约束范式来稳定地约束两层目标模式的再学习取得 - 全局（边界区域）和局部（份基础的）模型被用来约束（以及由此稳定）彼此

的网上再学习。但是，这种方法（和大多数早期方法例如[24]）更新以固定的速率目标外观模型，无论在当前目标观测置信（或缺乏）。这个问题是由先前描述的问题复杂化，这许多方法，例如[30]，结合的具有相等权重的所有功能，这可能会导致跟踪失败时的一个或多个特征的方式相比，其它的是判别很差的意见。如[30]，我们还采用多级的跨约束的方法来鲁棒性在线目标再学习，但是我们通过自适应地改变的再学习速率，在每一帧中的每个目标部分的基础上，实现显著改善了[30]性能评价结果从上下文信息导出电流跟踪置信度。

3. 多级目标模型

原则上我们的方法可以与的特征的任何组合使用，提供的像素的该特征值的图像区域可以与模型相关联（例如：直方图），并且一个合适的似然函数的存在是为了比较这些模型的相似性为两个这样的区域（原则的证据，我们使用基于 Bhattacharyya 距离的度量[5]尽管其他也是可能的）。在[2]和[34]，或方法的精神，在第一帧只使用一个给定的边界框初始化它的目标模式。像素（底部）的水平，基于份-（中）的水平，和边界框（顶部）层：目标是在三个不同的粒度级别分级建模。继模型构建的逻辑和模型更新的顺序，先介绍一下中等水平模式，因为最初的功能是从这个层面提取。后来，我们引进了顶级其次是底层。

顶层（总体目标外观模型）比较均匀采样一整套候选人边界框的位置，和数学期望算子用于提供候选边界框区域可以创建新的中间层次模型的位置。这个候选人的边界框然后分割成，使用 SLIC 像素发挥其分割方法 [3]。这些超像素然后用于标识一组合适的目标部件，形成中间层次

的模型组成的 M 部分，每个与 N 的特点，即 $\{f_i, m_k\} i = 1 \dots N, m = 1 \dots M$ 在哪里上网， m_k 代表 i th 零件在泰熙帧的年月特征。使用超像素作为基础中间层次特征直方图（简称 $\{h_i, m_k\} i = 1 \dots N, m = 1 \dots M$ ）比随机选择在前面的方法 [30] 的平方“修补程序”的几个优点。首先，像素发挥其是目标的更有可能对应于在语义上有意义的和同质的一部分。相比之下，随机（或均匀）选择的补丁程序是目标的可能包含两个或更多的不同（例如的颜色）部分，从而导致匹配问题的像素。其次后修补程序随机选择（或均匀）从最初的边界框，许多修补程序有可能包含来自目标和背景像素的信息，这也是可能产生负面影响跟踪性能。与此相反的是，由于超像素的均匀性质，从这些区域提取的特征是更有可能包括纯粹目标像素或纯粹背景像素。一旦开始跟踪，这些修补程序的错误对应的背景区域快速检测和排除目标模型，留下只有真正属于目标区域像素的这些修补程序。我们使用的 SLIC 像素发挥其分割方法 [3]，因为它提供了以下优点：可以提前知道超像素的数目、超像素有大小均匀、压实度可以定义该算法具有较高的计算效率。然后，中间层模型 k 是一整套的 N 个零件（小矩形），提取 N 超像素表示为：

$$\zeta_k = \{c_k^i, h_k^{i,m}\}_{i=1 \dots N, m=1 \dots M} \quad (1)$$

其中， k 为在帧 k 的部分 i 的图像坐标。每一个部分是由一组直方图，一个直方图对每个特征模式的表示。即 $\{h_i, m_k\} i = 1, \dots, N, M = 1 \dots M$ 表示一组 $M \times N$ 个直方图哪种模式 M 在各个不同的 N 中等水平的目标部分，从 N 个超像素得出不同的特点。目标模型的顶层被表示为 ζ_k 其包括关于边框区域总体信息：

$$\zeta_k = \{C_k, H_k^{B,m}, H_k^{F,m}\}_{m=1 \dots M} \quad (2)$$

其中， C_k 是边框图像坐标在帧 k （从零部件的中间级别车型的分配每个新帧重新计算）。背景特征直方图 $\{H_k^B, m_k\} M = 1 \dots M$ 是每个 M 个不同功能模式的直方图，通过缩放系数扩大目标的边框定义环形本地背景区域提取。 $\{H_k^F, m_k\} m = 1$ 时， \dots, M 是一组前景直方图，在每一个被从中间电平部分模型作为组合信息计算 M 个特征模式的：

$$H_k^{F,m} = \frac{1}{N} \sum_{i=1}^N h_k^{i,m} \quad (3)$$

底部级包括单个像素，每个都有其 RGB 值与所分配的可能性相关联。这个可能性是通过简单贝叶斯推理使用顶级前景和背景直方图计算。让 $FM(x)$ 为第 m 个要素中的像素位置 x 的值。此外，让 $\pi_k(X)$ （调频 (X) ）表示一个特殊的功能，它接受作为参数的特征值 $FM(x)$ 在图像位置 x ，并输出对应于该相关仓中的直方图 $H_m k$ 处的值的概率，它表示的图像帧 k 的 M 型特征的直方图。根据第 m 特征模式的统计信息，像素的对应于前景或背景区域的似然现在是 $P(X|F) = H_k^F, M_k(FM(X))$ 和 $P(X|B) = H_k^B, M_k(FM(X))$ ，其中的符号 F 和 B （以及标 H_V 和 H_B ）分别表示前景和背景信息。

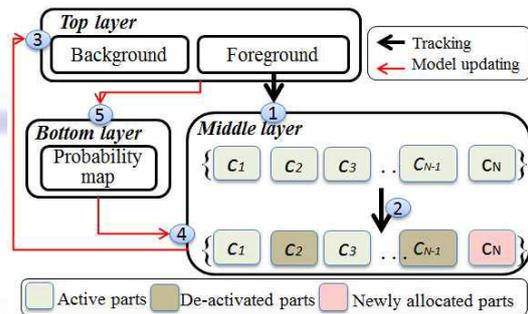
在任何特定的特征模式，一个特定像素的代表被跟踪的对象的整体概率可以计算为：

$$P_k^m(F|x) = \frac{\lambda P(x|F)}{\lambda P(x|F) + (1-\lambda)P(x|B)} \quad (4)$$

其中 λ 表示目标的图像区域的尺寸的尺寸的预期比率被搜索，以像素为单位。典型地，该搜索区域和包围盒不断调整以确保 λ 的大致恒定的值。请注意，这是微不足道的延长式的制定。4，包括任意多个功能的方式。

4. 追踪器传播和匹配

跟踪器的整体概略示于图 1. 跟踪设备信息匹配的顶级水平, 产生了中等水平的候选图像区域的第一传播。接着, 将该候选区域是由[3]被分割成大小相等的超像素。我们接下来提出了一种连续自适应群集树的方法, 它有效地找到最佳对应于在新的帧从先前帧的匹配中层目标模型部分, 到新分段超像素。连续自适应群集树的方法是有效的, 因为它适应利用的功能的最小数目的用于在每个图像帧中的每一目标部分匹配。最后, 应对目标的变形和外观的变化, 旧的目标部件的外观(中层补丁)可以自适应更新或严重漂流补丁可以被完全暂时关闭, 并通过新的修补程序取代。在这种情况下, 一个新的自适应聚集决策树被用来选择最合适的超像素用于形成新的部件。



跟踪处理的图 1 框图。通过前景信息在顶层配套 1-跟踪传播; 2-中等(超像素)与集群决策树电平匹配; 3, 更新顶层模型; 4-反馈底部(像素)的电平信息; 5 处于中等水平漂流零件重新采样。

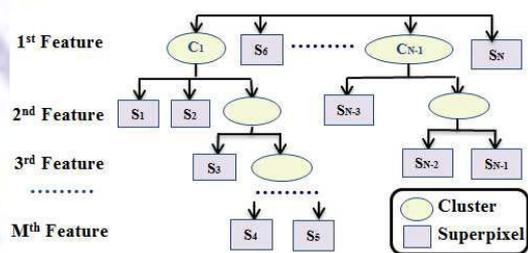
4.1 顶级传播

顶层使用一组均匀分布的样本来表示目标的后验密度函数传播, 具有相关联的权重。与[24], 总体目标位置是通过使用期望算子在一组样品, 其重量由前景特征匹配计算的估计。

4.2 自适应集群决策树的中度匹配

传播的顶层, 得到一候选边界框区域之后, 这被进一步扩大, 形成一个更广泛的搜索区域上, 然后再把段成超像素, [3]。现在这些超像素上一个图像匹配到中层的目标部分(补丁)提供的新形象候选位置的池。用我们提出的自适应聚集决策树方法执行此匹配。通常情况下, 一个决策树的结构将脱机通过训练, 这将是适合于具有动态外观跟踪目标获得。相反, 我们提出这是在网上重新学习对每一个新的图像中的每个新的目标的一部分, 通过明确考虑前景和背景信息的完全自适应集群树。

所提出的自适应聚类决策树方法如图 1 所示。2. 决策树的目的是图 2 群集决策树。每个树级别的代表功能模式(例如颜色, 运动等)。超像素的特征值构成在树上级叶片。如果任何两个叶子是在特征值足够相似然后将它们合并成一个群集。然后一个新的树级别(使用下一特征模式)被用于尝试消除歧义这种集群的成员。树继续生长(通过添加更多的功能更树级别), 直到所有的目标部分已被分配给超像素的独特选择。假定任何剩余部分不匹配已经成为闭塞, 并暂时关闭。



找到相应的超像素其中最佳匹配各中层目标部分, 而自适应选择的在每一帧鲁棒为此所需特征的最小组合。第一树级通过从集合的所有特征选择一个主要特征, 并标记所有超像素作为单个叶初始化。首先, 每片叶子是针对中等水平目标模型的一部分, 它匹配寻求比较。

$$\begin{aligned} \arg \max_{(i_p, i_c)} B(h_k^{i_p, m}, h_k^{i_c, m}) \quad s.t. \\ B(h_k^{i_p, m}, h_k^{i_c, m}) > 0 \end{aligned} \quad (5)$$

其中 i_p 表示处于中等水平目标模型的零件, i_c 是候选叶和 $B(\cdot, \cdot)$ 是相似性指标。接着, 将剩余的候选叶子相比所选择的叶, 按原样来表示, 在主动功能方式相似。类似叶子被分组为当:

$$B(h_k^{i_s, m}, h_k^{i_c, m}) > T_k^m \quad (6)$$

其中 $B(h_{is}, m, K, h_{ic}, m)$ 是一个选中的树叶之间的相似度量第 m 个特征直方图是另一个候选人叶 IC 和 TM_k 是一个集群的门槛。

标记 T_m 值有上每个功能的中级部分匹配过程中使用的程度有重要的影响。的 T_m 值高的值降低在各树层次聚类的量, 以确保大多数超像素将被表示为单独的叶, 即, 算法将区分单独使用此功能最超像素。相比之下, T_m 值低的值, 使其可能这个树形水平将增长许多集群, 以使过所有的算法将较少使用匹配该特定功能的效果, 并依靠后功能模式(更深在树中的级别)提供的歧视。换言之, T_m 为 k 的选择实际上可以视为在一特定特征的鉴别能力置信度的量度。

因此, 我们愿意为业绩不佳的功能(设置这些功能, 是当前帧的高度歧视性的 TM_k 的高值, 和 TM_k 的值较低例如, 那些其目标在本伪装映衬方式帧)。因此, 我们建议由 T_m 为 k 时在每一帧连续地重新学习每个功能的基础上, 评价判别相对于当前的上下文信息的能力的特征的方法。 T_m 为 k 被在线计算为:

$$T_k^m = \exp\left\{\frac{-B^2(H_k^{F, m}, H_k^{B, m})}{\sigma_m^2}\right\} \quad (7)$$

其中, 高频中, M_k 和 HB 中, M_k 为在方程限定的顶层前景和背景直方图。 2, 和式因此 7 是目标, 并在各自的特征模式的本地背景(即伪装的

量度)之间相似性的量度。 $B(\cdot, \cdot)$ 是一样的相似性量度如等式。 5。

如果任何候选超像素既是 i) 从别人足够不同, 它形成其自己的叶和不位于集群(例如 S_6 或 SN 图 2)的内部和 ii) 强烈的目标部分相匹配, 满足等式 5, 则决策树停止生长及中等水平目标部分被标记为匹配的候选超像素。一旦所有的中间电平超像素在当前图像中的被匹配到新的位置, 它们的分布是用来输出一个新的边界框的位置, 并且顶层目标模型被相应地更新。偶尔, 在模型中的一些部分将不能找到一个强匹配候选超像素, 甚至耗尽所有可能的图像的功能(对应于所有可能的树的水平)之后。在这种情况下, 可以推测该部分不再可见, 它被关闭。在文献中的其它方法(例如, [30])删除不匹配的补丁, 从而闭塞, 这是不能以后恢复期间永久失去模型的部分。相比之下, 我们的提议的暂时切断无与伦比份的方法提供了强大的跟踪存储器自动处理阻塞的情况下, 无需特别附加闭塞例程。

4.3 模型更新

用于鲁棒跟踪, 它是必要的, 因为它改变其随时间的外观, 不断更新的目标模式。建议跟踪通过两种机制做到这一点: 适应旧的底部和顶部的水平的目标模式, 并加入新的中等水平的部分新型。

在每一帧, 我们检查所有中层部件和检测那些漂流(移动从目标重心太远), 用改编自法[35], 它考察了部分地区的边缘分布。要更换零件漂流, 我们选择在当前帧这是最有可能代表目标部分的那些超像素。而另一种自适应聚集决策树的使用方法如下执行此选择。我们首先使用的主要特征直方图, 喜, 男 K , 最初排名根据所有无与伦比的超像素:

$$Rank = \text{sort}(s_k^{i, m}, \text{descend}) \quad (8)$$

其中 S_k 是计算为特征米超像素候选人的优先分数:

$$s_k^{i,m} = \sum_{x=1}^{N_p^i} P_k^m(F|x)/N_p^i \quad (9)$$

其中, $P_k(F|x)$ 是可能性, 从方程计算。4, 在图像位置 x 像素, 第 i 个候选超像素内的所有像素。如前一节为匹配中描述的本排名列表可视为上的自适应聚集的决策树结构的第一树级 (第一特征) 的叶子。接下来, 当它们满足下面的约束叶 (超像素) 具有相似的优先级得分都聚集:

$$\|s_k^{i,m} - s_k^{j,m}\| < \lambda_{rank} \sigma_{all}^m \quad (10)$$

其中 λ_{rank} 是等级为预先定义的参数, 而 σ_m 该是所有超像素预期可能性的标准偏差。

再次, 在树上级集群表明, 这棵树级的功能模式是不足以判别, 实现了稳健的排名。因此, 二次特征被选择并用于排列在群集内的所有一致的超像素, 形成在树的第二树级。树生长 (通过添加连续树的水平, 使用连续功能), 直到一个独特的排名已被分配给所有超像素。比, 排名最高的超像素 n 被选择以初始化 N 新零件。这些新的部件更换经鉴定为漂流 n 个旧部分。

对于 (即是在新的框架强烈匹配到超像素的) 目标 “部分” 或 “打补丁” 的模型是根据新的观测结果更新了非漂流的部分。需要注意的是任何一种目标模式再学习的是潜在的危险, 因为即使很小的跟踪误差容易造成背景像素要学习到目标模式, 从而导致不稳定性成倍增加的错误。早期的彩色微粒过滤器的工作 [24] 和国家的最先进的基于近期补丁的方法 [30] 执行模型在一个固定的更新速度再学习。相比之下, 我们不断地重新计算单独的更新速度在每帧中的每个中层的一部分。我们的前提是, 部件可以重新学习迅速时, 有一个高置信在他们的匹

配, 而再学习率应不确定性的情况下被降低。因此, 我们更新每一部分, 采用了连续重新学习参数率 μ_i, m_k :

$$h_k^{i,m} = (1 - \mu_k^{i,m}) h_{k-1}^{i,m} + \mu_k^{i,m} h_{obs}^{i,m} \quad (11)$$

$$\mu_k^{i,m} = B(h_{k-1}^{i,m}, h_{obs}^{i,m}) \quad (12)$$

其中, $h_{i,m,k}$ 是在帧 k 西斯部分的第 m 特征直方图, 而 $h_{i,m,OBS}$ 是在新的框架匹配的超像素的第 m 个特征直方图。再次, $B(\cdot)$ 是度量如等式所描述的第 m 个特征的相似性。5. 在每一帧, 一旦所有中层部分被任一关闭, 更新或替换, 则顶 (全局) 水平目标模式是根据公式更新。如在第 3 节所述。

4.4 遮挡处理

建议跟踪器利用了能够记忆最新的跟踪状态, 包括所有的中层部分机型的内存。如在节所述。4.2, 部分遮挡被暂时关闭匹配不佳中等水平的部分, 但在内存中保留这些, 一旦重新获得他们的目标闭塞部位在后面的视频帧再现处理。如果部分大比例 (由门限参数定义) 保持秒的匹配过程后无法比拟的。4.2, 则跟踪器被视为完全堵塞的一种特殊状态存在。

在完全堵塞状态下, 所有的目标模型更新 (在所有模型级别) 被关断, 并且传播范围 (围绕估计目标边界框候选图像区域) 被扩大。跟踪器返回到正常 (非闭塞) 状态的一次中等水平的地区有足够比例的再次强烈匹配候选超像素。

5. 实验

我们已经从公开提供的数据集 VOT2013, VOT2014 CVPR2013 基准数据集 [2], [34], 它们共同构成总共 70 序列测试了我们的跟踪器的性能上的序列。数据集的更多细节可以从网页中找到 [2] 和 [1]。

5.1 实现

所提出的自适应聚集的决策树结构设计来处理，原则上，在一个强大的和高效率的方式任意多的功能。对于原则的证明，我们已经实现了一棵树，只有两个树平（利用只有两个功能模式），但是这已经在提供基准测试数据强劲的竞争力的表现。对于这两种自适应聚类树结构（中间电平部分的匹配二段 4.2，和新的部件学习秒。4.3），我们使用简单的颜色直方图作为主要特征直方图，与常用的巴氏度量作为匹配似然测度。对于辅助功能，我们使用一个简单的运动测量，在那里，如果他们意味着对部分小型帧到帧运动相匹配的候选超像素被赋予较高的匹配可能性。该跟踪算法已实施了英特尔酷睿 i5-3570 的 CPU，用 Matlab 代码（也与一些 C++ 组件）。这个未优化实施方式中，旧的 PC 机上，以达到近于实时 8 帧每秒的性能（与不同的测试视频有所不同）。在第一帧初始化的关键参数都列在表 1。

Table 1. Values of key algorithmic parameters

Section	Equation	Value
Initialization	λ in Eq. 4	0.1
Decision tree	σ_m^2 in Eq. 7	0.05
	λ_{rank} in Eq. 10	0.1
Occlusion	the ratio of unmatched parts	40%

Table 2. VOT challenge results: comparing against best 4 trackers

	VOT 2013 (16 sequences)				
	Ours	PLT13 [15]	LGT++ [35]	EDFT [11]	FoT [31]
Fail.	0	0	1.53	14	22
Acc.	0.59	0.58	0.57	0.58	0.63
	VOT 2014 (25 sequences)				
	Ours	PLT14 [15]	DGT [7]	DSST [9]	SAMF [20]
Fail.	1	4	25	29	32
Acc.	0.52	0.56	0.58	0.62	0.61

5.2 进化

我们首先使用 ICCV2013 年和 2014 年 ECCV “VOT 挑战” [2] 测试平台评估我们的跟踪器。标签。2 我们对最好的 4 VOT 跟踪，出约 30 纤夫的那些难题评估跟踪器的性能进行比较。结果显示于鲁棒性方面平均超过所有帧（失败实例的总数）和准确度（跟踪器输出的边界框与地面实况边界框之间的百分比重叠）。

用于鲁棒性，我们的跟踪器具有 VOT2013 零故障，只有一个在 VOT2014 故障。下一个最好的算法是 PLT 这 VOT2013 也实现了零故障。但是，请注意，在 VOT2013 PLT 测试的版本中使用固定的边界框的大小。因此该算法是无法适应该追踪过程中改变它们的大小的目标。由于在大多数测试序列最幸运的对象大致保持不变的尺寸，这个刚性约束帮助算法实现了高鲁棒性的分数。对于 VOT2014，不同版本的 PLT 的提交，这确实能够适应不断变化的目标大小。在这种情况下，血小板的健壮恶化到四个故障。

需要注意的是精度分数有时会产生误导。在 VOT 测试方法，地面实况用来跟踪每次失败后重新初始化跟踪器（完全准确）。因此，这通常不能跟踪器将显示精度高的分数，即使他们不是“好”的跟踪。因此，比较具有相同的鲁棒性得分两个跟踪器主要是当精度比是有意义的。在 VOT2013 我们的精度比共享相同的鲁棒性，往往微不足道的 VOT2014 没有其他跟踪器能够达到火焰鲁棒性的唯一疗法算法（PLT）更好。

为了更广泛的比较，我们还与所有那些来自 CVPR2013 跟踪基准数据集 [34] 结合 VOT 测试序列。使用该 70 序列数据集，我们比较了针对已在任一 VOT 或 CVPR 跟踪挑战，即表现出较强的性能可公开获得的跟踪方法：触击 [14]，单片机 [36]，LGT [35]，CSK [16]，IVT [26] 中，L 1 [23]，和 PF [24]。由于该数据集包含完整的闭塞情况下，评价是不重新初始化跟踪失败后进行。我们显示的结果作为权衡曲线由 [34] 的建议。

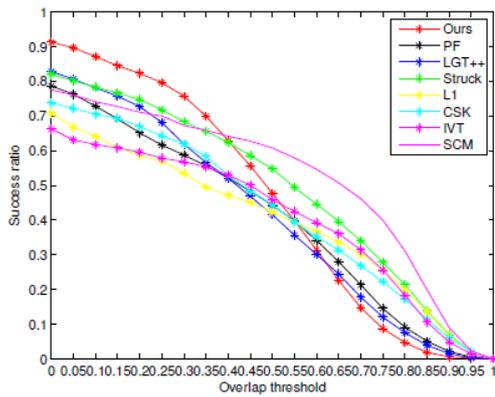


图 3

如示于图 3，我们的跟踪器达到显著埃特稳健性精度范围可达 0.4。我们主张帽子鲁棒性是最重要的这些度量的：首先，这些方法适用于高度可变形的目标（如个人）的量很难有意义解释的矩形边界框的“准确性”，其中包括许多非目标像素即使在良好的性能；其次，条件是一个跟踪器是健壮的，精度总是可以进一步通过将关于特定目标的附加的先验知识改善[9]。

为了进一步评估性能，我们还展示了在基准挑战确定为包含的类别的测试视频的权衡曲线：显著目标变形，图 4；严重的光照变化图。5；和闭塞，图 6。我们的跟踪显著优于高度变形目标跟踪的其他方法。我们赋予聚集决策树方法网上再学习模型的灵活性。跟踪器还实现了在光照变化和遮挡的情况有竞争力的结果。我们的属性在光照下更改交叉约束的多层次目标模型的稳健表现强劲。我们的属性闭塞测试了该方法的通用性和适应性的结果。当一个方法被设计为针对戏剧性靶的外观和形状的变化鲁棒，它可能不总是可能的外观变化和遮挡区分，因此，我们的方法牺牲赞成健壮性在这种情况下一些精度。

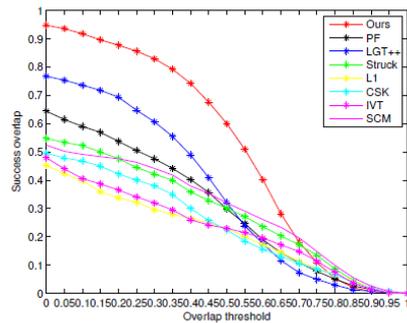


Figure 4. The success ratio versus overlap threshold curve in 19 sequences with deformation.

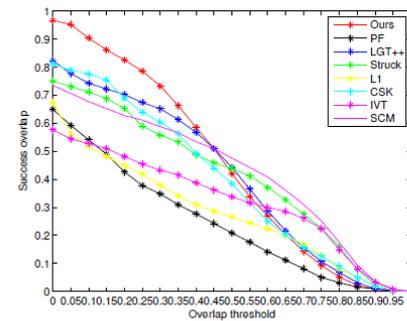


Figure 5. The success ratio versus overlap threshold curve in 18 sequences with illumination change.

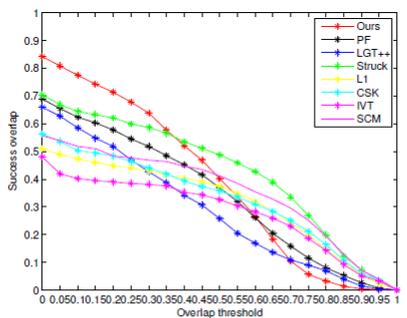


Figure 6. The success ratio versus overlap threshold curve in 22 sequences with occlusion.

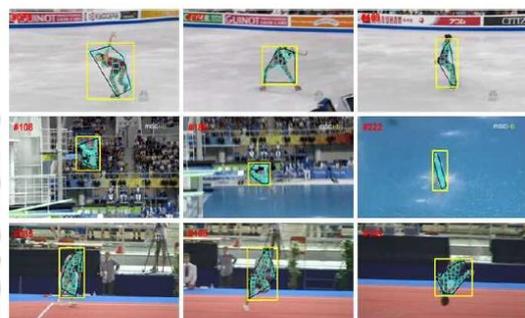


Figure 7. Visualization of results on sequences: Iceskating, Diving, Gymnastics, showing extreme target deformations and significant clutter.

6. 总结

在本文中，我们提出了多层次的目标模式，并为匹配自适应聚集决策树方法和再学习中等级水平的目标部分在连续的图像帧。由此产生的跟踪算法显著优于从每个 VOT2013 和 VOT2014 基准跟踪挑战的最佳 4 的算法，并优于国家的最先进的 7 额外 CVPR2013 基准的跟踪数据集内跟踪器。该方法对付大的目标变形有挑战性跟踪的条件下，迅速光照变化，和闭塞尤为强劲。自适应决策树：1) 在线生成，克服离线分类产生的过拟合；2) 有效地利用仅需要在每帧中的每个目标部分的功能的最小数目；3) 有利于最挑剔的特点自适应的权重，动态响应不断变化的伪装量在不同的功能模式。未来的工作将评估纳入附加功能的方式，从而有望提高性能。

7. 致谢

- a) J. Xiao is supported by the China Scholarship Council
- b) and also by a University of Birmingham school scholarship.
- c) This work was supported in part by EU H2020 RoMaNS,645582, and EPSRC EP/M026477/1.

8. 参考资料

- [1] CVPR benchmark dataset. <https://sites.google.com/site/trackerbenchmark/benchmarks/v10>. 6
- [2] The VOT challenge. <http://www.votchallenge.net/>. 3, 6, 7
- [3] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *PAMI*, 34(11):2274–2282, 2012. 3, 4
- [4] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. Robust object tracking with online multiple instance learning. *PAMI*, 33(8):1619–1632, 2011. 3
- [5] Anil Kumar Bhattacharya On a measure of divergence between two statistical populations defined by their probability distributions. In *Bulletin of the Calcutta Mathematical Society*, issue 35, pages 99–110, 1943. 3
- [6] Paul Brasnett, Lyudmila Mihaylova, David Bull, and Nishan Canagarajah. Sequential monte carlo tracking by fusing multiple cues in video sequences. *Image and Vision Computing*, 25(8):1217–1227, 2007. 2
- [7] Zhaowei Cai, Longyin Wen, Jianwei Yang, Zhen Lei, and Stan Z Li. Structured visual tracking with dynamic graph. In *ACCV*, pages 86–97. Springer, 2013. 7
- [8] Robert T Collins, Yanxi Liu, and Marius Leordeanu. Online selection of discriminative tracking features. *PAMI*, 27(10):1631–1643, 2005. 2
- [9] Martin Danelljan, Gustav Håger, Fahad Shahbaz Khan, and Michael Felsberg. Accurate scale estimation for robust visual tracking. In *BMVC*, 2014. 7
- [10] François Ennesser and Gérard Medioni. Finding waldo, or focus of attention using local color information. *PAMI*, 17(8):805–809, 1995. 2
- [11] Michael Felsberg. Enhanced distribution field tracking using channel representations. In *ICCV visual object tracking workshop*, pages 121–128. IEEE, 2013. 7
- [12] Helmut Grabner and Horst Bischof. On-line boosting and

- vision. In *CVPR*, volume 1, pages 260–267. IEEE, 2006. [3](#)
- [13] Helmut Grabner, Christian Leistner, and Horst Bischof. Semi-supervised on-line boosting for robust tracking. In *ECCV*, pages 234–247. Springer, 2008. [3](#)
- [14] Sam Hare, Amir Saffari, and Philip HS Torr. Struck: Structured output tracking with kernels. In *ICCV*, pages 263–270. IEEE, 2011. [7](#)
- [15] Cher Keng Heng, Sumio Yokomitsu, Yuichi Matsumoto, and Hajime Tamura. Shrink boost for selecting multi-lbp features in object detection. In *CVPR*, pages 3250–3257. IEEE, 2012. [7](#)
- [16] Joao F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *ECCV*, pages 702–715. Springer, 2012. [7](#)
- [17] Seunghoon Hong and Bohyung Han. Visual tracking by sampling tree-structured graphical models. In *ECCV*, pages 1–16. Springer, 2014. [2](#)
- [18] Xu Jia, Huchuan Lu, and Ming-Hsuan Yang. Visual tracking via adaptive structural local sparse appearance model. In *CVPR*, pages 1822–1829. IEEE, 2012. [2](#)
- [19] Junseok Kwon and Kyoung Mu Lee. Highly nonrigid object tracking via patch-based dynamic appearance modeling. *PAMI*, 35(10):2427–2441, 2013. [2](#)
- [20] Yang Li and Jianke Zhu. A scale adaptive kernel correlation filter tracker with feature integration. In *ECCV visual object tracking workshop*, pages 254–265. 2014. [7](#)
- [21] Yang Lu, Tianfu Wu, and Song-Chun Zhu. Online object tracking, learning and parsing with and-or graphs. In *CVPR*, pages 3462 – 3469. IEEE, 2014. [2](#)
- [22] Iain Matthews, Takahiro Ishikawa, and Simon Baker. The template update problem. *PAMI*, 26(6):810–815, 2004. [2](#)
- [23] Xue Mei and Haibin Ling. Robust visual tracking using l1 minimization. In *ICCV*, pages 1436–1443. IEEE, 2009. [7](#)
- [24] Katja Nummiaro, Esther Koller-Meier, and Luc Van Gool. An adaptive color-based particle filter. *Image and vision computing*, 21(1):99–110, 2003. [2, 3, 4, 6, 7](#)
- [25] Federico Pernici and Alberto Del Bimbo. Object tracking by oversampling local features. *PAMI*, 36:2538 – 2551, 2013. [2, 3](#)
- [26] David A Ross, Jongwoo Lim, Rwei-Sung Lin, and Ming-Hsuan Yang. Incremental learning for robust visual tracking. *IJCV*, 77(1-3):125–141, 2008. [7](#)
- [27] Rustam Stolkin, Ionut Florescu, and George Kamberov. An adaptive background model for camshift tracking with a moving camera. In *6th International Conference on Advances in Pattern Recognition*, pages 147–151. Citeseer, 2007. [2, 3](#)
- [28] Rustam Stolkin, David Rees, Mohammed Talha, and Ionut Florescu. Bayesian fusion of thermal and visible spectra camera data for region based tracking with rapid background adaptation. In *IEEE Conference on Multisensor Fusion and*

Integration for Intelligent Systems, pages 192–199.

IEEE,

2012. [2](#)

[29] Rustam Stolkin and Mohammed Talha. Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data. *IEEE Sensors*, 2014. [2](#)

[30] Luka Čehovin, Matej Kristan, and Aleš Leonardis. Robust visual tracking using an adaptive coupled-layer visual model.

PAMI, pages 941 – 953, Apr 2013. [2](#), [3](#), [5](#), [6](#)

[31] Tomáš Vojtíšek and Jiří Matas. Robustifying the flock of trackers.

In *16th Computer Vision Winter Workshop*, pages 91–97,

2011. [7](#)

[32] Qing Wang, Feng Chen, Wenli Xu, and Ming-Hsuan Yang.

Online discriminative object tracking with local sparse representation.

In *WACV*, pages 425–432. IEEE, 2012. [2](#)

[33] Shu Wang, Huchuan Lu, Fan Yang, and Ming-Hsuan Yang.

Supapixel tracking. In *ICCV*, pages 1323–1330.

IEEE,

2011. [2](#)

[34] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object

tracking: A benchmark. In *CVPR*, pages 2411–2418.

IEEE,

2013. [3](#), [6](#), [7](#)

[35] Jingjing Xiao, Rustam Stolkin, and Aleš Leonardis. An enhanced adaptive coupled-layer LGTracker++. In *ICCV visual object tracking workshop*, volume 2, page 5. 2013.

[6](#),

[7](#)

[36] Wei Zhong, Huchuan Lu, and Ming-Hsuan Yang. Robust

object tracking via sparsity-based collaborative model. In

CVPR, pages 1838–1845. IEEE, 2012. [7](#)