目标跟踪算法的研究与分析

摘要：目标跟踪是计算机视觉的众多应用中最重要的组成部分之一。近年来，随着众多共享代码和数据集的出现，目标跟踪领域取得了很多重大进展，因此，对各类目标跟踪算法进行对比分析，建立一套评价标准来衡量各类目标跟踪算法的性能是非常重要的。本文简要回顾了在线目标跟踪的最新进展后，又进行了大规模的实验，并通过各种评价标准来了解这些算法的性能。测试图像序列用不同的属性注释以用于性能评估和分析。通过定量分析结果，本文确定了有效和鲁棒的跟踪算法，并且为这一领域未来的研究奠定了基础。

**1 绪论**

目标跟踪是计算机视觉的广泛应用中最重要的组成部分之一，例如智能监控、人机交互和医学成像。给定目标物体在一段视频的某一帧中的初始状态（例如：位置和大小），则目标跟踪就是估计出目标物体在后续视频帧中的状态。尽管目标跟踪已经被研究了几十年，而且今年来也取得了很多进展，但是目标跟踪仍然是一个非常有挑战性的课题。许多因素影响着目标跟踪算法的性能，例如光照变化、遮挡以及背景干扰，并且没有可以成功处理所有跟踪场景的单一跟踪方法。因此，对目前的跟踪器的性能进行研究分析，提取其优点与不足，对于在该领域的未来研究中设计更加鲁棒的算法是至关重要的。

对于算法综合性能的评估，收集具有代表性的数据是十分重要的。目前在监控场景中存在若干个用于视觉跟踪的数据集，例如VIVID、CAVIAR以及PETS数据集。然而，在监控场景中目标物体通常是行人或汽车等小型物体，而且背景一般是静态背景。虽然一些数据集对通用场景用边界框进行了标记，但是大部分是没有进行标记的。对于没有标记的地面实况序列，由于报告的结果基于不一致的标注的目标位置，因此很难用来评估跟踪算法。

近些年来，很多跟踪算法已经被公开，例如OAB、IVT、MIL、LI以及TLD算法，已经经常被用于跟踪算法的评价。但是，大多数跟踪器的输入和输出格式是不一样的，因此对大规模的性能评价不方便。本文中建立了一个包含大多数公开可用的跟踪器的代码库以及一个具有标注数据的测试数据集用于评估任务。此外，数据集中的每个序列用通常影响跟踪性能的属性注释，例如遮挡、快速运动以及照明变化。

评估跟踪算法的一个常见问题是，结果仅仅基于具有不同初始条件或参数的几个序列来报告。因此，评价结果不能体现算法的整体性能。为了公平和综合地进行算法性能评估，本文从标注的真实目标位置在空间和时间上扰乱初始状态。虽然鲁棒的初始化在此研究领域是众所周知的问题，但是这个问题在文献中很少涉及。众所周知，这是第一个全面的工作来解决和分析目标跟踪的初始化问题。我们使用基于位置误差度量的精度图和基于重叠度量的贡献图来对每种算法的性能进行分析。

这项工作的贡献有以下三项：

数据集：本文构建一个具有50个完全注释序列的跟踪数据集，以便于跟踪评估。

代码库：本文的代码库中整合了大部分的公开可用的跟踪器源码，并且归一化了输入输出格式以便进行大规模的性能评估。目前，它包含了29中跟踪算法。

鲁棒性评估：用于跟踪的初始化边界框在空间和时间上被采样以评估跟踪器的鲁棒性以及特性。每个跟踪器通过分析超过660000个边界输出框进行广泛评估。

这项工作主要集中在单个目标的在线跟踪。代码库，注释数据集和所有跟踪结果可在网站<http://visualtracking.net>上获得。

**2 相关工作**

在这个部分，本文回顾了最近的目标跟踪算法的几个主要模块：目标表示方法、搜索机制和更新模型。此外，已经提出了一些基于组合一些跟踪器或挖掘上下文信息的方法。

目标表示方法：目标描述是任何视觉跟踪器中的主要组件之一，并且已经给出了许多方案。自从Lucas和Kanade的开创性工作以来，整体模板（原始强度值）已被广泛用于跟踪算法。后来，基于子空间的跟踪方法以更好地考虑外观变化也被提出。此外，Mei和Ling提出了一种基于稀疏表示的跟踪方法来处理损坏的外观，并且最近进一步改进。除了模板，许多其他视觉特征已经被采用在跟踪算法中，例如颜色直方图、定向梯度直方图（HOG），协方差区域描述子，以及Haar-like特征。最近，判别模型也被广泛应用于目标跟踪中，其中在线学习二元分类器以将目标与背景区分开。大量的机器学习方法被适用于跟踪问题中，例如支持向量机（SVM）、结构化输出SVM、Ranking SVM、Boosting、Semi-Boosting和Multi-instance Boosting。为了使跟踪器更加鲁棒以应对光照变化以及部分遮挡，对象可以由其中每一个由描述符或直方图表示的部分表示。在几个局部直方图中用于表示预定义网格结构中的对象。Kwon和Lee提出了一种方法自动更新本地修补程序的拓扑以处理大的姿势更改。为了更好地处理外观变化，最近已经提出了关于集成多种表示方法的一些方法。

搜索机制：为了估计目标物体的状态，确定性或随机的方法已经被应用。当跟踪问题在优化框架内提出时，假设目标函数相对于运动参数是可微分的，梯度下降法可以用于有效地定位目标。但是，这些目标函数通常不是线性的而且包含很多局部最小值。为了缓解这个问题，已经以牺牲高计算负荷的方式采用了密集采样方法。另一方面，诸如粒子滤波器的随机搜索算法已经被广泛使用，因为它们对局部最小值和计算效率相对不敏感。

模型更新：更新目标表示或模型以考虑外观变化是至关重要的，Matthews 解决了Lucas-Kanade算法的模板更新问题，其中使用从第一帧提取的固定参考模板和来自最近帧的结果的组合来更新模板。有效的更新算法也已经通过在线混合模型提出，在线Boosting和增量子空间更新。对于辨别模型，最主要问题是改进样本采集模块，使在线训练分类器更加鲁棒。虽然已经取得了很多进展，但是仍然难以获得自适应外观模型以避免漂移。

上下文和融合的跟踪器：上下文信息对于跟踪同样十分重要。最近，已经提出了通过挖掘辅助对象或围绕目标的局部视觉信息来辅助跟踪的一些方法。当目标完全遮挡或离开图像区域时，上下文信息特别有用。为了提高跟踪性能，最近提出了一些跟踪器融合方法。Santner提出了一种结合静态、中度自适应和高度自适应跟踪器来解决外观变化的方法。甚至多个跟踪器或多个特征集在贝叶斯框架中被提及和选择以更好地考虑外观改变。

**3 评估算法和数据集**

为了进行公平的评价，我们测试其原始源代码或二进制代码公开可用的跟踪算法，因为所有实现不可避免地涉及技术细节和特定参数设置。表一展现了跟踪算法的评估结果。本文也评估了VIVID测试平台中的跟踪器包括Mean-Shift、模板匹配、Radio-Shift和峰值差的方法。

近年来，许多基准数据集已被开发用于各种视觉问题，例如伯克利segmentation,FERET人脸识别和光流数据集。同在在监控场景中存在一些用于跟踪的数据集，例如VIVID和CAVIAR数据集。为了实现通用的视觉跟踪，更多的视频序列被用于评估。然而，大多数视频序列没有基准数据标注，并且可以用不同的初始条件产生定量评估结果。为了使算法性能评估更加公平，本文收集并注释了最常用的跟踪序列。图一展现了每个视频序列在第一帧时目标物体被初始化的边界框。

测试序列的属性：准确地评估一个跟踪器的性能是比较困难的，因为其性能可能受到多种因素的影响。为了更好地评估和分析跟踪方法的优点和缺点，本文通过用表2中所示的11个属性注释来对序列进行分类。

我们的数据集中的属性分布如图2（a）所示。一些属性更频繁地出现，例如OPR和IPR相对于其他属性更加频繁地出现。同时还显示了一个序列通常用几个属性注释。除了总结整个数据集的性能，我们还构建了对应于属性的几个子集以体现具体的挑战性条件。例如，OCC子集包含了29个序列来分析跟踪器解决遮挡问题的性能表现。OCC子集中的属性分布如图2（b）所示，其他属性在补充材料中可用。

|  |  |
| --- | --- |
| Attributes | Description |
| IV | 光照变化—目标区域中的光照显着改变 |
| SV | 尺度变化—第一帧和当前帧的边界框比率在范围[1/ts:ts]之间，ts>1(ts=2) |
| OCC | 遮挡 - 目标被部分或完全遮挡 |
| DEF | 变形 - 非刚性物体变形 |
| MB | 运动模糊 - 目标区域由于目标或相机的运动而模糊 |
| FM | 快速运动 - 地面实况的运动大于tm像素（tm = 20）（帧率大于20fps） |
| IPR | 平面内旋转 - 目标在图像平面中旋转 |
| OPR | 平面外旋转 - 目标旋转出图像平面 |
| OV | 离开视野 - 目标的一部分离开视图 |
| BC | 背景干扰 - 目标附近的背景具有与目标类似的颜色或纹理 |
| LR | 低分辨率 - 地面实况边界框内的像素数小于tr（tr = 400）。 |

**4 评价方法**

在这项工作中，本文使用精度和跟踪成功率进行定量分析。此外，本文通过两个方面来评估跟踪算法的鲁棒性。

精度图：一种广泛使用的关于跟踪精度的评估度量是中心位置误差，其被定义为被跟踪目标的中心位置与手动标记的地面实况之间的平均欧氏距离。然后，使用一个序列的所有帧上的平均中心位置误差来概括该序列的总体性能。但是，当跟踪器丢失目标时，输出位置可以是随机的，并且平均误差值可能不能正确地测量跟踪性能。最近，采用精度图来测量总体跟踪性能。它显示其估计位置在地面实况的误差在给定阈值距离内的帧的百分比。作为每个跟踪器的代表性精确得分，我们使用阈值= 20像素的得分。

跟踪成功率图：另一个评估度量是边界框重叠。给定跟踪得到的边界框rt和基准边界框ra，重叠分数定义为，其中表示两个区域的交集和并集，……表示重叠区域中像素的数量。为了测量帧序列上的性能，我们统计其重叠S大于给定阈值t0的成功帧的数目。跟踪成功图表示在阈值从0到1变化的成功帧的比率。在特定阈值（例如to = 0.5）处使用用于跟踪器评估的一个成功率值可能不公平或不具有代表性。相反，我们使用每个成功图的曲线下面积（AUC）来对跟踪算法进行排名。

鲁棒性评价：评估跟踪器的常规方式是在初始化第一帧中的地面实况位置的整个测试序列中运行它们，并报告平均精度或成功率。我们称之为一次通过评估（OPE）。然而，跟踪器可能对初始化敏感，并且其在不同起始帧处具有不同初始化的性能可能变得更差或更好。因此，我们提出两种方法来分析跟踪器的鲁棒性初始化，通过在时间上扰动初始化（从不同的帧开始）和从空间上扰动初始化（从不同的初始边界框开始）。这些测试分别被称为时间鲁棒性评估（TRE）和空间鲁棒性评估（SRE）。

所提出的测试场景在现实世界的应用中发生很多，因为跟踪器通常由对象检测器初始化，这可能在位置和尺度方面引入初始化误差。另外，对象检测器可以用于在不同的时间实例重新初始化跟踪器。通过研究鲁棒性评估中的跟踪器的特性，可以对跟踪算法进行更透彻的理解和分析。

时间鲁棒性评估：给定一个初始帧与目标的地面实况边界框，一个跟踪器被初始化并运行到序列的末尾，例如，整个序列的一个片段。在每个片段上评估跟踪器，并计算整体统计数据。

空间鲁棒性评估：本文通过移动或缩放地面实况来对第一帧中的初始边界框进行采样。这里，我们使用8个空间移位，包括4个中心移位和4个角位移，和4个比例变化（补充）。移动量是目标尺寸的10％，并且尺度比在0.8,0.9,1.1和1.2之间变化与地面实况。 因此，我们评估每个跟踪器12次SRE。

**5 评价结果**

对于每个跟踪器，带有默认参数的源代码将用于所有评估。表1列出了在使用Intel i7 3770 CPU（3.4GHz）的PC上运行的OPE中每个跟踪器的平均FPS。更多详细的速度统计数据，如最小和最大，可在补充。

对于OPE，每个跟踪器测试超过29,000帧。对于SRE，每个跟踪器在每个序列上评估12次，其中生成超过350,000个边界框结果。对于TRE，每个序列被分割成20个段，因此每个跟踪器在大约310,000帧上执行。据我们所知，这是视觉跟踪的最大规模的性能评估。在这份手稿中报道最重要的发现，更多的细节和数字可以在补充中找到。

**5.1 整体表现**

所有跟踪器的总体性能由成功和精度图概括，如图3所示，其中为了清楚仅示出前10个算法，并且在补充材料中显示完整的图。对于成功图，我们使用AUC分数来总结和排名跟踪，而对于精度图，我们使用误差阈值为20的结果进行排名。在精度图中，一些跟踪器的排名与成功图中的排名略有不同，因为它们基于测量跟踪器的不同特性的不同度量。因为成功图的AUC分数测量比在图的一个阈值处的分数更准确的总体性能，在下面我们主要分析基于成功图的排名，但使用精度图作为辅助。

平均TRE性能高于OPE，其中帧的数量从TRE的第一段到最后段减少。由于跟踪器趋向于在较短的序列中表现良好，TRE中的所有结果的平均值倾向于更高。另一方面，SRE的平均性能低于OPE。初始化错误倾向于导致跟踪器用不精确的外观信息更新，从而导致逐渐漂移。

跟踪成功率图表中，OPE中排名最高的跟踪器SCM的表现优于Struck 2.6％，但比SRE中的Struck低1.9％。结果还表明，OPE不是最佳性能指标，因为OPE是SRE或TRE的一个试验。TRE中TLD的排名低于OPE和SRE。这是因为TLD在具有重新检测模块的长序列中表现良好，而在TRE中存在大量短段。Struck在TRE和SRE中的成功图表明，当重叠阈值小时，Struck的成功率高于SCM和ALSA，但是当重叠阈值大时小于SCM和ALSA。这是因为Struck只估计目标的位置，不处理尺度变化。

稀疏表示在SCM，ASLA，LSK，MTT和L1APG中使用。这些跟踪器在SRE和TRE中表现良好，这表明稀疏表示是考虑外观变化的有效模型。我们注意到，SCM，ASLA和LSK优于MTT和L1APG。结果表明，局部稀疏表示比具有整体稀疏模板的表示更有效。ASLA的AUC分数小于从OPE到SRE的其他前5个跟踪者的AUC分数，并且ASLA的排名也增加。它表明ASLA采用的对齐池技术对于失准和背景杂波更加鲁棒。

在前10个跟踪器中，CSK具有最高速度，其中所提出的循环结构起着关键作用。VTD和VTS方法采用混合模型来提高跟踪性能。与其他更高排名的跟踪者相比，它们的性能瓶颈可归因于它们基于稀疏主成分分析所采用的表示，其中使用整体模板。由于空间限制，在以下部分中提供了SRE的图以供分析，并且补充中包括更多结果。

**5.2 基于属性的性能分析**

通过注释每个序列的属性，我们构建具有不同主导属性的子集，这有助于分析跟踪器的性能对于每个挑战性因素。由于空间的限制，我们只说明与如图4分析SRE的成功图和精度图测试属性为OCC，SV和FM，并且在补充材料中呈现更多的结果。

当对象快速移动时，基于密集采样的跟踪器（例如，Struck，TLD和CXT）执行得比其他更好。一个原因是搜索范围大，并且辨别模型能够将目标与背景杂波区分开。然而，由于较差的动态模型，具有高整体性能的基于随机搜索的跟踪器（例如，SCM和ASLA）在该子集中不能很好地执行。如果这些参数设置为大值，需要更多的粒子来使跟踪器稳定。这些跟踪器可以用具有更有效的粒子滤波器的动态模型进一步改进。

在OCC子集上，Struck，SCM，TLD，LSK和ASLA方法优于其他方法。结果表明，结构化学习和局部稀疏表示有效处理遮挡。在SV子集上，ASLA，SCM和Struck执行得最好。结果显示，具有仿射运动模型（例如，ASLA和SCM）的跟踪器通常比被设计成仅考虑具有少数例外（例如Struck）的平移运动的其他跟踪器更好地处理尺度变化。

**5.3 不同尺度的初始化**

已知跟踪器通常对初始化变化敏感。图5和图6显示了在不同尺度下初始化的总结跟踪性能。当计算重叠分数时，我们重新调节跟踪结果，使得性能概要可以与原始尺度，即图3中的OPE的图相当。图6示出了对于每个比例的所有跟踪器的平均性能，其示出了当比例因子较大时（例如，α1:2），性能通常显着减小，因为许多背景像素不可避免地包括在初始表示中。TLD，CXT，DFT和LOT的性能随着初始化规模的增加而降低。这表明这些跟踪器对背景杂波更敏感。一些跟踪器在比例因子较小时表现更好，，例如LIAPG，MTT,LOT和CPF。在L1APG和MTT的情况下这一点的一个原因是模板必须被扭曲以适合通常较小的规范模板的大小，使得如果初始模板小，更多的外观细节将保持在模型中。另一方面，当初始边界框被放大时，一些跟踪器表现良好或甚至更好，例如Struck,OAB,SemiT和BSBT。这表明由于计算特征时的求和操作，Haar特征对于背景杂波有些鲁棒。总的来说，Struck对尺度变化不如其他性能良好的方法敏感。

**6 结语**

在本文中，我们进行大规模实验来评估最近在线跟踪算法的性能。根据我们的评估结果和观察，我们强调一些跟踪组件，这对于提高跟踪性能是必不可少的。 首先，背景信息对于有效跟踪至关重要。它可以通过使用高级学习技术来隐式地（例如，Struck）在辨别模型中对背景信息进行编码来利用，或者明确地用作跟踪上下文（例如，CXT）。第二，与整体稀疏表示（例如，MTT和L1APG）相比，局部模型对于如本地稀疏表示（例如，ASLA和SCM）的性能改进中所示的跟踪是重要的。当目标的外观部分改变时，它们特别有用，例如部分遮挡或变形。第三，运动模型或动态模型对于对象跟踪至关重要，特别是当目标的运动较大或突然时。然而，我们大多数评估的跟踪器不关注这个组件。基于动态模型的良好的位置预测可以减少搜索范围，从而提高跟踪效率和鲁棒性。改进这些组件将进一步提高在线对象跟踪的技术水平。

评估结果表明，在过去十年中，在对象跟踪领域取得了重大进展。我们从几个角度提出和演示了对跟踪算法的深入分析的评估指标。这种大规模的性能评估有助于更好地理解最先进的在线对象跟踪方法，并提供用于测量新算法的平台。我们正在进行的工作重点是扩展数据集和代码库，以包含更多完全注释的序列和跟踪器。