

指导教师： 杨涛

提交时间： 2016/3/20

CVPR2015

Paper Translation

No: 01

姓名： 安伟

学号： 2013302595

班号： 10011306



结构稀疏跟踪

Tianzhu Zhang^{1;2} Si Liu³ Changsheng Xu^{2;4} Shuicheng Yan⁵

Bernard Ghanem^{1;6} Narendra Ahuja^{1;7} Ming-Hsuan Yang

¹ 高级数字科学中心 ² 中科院自动化研究所 ³ 中科院信息工程研究所
⁴ 中国-新加坡数字媒体学院 ⁵ 新加坡国立大学 ⁶ 阿卜杜拉国王科技大学
⁷ 伊利诺伊大学香槟分校 ⁸ 加州大学莫塞德校区

摘要

稀疏表示已通过使用目标模板搜索重构误差最小的最佳目标候选人应用到视觉跟踪。然而,大多数稀疏表示都基于只考虑整体或局部表示,不充分利用目标候选人及其内部固有结构的追踪器,从而在类似对象出现或遮挡情况下不那么有效。在本文中,我们提出一种新颖的结构稀疏跟踪(SST)算法,它不仅利用目标候选人和局部区块之间的内在关系共同学习稀疏表示,还保留了局部区块里每个目标候选人的空间布局结构。我们将展示出我们的结构稀疏跟踪算法适应大多数现有的稀疏追踪器及其具有的个别优点。通过对具有挑战性的基准图像序列的定性和定量评估,我们证明出结构稀疏追踪算法比其他几种先进算法计算效果更好。

1. 简介

视觉跟踪的目的是估算视频内某个移动目标的状态。它长期以来都是计算机视觉中最重要和最基本

的话题,并且有着大量的应用,如监控、车辆导航、人机界面、人类的运动分析,等等。尽管最近几年里许多对象跟踪方法被提出,针对复杂和动态场景开发出一个鲁棒算法仍然是一个具有挑战性的任务,这是因为诸如部分遮挡、照明、姿势、规模、相机运动、背景杂斑和视角等因素的影响。

大多数跟踪算法都是从识别和生成的角度发展的。判别方法是制订一个类似二进制分类任务的追踪问题,从而找到从背景中分离目标物体的判定边界。Avidan 将一组弱分类器组合成一个强分类器并提出一种整体跟踪方法。柯林斯等人证明将目标物体从背景中分离可以找到最明显的差异特征。Grabner 等人提出一种在线增加方法来更新差异特征,同时提出一种半在线增加算法来处理对象跟踪中的漂移问题。Babenko 等人在样本被认为是正包或负包时将多个实例学习引入到在线跟踪对象中。Kalal 等人提出 pn 学习算法来利用正负样本的底层结构

作为对象追踪的学习分类器。

与之相反,通常生成的跟踪方法先学习一个模型来表示目标对象,然后使用模型来搜索重建误差最小的图像区域。Black 等人利用一个离线子空间模型来表示跟踪的兴趣对象。均值漂移跟踪算法模拟一个有非参数分布的特性(如目标的颜色像素)和定位对象模式转变的目标。Frag 追踪通过建模对象外观与局部区块的直方图解决了部分遮挡的问题。早期诊断方法利用增量子空间模型来调整外观变化。Kwon 等人使用多个观测模型涵盖了范围广泛的在跟踪过程中由姿势和光照变化引起的外观变化。大多数这类方法使用整体的表现来描述对象,因而不能

很好地处理遮挡或错误。

最近,基于稀疏表示的生成跟踪方法已经在目标跟踪方面有了发展。如图 1 所示,这些追踪器可以在表示方案的基础上分类成全局的、局部的、以及联合的稀疏外观模型。稀疏追踪器将每个目标候选人 x_i 表示为一个能根据外观变化动态更新的目标模板 T 的稀疏线性组合。这些模型已被证明能够可靠的解决追踪中的局部遮挡问题。然而,正如图 1(a)所示,所有这些方法都模拟一个目标对象作为单一的实体,并采用全局稀疏的外观模型。因此,这些方法在处理严重的遮挡不是十分有效。

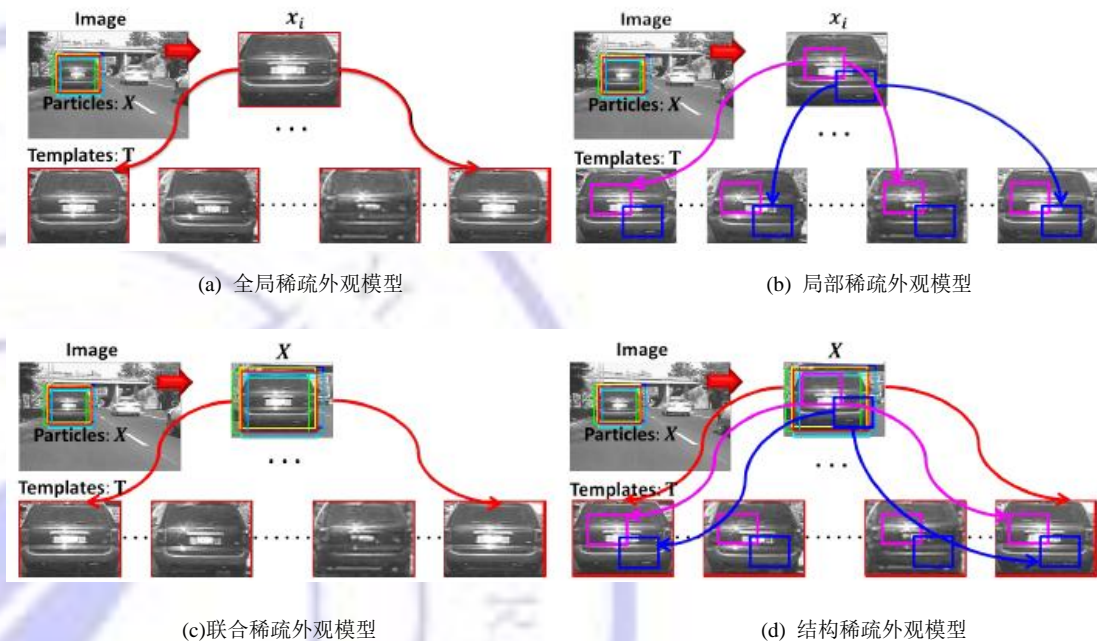


图1. 基于稀疏表示的追踪器。这些方法根据稀疏外观模型进行分组。给定一个有 n 个采样粒子 $X = [x_1, \dots, x_i, \dots, x_n]$ 的图像和字典模板 T 。(a)全局稀疏外观模型。这些追踪器采用目标的整体表现作为外观模型并通过解决 e_1 最小化问题进行追踪。因此,目标候选人 x_i 由 T 中一个稀疏的元素数量表示。(b)局

部稀疏外观模型。这些追踪器通过 T 中局部区块的一个稀疏线性组合表示每个局部区块里面一个可能的目标候选人 x_i 。注意,目标候选人 x_i 内部的局部区块可能被其他字典模板内相配的局部区块稀疏表示。(c)联合稀疏外观模型。这些追踪器利用粒子 X 间的内在关系研究它们的联合稀疏表示。联合稀疏约束支持所有粒子表征成为联合稀疏的并且共享相同的(一些)能可靠代表它们的字典模板。(d)我们提出的结构稀疏外观模型将上述三个模型结合在了一起。我们的模型利用了粒子 X 之间和局部区块之间的内在关系来研究他们的联合稀疏表示。此外,我们的方法也在每个目标候选人内部的局部区块保留了空间布局结构,而这正是上述三个模型所忽略的。使用我们的模型,所有粒子 X 和它们的局部区块都会用联合稀疏表示,即:只有少数(但相同)字典模板是用来表示每一帧所有的粒子和它们的局部区块。注意,所有粒子 X 里面的局部区块都通过用表示 X 的相同字典模板内的相同的局部区块来联合稀疏表示。

局部稀疏的外观模型的基本思路如图 1 所示(b),可能的目标候选人 x_i 内部的局部区块用字典模板内的局部区块来稀疏表示。Liu 等人提出一种基于局部稀疏模型的跟踪算法,该算法采用直方图的稀疏系数和均值漂移算法来进行对象跟踪。然而,这种方法是基于静态局部稀疏字典的并且当类似的物体出现在场景中可能失败。Jia 等人开发了一个基于局部稀疏外观模型的跟踪方法,该方法使用目标区域内一组重叠的图像区块及其空间布局。这些局部区块用于形成一个可能的目标候选人内部的编码区的字典。对一个目标候选人来说,其局部区块可以用同样的方式提取。因为每个局部区块代表一个目标对象的某个固定部分,所有的局部区块可以代表完整的目标结构。在稀疏设想下:目标候选人内的局部区块可以用少数基于解决 e_i 最小化问题的字典的线性组合来表示。尽管这个模型解决了全局稀疏外观模型的一些

问题,这种跟踪算法不考虑每个目标候选人内部的局部区块间的空间布局结构和所有目标候选人与局部区块间的相关性。例如,如图 1(b)所示,目标候选人 x_i 内部的局部区块也行能用其它字典模板的局部区块来稀疏。理想情况下,为了保持局部区块间的空间布局, x_i 的紫色局部区块最好在第一个字典的基础上用局部区块来表示,并且 x_i 的蓝色局部区块也应该用第一个词典的相同的蓝色局部区块来表示。

在粒子基于过滤器的跟踪方法中,粒子在目标和上或目标周围是随机抽样的,这种随机抽样采用基于之前状态的零均值高斯分布。每个粒子与其他粒子共享依赖,并且它们的对应影像点可能会是相似的。研究每个粒子的表征被视为一个单独的任务,一个关于所有粒子学习采用联合稀疏的多任务被采用了。低阶稀疏研究应用于学习所有粒子的稀疏表示。多任务多视点的联合稀疏表示被应用于跟踪。基于联合稀

疏外观模型的方法旨在利用粒子之间的内在关系提高跟踪性能。此外，由于所有粒子 X 之间的联合优化，这个模型计算很有效率。然而，这些模型仍然使用整体表象来描述对象的外观。

基于上述三个模型，我们提出一种新颖的结构稀疏外观模型如图 1(d) 所示，具有以下优点。首先，提出的结构稀疏外观模型将三个模型结合在一起，对部分遮挡不太敏感，通过考虑目标之间的相关性计算效率很高。其次，该模型利用粒子之间 X 和相应的局部区块之间的内在联系来研究他们的稀疏表示。第三，该模型保存了每个目标候选人内的局部区块的空间布局结构，这是前面的稀疏追踪器所忽略的。如图 1(d) 所示，因为所有的粒子 X 和它们的局部区块用联合稀疏表示，只有少数(但相同的)字典模板是用来代表在每一帧下的粒子及它们的局部区块。请注意：所有的粒子 X 内的局部区块用相同的字典模版内的相同的局部区块代表 X 来联合表示。

基于结构稀疏外观模型，我们提出一种在粒子滤波框架内计算高效的结构稀疏的跟踪 (SST) 算法。在这里，所有粒子和它们的局部区块通过提出的结构稀疏外观模型来表示，并且下一个目标状态是自身及局部区块与目标字典模板有最高相似度的粒子。与之前的方法不同，我们提

出的结构稀疏跟踪算法不仅利用了粒子和它们的局部区块间的内在关系来研究它们的联合稀疏表示，还保留了每个目标候选人内的局部区块间的空间布局结构。这有助于更准确地目标定位和对部分遮挡不敏感。在我们的结构稀疏跟踪说明，我们使用 “ p, q mixed-norm 调整，优化使用加速近端梯度 (APG) 方法的快速收敛性。此外，我们表明，现有的 e_1 追踪、LST 和 MTT 方法海温公式的特殊情况。



图 2. 局部区块采样的空间布局。注意，其他局部区块抽样方法也可以采用。

2. 结构稀疏跟踪

在本节中，我们详细描述我们的基于用外观模型表示粒子和它们的局部区块联合的跟踪方法的粒子滤波器。粒子在早先的对象位置及其附近采样来预测在时间 t 目标的状态 S_t ，从当前图像开始我们发现感兴趣区域 Y_t 并标准化模板大小。状态转移函数函数 $p(S_t | S_{t-1})$ 由一个仿射运动模型斜高斯分布模拟。观察模型 $p(Y_t | S_t)$ 反映了观察到的图像区域 Y_t 到粒子 S_t 和当前字典模板之间的相似性。在这个工作中， $p(Y_t | S_t)$ 计算获得的重建误差的函数通过线性代表 Y_t 和局部区块使用

字典模板。最大化这个函数选择的粒子被跟踪目标在每一次实例。接下来,我们展示了如何使用结构稀疏外观型代表粒子和局部区块的细节。

2.1 结构稀疏外观模型

给定一组图像目标模板 $T =$

$[T_1, T_2, \dots, T_m]$, 我们通过空间布局采样每个目标区域内的局部区块 K 。为简单起见,空间布局如图 2 所示。注意,其他局部区块抽样方法也可以采用。抽样局部区块后,他们被用来形成一个字典对当任何候选人地区内的局部区块进行编码。对这些目标模版 m 中的一个 k 阶的局部图像区块,我们得到相应的字典模版 $D^k = [d_1^k, d_2^k, \dots,$

$d_m^k] \in R^{d_k \times m}$, 当 $k=1, \dots, K$ 时; K 是目标区域内采样的局部区块的数量, d_k 是 k 阶图像区块矢量的维度, m 是目标模版的数量。 D^k 中的每一列可以通过从 T 中观察提取灰度图像的 e_2 矢量归一化求出。每个局部区块代表目标的某个固定部分,因此所有的局部区块可以代表完整的目标结构。因为图像区块是从很多模板中收集的,这本字典提取不同模板的共性并能代表这些部分的各种形式。为了体现鲁棒性追踪中物体外观的变化,字典 T 逐步更新

模板。这项工作中的字典更新计划是相似的。每个目标模板 T 分配重量,表明代表模板。当一个模板被频繁的用来表示跟踪结果,它有更高的重量。当一组 T 不能很好的代表粒子,重量最低的目标模板由当前跟踪结果所取代。为了初始化目标模板 m ,我们在目标的初始位置及其周围采样等尺寸的区块。

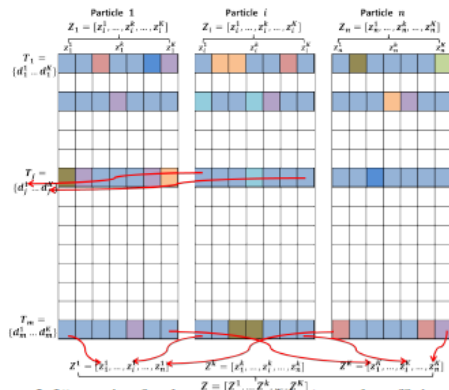


图 3. 说明了系数矩阵的结构 Z , 该结构不同的条目颜色代表不同的学习值, 白色的条目表明 0 行和 0 列。

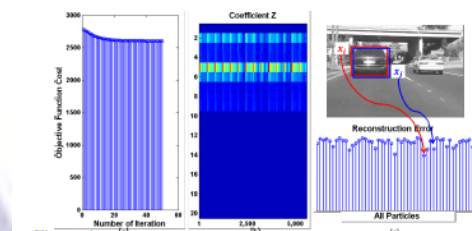


图 4. 用一个例子来说明我们的追踪工作。(a) 目标成本函数和迭代的数量。我们的算法能在几个迭代后收敛。(b)学习矩阵 $X \in R^{(20 \times 5600)}$ 。其中 $m = 20, K = 14, n = 400$ 。注意, Z 联合稀疏的列,即。一些(但相同)字典模板是用来代表在一起的所有图像区块。(c)粒子 x_i 根据跟踪的结果在所有其他的粒子选中,因为它有最小的重建误差。

在时间 t , 我们有 n 个采样粒子

和从矩阵 $X = [x_1, x_2, \dots, x_n]$ 观察得到的相应的矢量化灰度图像，观察到的第 i 个粒子用 x_i 表示。对目标候选人 x_i ，我们用它提取局部区块 K 并用和构造字典模版 D^k 相同的方式将它们转换成向量。对这些 n 个采样粒子的 k 阶局部图像区块，它们的相应的矢量化灰度图像是从矩阵 $X = [x_1, x_2, \dots, x_n] \in R^{d_k \times m}$ 中观察得到的。

我们从字典 D^k 观察到 X^k 的线性组合模板，例如 $X^k = D^k Z^k$ 。我们用 $Z^k = [z_1, z_2, \dots, z_n] \in R^{m \times n}$ 表示关于 D^k 的 k 阶局部区块。把表示所有 K 的局部区块放在一起，我们得到 $Z^k = [z_1, z_2, \dots, z_n] \in R^{m \times nK}$ 。对第 i 个粒子，它的局部区块的相应表征构成了一个矩阵 $Z_i = [Z_i^1, Z_i^2, \dots, Z_i^K] \in R^{m \times K}$ 。

对 Z^k 和 Z_i ，我们有如图 3 所示的发现：(1) Z^k 描述了所有的 n 个采样粒子的 k 阶图像区块。基于这些粒子在目标周围采样的事实，这些粒子可能是相似的，他们的 k 阶图像区块也能预测到是相似的。因此，应该利用底层的局部区块间的关系，而现有的基于局部稀疏表示的方法不考虑这些属性。(2) 它显示了对应的 i 阶粒子的局部区块的表征。因为这些局部区块是从目标候选人里面取样的，其空间布局结

构被保存了下来。

如图 2 所示，采样后，通过这些局部区块的空间布局，他们的表征应该满足以下约束：如果在第 i 个粒子内的 k 阶图像区块选择目标模版的第 k 个元素 $T_j = \{d_j^1, d_j^2, \dots, d_j^K\}$ 来表示 Z_i^k ，其他图像区块也应该在相同的目标模板 T_j 中选择相应的元素来表示。

处于以上的观察，我们使用凸的 $e_{p,q}$ 混合标准。特别地，我们使用 $e_{2,1}$ 来模拟 Z^k 和 Z_i 的结构并获得如下的对象追踪的结构稀疏外观模型：

$$\min_Z \frac{1}{2} \sum_{k=1}^K \|X^k - D^k Z^k\|_F^2 + \lambda \|Z\|_{2,1} \quad (1)$$

当 $Z = [Z^1, Z^2, \dots, Z^K] \in R^{m \times nK}$ ， $\|\cdot\|_F$ 表示弗罗贝尼乌斯规范， λ 是可靠的重建与联合稀疏正则化之间的权衡参数。 $e_{p,q}$ 混合标准的定义是 $\|Z\|_{p,q} =$

$$\left(\sum_i (\sum_j |Z_{i,j}|^p)^{\frac{q}{p}} \right)^{\frac{1}{q}}$$

和 $[Z]_{i,j}$ 表示 Z 第 i 行和第 j 列的条目。

图 3 说明了学习矩阵 Z 的结构。学习 Z 后，跟踪候选人 i 的观察似然度有如下定义： $p(y_t | s_t) = \frac{1}{\beta} e^{-\alpha \sum_{k=1}^K \|x_t^k - D^k z_i^k\|_F^2}$ ， z_i^k 是第 i 个目标候选人对应的 k 阶局部区块

的系数， α 和 β 是归一化参数。跟踪的结果是有最大观察可能性的粒子。

为了提出更直观的观点，我们设想一个学会了图 4 中的学习稀疏表示的实例，并展示出 SST 追踪器如何工作的例子。所有给定的粒子 X (在跟踪汽车周围采样)，它们的局部区块 ($X^k, k = 1, \dots, K$) 可以如图 2 所示基于空间布局进行采样。然后基于相应的字典模版 ($D^k, k = 1, \dots, K$)，我们可以学习矩阵 $Z = [Z^1, Z^2, \dots, Z^K]$ 。注意，一个更加明亮的色彩块 Z 表示相应条目中一个更大的值。显然， Z 的列是联合稀疏的，一些 (但相同) 的字典模版被用来表示所有的图像区块。确定粒子 x_i 作为当前跟踪结果 y_t 因为它的图像区块对目标模板的重建误差与其它粒子相比是最小的。由于粒子 x_j 对应于一个偏差幅目标，它有一个更大的重建误差并且不能很好的用 T 表示。

2.2 讨论

正如图 1 所介绍的，现有的跟踪方法可分为全局的、局部的、联合的基于表示形式的稀疏模型。出于上述三个模型，我们提出一种新颖的结构稀疏外观模型如图所示的进行对象跟踪。我们的方法是通用的，将上述三个模型组合在一起，

有自己的特点。值得强调的是 SST 算法和其它几个相关的跟踪方法之间的差异。

全局稀疏外观模型的追踪。当 $K=1$ (只有如图 2 所示的 $1*1$)，在采用 $e_{1,1}$ 的情况下，构想降低到由一个单一的实体模拟目标的全局稀疏外观模型，这种全局稀疏外观模型学习独立目标候选人的稀疏表示而不考虑它们的内在联系。

局部稀疏外观模型的追踪。通过图像区块采样方法和 $e_{1,1}$ 混合范数的使用，我们的构想降低到一个局部稀疏外观模型，这种局部稀疏外观模型不考虑多个目标候选人间图像区块的相关性和每个目标候选人内部图像区块的布局结构。

联合稀疏外观模型的追踪。当 $K=1$ (只有如图 2 所示的 $1*1$)，在采用 $e_{2,1}$ 的情况下，构想降低到一个联合稀疏外观模型，该联合稀疏外观模型考虑了目标候选人间的内在联系。然而，这种模型使用整体对象表示。

结构稀疏外观模型的追踪。我们的 SST 追踪器有如下特性：(1) 它考虑了全局的和局部的稀疏约束。(2) 它考虑到了目标候选人和它们的局部区块间的内在联系。

(3) 它考虑到了每个目标候选人内的图像区块的空间布局结构。

基于上述讨论，很明显我们的方法是通用的，大多数基于稀疏表示的跟踪方法是它的特殊情况。我们的构想不仅能维持现有三个模型的稀疏特性，还能利用每个目标候选人内图像区块的空间布局结构。

3. 实验结果

为了评估我们的跟踪器性能，我们在 20 日公开挑战图像序列中进行了广泛的实验研究。这些序列包含复杂的场景和视觉跟踪中的挑战性因素：例如：凌乱的背景下、移动相机、快速运动、大的构成和规模的变化、部分遮挡、形状变形和失真(参见图 5)。相比之下，我们运行 14 种目标初始位置相同的艺术算法。这些算法有在线多实例学习 (MIL)，在线 Adaboost 增加 (OAB)，通过检测跟踪 (TLD)，循环结构跟踪 (CST)，部分原因视觉跟踪 (PT)，实时压缩跟 (RTCT)， e_1 跟踪 (e_1T)，当地稀疏跟踪 (LST)，多任务跟踪 (MTT)，增量视觉跟踪 (溶)，分布领域跟踪 (DFT)，fragments-based (碎片弹)，和 local-global 跟踪 (LGT) 方法。其中，MIL、OAB、TLD、Struck、CST 和 PT 是有识别力的追踪器；其它的 (IVT、DFT、Frag、LGT、RTCT、 e_1T 、MTT、和 LST) 则是生成追踪器。此外，RTCT 和 e_1T 、LST 和 MTT 分别是基于全局、局部和联合的稀疏模型。通过比较这些不同的方法，我们证明

我们提出 SST 的有效性。为了公平的比较，我们使用由作者提供的公开的源代码或二进制代码，并将默认参数设置初始化。

对所有的实验报告，我们设置 $\eta = 0.1$ ， $\tilde{\lambda} = 0.5$ ，如图 2 所示，设置图像区块的数量 $K=14$ ，模版的数量 $m=20$ ，粒子的数量 $n=400$ 。粒子采样的仿射参数的方差设置为

(0.01, 0.0005, 0.0005, 0.01, 4, 4)。模版尺寸 d 被设置为目标对象第一次手动初始化尺寸的一半。该算法在 MATLAB 中实现并且在 2.80GHz 英特尔嵌件 8 GB 内存的双核机上运行时 0.45 秒每帧。我们将向公众开放源代码。

3.1 量化评估

为了定量评估每个跟踪器的性能，我们使用两个指标包括中心位置误差和重叠率。中心位置误差是中心的跟踪结果和每个框架的地面真值之间的欧几里得距离。重叠率是基于帕斯卡的挑战对象检测得分。给定跟踪边界框 ROI_T 和地面真值边界框 ROI_{GT} ，重叠的分数计算是通过

$$\text{分数} = \frac{\text{面积}(ROI_T \cap ROI_{GT})}{\text{面积}(ROI_T \cup ROI_{GT})} \text{得到的。}$$

对跟踪性能排名，我们在每个图像序列帧计算中心位置误差和重叠的平均分数。表 1 总结了这些 20 个图像序列结果。总的来说，提出的 SST 算法在所有测试序列中比其他先进的算法执行的更良好。由于空间限制，

平均中心位置误差和更多的实验结果以及视频在补充材料中可找到。

Video	SST	RTCT	IVT	MIL	OAB	Frag	Struck	MTT	e_1T	TLD	CST	DFT	LST	PT	LGT
tunnel	0.64	0.29	0.21	0.08	0.09	0.04	0.32	0.23	0.15	0.34	0.32	0.23	0.63	0.31	0.15
tud	0.87	0.32	0.56	0.38	0.56	0.68	0.61	0.67	0.84	0.71	0.36	0.67	0.44	0.57	0.24
trellis70	0.61	0.22	0.39	0.35	0.46	0.29	0.50	0.60	0.38	0.21	0.72	0.32	0.62	0.39	0.63
surfing	0.88	0.78	0.84	0.79	0.82	0.50	0.87	0.84	0.85	0.60	0.79	0.40	0.73	0.82	0.48
surfer	0.34	0.15	0.16	0.57	0.59	0.03	0.56	0.27	0.16	0.41	0.21	0.03	0.04	0.41	0.07
sphere	0.70	0.42	0.54	0.36	0.60	0.08	0.68	0.56	0.18	0.49	0.68	0.06	0.11	0.64	0.66
singer	0.78	0.45	0.48	0.41	0.18	0.26	0.46	0.86	0.70	0.40	0.47	0.47	0.73	0.46	0.29
girl	0.73	0.32	0.68	0.45	0.53	0.60	0.41	0.71	0.68	0.59	0.35	0.38	0.73	0.71	0.25
football	0.65	0.02	0.64	0.52	0.23	0.59	0.60	0.66	0.45	0.60	0.57	0.68	0.58	0.56	0.35
faceocc	0.76	0.73	0.84	0.58	0.77	0.87	0.85	0.84	0.86	0.57	0.92	0.91	0.30	0.87	0.57
faceocc2	0.73	0.54	0.79	0.72	0.59	0.38	0.77	0.74	0.67	0.57	0.77	0.78	0.77	0.77	0.46
david	0.60	0.41	0.36	0.42	0.43	0.23	0.38	0.53	0.50	0.60	0.50	0.57	0.45	0.64	0.58
carchase	0.87	0.29	0.44	0.53	0.82	0.60	0.85	0.58	0.59	0.80	0.84	0.40	0.79	0.72	0.31
car4	0.89	0.24	0.74	0.27	0.22	0.23	0.49	0.80	0.62	0.57	0.47	0.23	0.87	0.49	0.15
car11	0.77	0.00	0.51	0.22	0.55	0.10	0.83	0.80	0.52	0.28	0.80	0.52	0.79	0.82	0.43
biker	0.68	0.45	0.31	0.43	0.44	0.27	0.38	0.44	0.39	0.30	0.45	0.27	0.39	0.37	0.42
bicycle	0.59	0.33	0.32	0.54	0.31	0.11	0.40	0.64	0.29	0.39	0.25	0.25	0.54	0.28	0.35
human	0.78	0.33	0.66	0.48	0.54	0.47	0.53	0.65	0.73	0.08	0.53	0.31	0.74	0.49	0.23
osow	0.92	0.56	0.83	0.56	0.71	0.77	0.81	0.89	0.91	0.65	0.81	0.82	0.90	0.80	0.54
olsr	0.81	0.29	0.44	0.35	0.47	0.27	0.50	0.76	0.78	0.28	0.46	0.40	0.34	0.50	0.29

表 1. 15 种不同的追踪器在 20 个不同的视频的平均重叠分数。平均而言,提出的 SST 追踪优于其他 14 个最先进的追踪器。每个视频中最小的和第二个最小的距离用红色和蓝色分别表示。

基准的比较结果如图 6 所示。结果表明,我们的 SST 追踪比其他相关稀疏的追踪器有更良好的性能。与其他的方法相比,我们的 SST 有第二好的整体性能。如图 7 所示,我们注意到在背景杂乱,光照变化,低分辨率和基于精确度量遮挡属性表现良好。同样,结构稀疏追踪算法对有上面属性的视频达到良好的结果,成功率指标如图 8 所示。

3.2 定性评估

图 5 显示了 15 个追踪器在 20 日图像序列的一些跟踪结果。下面讨论的跟踪结果将基于每个视频中的主要挑战因素。

遮挡。在 olsr 序列中,所有其他的追踪器在 200 帧时丢失跟踪目标的轨迹因为她被一个男人遮挡住了。当其他追踪器锁定哪个男人的时候,如图所示图 5,其余序列的错误增加。在立筋序列,目标车辆被

过路的行人遮挡。MIL、OAB、CST、LST、Struck 和 RTCT 算法在遮挡时会从目标对象上渐渐远去。另一方面, e_1T 、DFT、Frag 以及提出的 SST 算法在这个序列表现良好。在其它序列的遮挡中,osow 和提出的 SST 表现至少是第二好的。

规模变化。人类和球的视频包含重大的规模变化。在人类的序列中,一个人远离相机并在极短时间内被遮挡。TLD 方法从一开始就失去跟踪的目标,RTCT 在 1674 帧漂走,DFT 和 LGT 追踪器在 1600 帧开始对目标漂移,最后失去了跟踪的目标。所有其他的方法都能成功跟踪目标,但 SST、 e_1T 、LST 和 IVT 方法实现更高的分数。在球序列中,球体的规模显著变化。虽然大多数器跟踪球失败,但 SST、CST 和 Struck 方法可以在整个序列跟踪目标。



图 5. 15 个追踪器(用不同颜色和线条表示)在 20 日图像序列上的跟踪结果。每个图左上角的黄色显示了帧索引。有关详细信息,请参阅文本。结果最好在高分辨率显示器上观看。

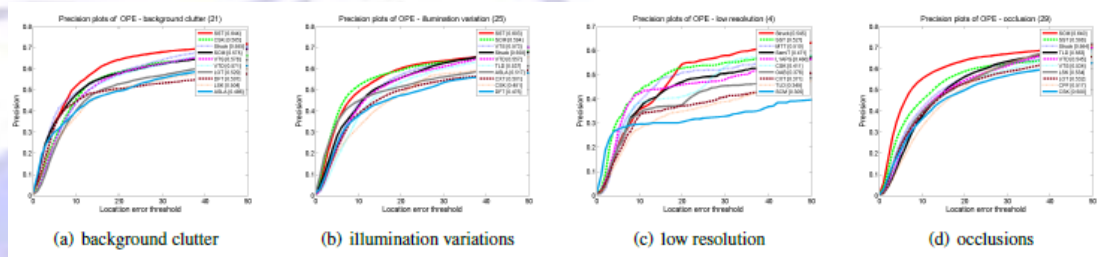


图 7. 基于精度指标的块开放属性。

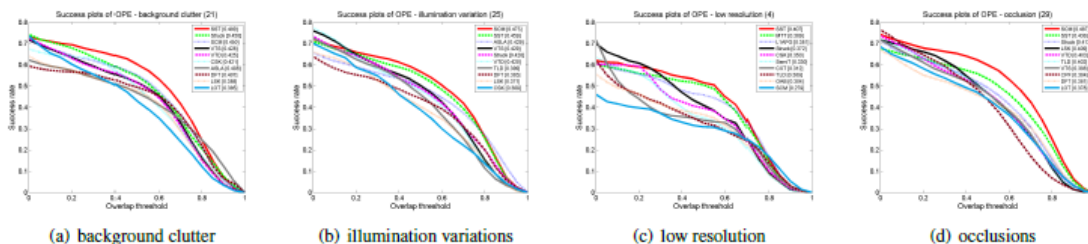
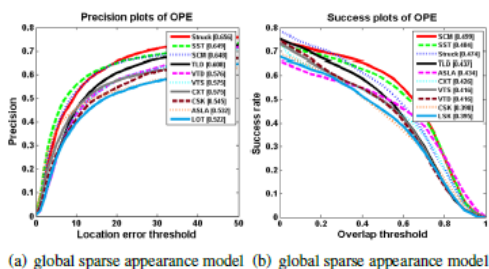


图 8. 使用成功率指标的块开放属性。



突然运动。在足球和隧道序列，杂乱背景的目标对象接受突然的运动。在足球序列，几个类似于目标的对象出现在镜头。OAB、和 LGT 算法分别在 100 帧、246 帧和 276 帧漂移。当近距离出现时(如。362 帧)，大多数方法从预定目标漂移到其他对象。提出的 SST、TLD、DFT 和 MTT 方法成功地跟踪整个序列的目标。

照明和规模的变化。car4、car11 和歌手图像序列包含大型照明和规模的变化。对 car4 序列，OAB、Frag 和 LGT 方法从 185 帧目标车辆经过天桥的时候开始漂移。MIL、DFT 和 RTCT 算法在 210 帧漂移到远离目标对象。 e_1T 、PT、CST、Struck 和 TLD 方法能够在大多数帧跟踪目标。另一方面，尽管有大型照明的变化，目标对象被 LST、IVT、MTT 和 SST 算法成功在整个序列跟踪。

遮挡和姿势变化。faceocc2 和自行车序列包含重要的遮挡和姿势变化。faceocc2 序列显示大多数追踪器在严重遮挡时漂移远离目标。因为提出的 SST 算法利用结构稀疏外观模型解决遮挡问题，它在这个序列执行得很好。自行车序列包含了有严重遮挡和姿势变化的场景。The SST、LST、Struck 和 MTT 方法有在整个序列跟踪目标的能力。其他追踪经验在不同时间实例漂移 (Frag 在 40 帧， e_1T 在 132 帧，DFT 在 135 帧，MIL 和 CST 在 144 帧，其他追踪器由于遮挡在 177)。

突然运动、姿势变化和遮挡。女孩和 carchase 序列包含有突然运动，姿势变化和遮挡的对象。在女孩的序列，SST、LST、MTT 和 PT 方法能够在整个序列跟踪目标，而其他方法逐渐慢慢漂移远离。在 carchase 序列，DFT 方法在 139 帧开始漂移。IVT、LGT 和 MTT 算法在 169 帧遮挡发生时漂移。OAB、CST、TLD、Struck 和 SST 方法在整个序列表现良好。大多数其他的追踪器由于突然运动和姿态变化都会有漂移问题。

4. 结论

在本文中，我们提出一个新颖的用来对象跟踪的结构稀疏外观模型，在粒子滤波框架下，目标候选人和他们图像区块的表征，正规化的稀疏 $e_{2,1}$ ，混合标准，都是使用一个有效的加速梯度共同学习 (APG) 方法来共同学习的。我们表明，该模型不仅利用目标候选人之间的内在关系，也利用相应的局部区块来研究它们的联合稀疏表示。此外，提出的模型保留了每个目标候选人内局部区块的空间布局结构。该算法一般可以适应大多数现有的稀疏追踪器。对几种先进的方法在挑战性的图像序列上的实验结果与评估证明了提出的 SST 跟踪算法的有效性和鲁棒性。在未来，我们将通过考虑噪声、随机抽样等方法对对象跟踪扩展结构稀疏外观模型。

感谢

这项研究是由了来自新加坡的科学、技术和研究机构——人类第六感研究资助项目的高级数字科学中心的研究资金支持。C. Xu 由 973 计划项目 No. 2012CB316304 和国家自然科学基金委 61225009, U1435211, 61432019, 61303173, 173211kysb20130018 支持。M.-H. Yang 的部分支持由 NSF 事业拨款# 1149783 和 NSF 互联网信息拨款# 1152576 提供。

参考文献

- [1] A. Adam, E. Rivlin, and I. Shimshoni. Robust fragmentsbased tracking using the integral histogram. In CVPR, pages 798 - 805, 2006. 1, 5
- [2] S. Avidan. Ensemble tracking. In CVPR, pages 494 - 501, 2005. 1
- [3] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In CVPR, 2009. 1, 5
- [4] C. Bao, Y. Wu, H. Ling, and H. Ji. Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, 2012. 1, 2, 3, 6
- [5] M. J. Black and A. D. Jepson. Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. IJCV, pages 63 - 84, 1998. 1
- [6] R. T. Collins and Y. Liu. On-line selection of discriminative tracking features. In ICCV, pages 346 - 352, 2003. 1
- [7] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based Object Tracking. TPAMI, 25(5):564 - 575, Apr. 2003. 1
- [8] M. Everingham, L. Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object class (voc) challenge. IJCV,

- 88(2):303 - 338, 2010. 6
- [9] H. Grabner, M. Grabner, and H. Bischof. Real-Time Tracking via On-line Boosting. In BMVC, 2006. 1, 5
- [10] H. Grabner, C. Leistner, and H. Bischof. Semi-supervised online
- [11] S. Hare, A. Saffari, and P. Torr. Struck: Structured output tracking with kernels. In ICCV, 2011. 5
- [12] J. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In ECCV, 2012. 5
- [13] Z. Hong, X. Mei, D. Prokhorov, and D. Tao. Tracking via robust multi-task multi-view joint sparse representation. In ICCV, 2013. 1, 2, 3, 5, 6
- [14] A. Jepson, D. Fleet, and T. El-Maraghi. Robust on-line appearance models for visual tracking. TPAMI, 25(10):1296 - 1311, 2003. 1
- [15] X. Jia, H. Lu, and M.-H. Yang. Visual tracking via adaptive structural local sparse appearance model. In CVPR, 2012. 1, 2, 3, 4, 5, 6
- [16] Z. Kalal, J. Matas, and K. Mikolajczyk. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In CVPR, 2010. 1, 5
- [17] M. Kristan, L. Cehovin, and et al. The visual object tracking vot 2013 challenge results. In ICCV2013 Workshops, Workshop on Visual Object Tracking Challenge, 2013. 1
- [18] J. Kwon and K. M. Lee. Visual tracking decomposition. In CVPR, 2010. 1
- [19] H. Li, C. Shen, and Q. Shi. Real-time visual tracking with compressed sensing. In CVPR, 2011. 1, 2, 3, 5
- [20] B. Liu, J. Huang, L. Yang, and C. Kulikowski. Robust visual tracking with local sparse appearance model and k-selection. In CVPR, 2011. 1, 2, 3, 4, 5
- [21] B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, and C. Kulikowski. Robust and fast collaborative tracking with two stage sparse optimization. In ECCV, 2010. 1, 2, 3, 5
- [22] X. Mei and H. Ling. Robust Visual Tracking and Vehicle Classification via Sparse Representation. TPAMI, 33(11):2259 - 2272, 2011. 1, 2, 3, 4, 5, 6
- boosting for robust tracking. In ECCV, 2008. 1
- [23] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai. Minimum error bounded efficient l1 tracker with occlusion detection. In

- CVPR, 2011. 1, 2, 3, 5, 6
- [24] Y. Pang and H. Ling. Finding the best from the second bests
- inhibiting subjective bias in evaluation of visual tracking algorithms. In ICCV, 2013. 1
- [25] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental Learning for Robust Visual Tracking. IJCV, 77(1):125 - 141, 2008. 1, 5
- [26] S. Salti, A. Cavallaro, and L. D. Stefano. Adaptive appearance modeling for video tracking: Survey and evaluation. TIP, 21(10):4334 - 4348, 2012. 1
- [27] L. Sevilla-Lara and E. Learned-Miller. Distribution fields for tracking. In CVPR, pages 1910 - 1917, 2012. 5
- [28] L. Chevin, M. Kristan, and A. Leonardis. Robust visual tracking using an adaptive coupled-layer visual model. TPAMI, 35(4):941 - 953, 2013. 5
- [29] D. Wang, H. Lu, and M. Yang. Online Object Tracking with Sparse Prototypes. IEEE Transaction on Image Processing, 22(1):314 - 325, 2013. 1
- [30] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In CVPR, 2013. 1, 6
- [31] M. Yang, Y. Wu, and G. Hua. Context-aware visual tracking. TPAMI, 31(7):1195 - 1209, 2009. 1
- [32] R. Yao, Q. Shi, C. Shen, Y. Zhang, and A. van den Hengel. Part-based visual tracking with online latent structural learning. In CVPR, 2013. 5
- [33] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. ACM Comput. Surv., 38(4):13, Dec. 2006. 1
- [34] K. Zhang, L. Zhang, and M.-H. Yang. Real-time compressive tracking. In ECCV, 2012. 1, 2, 3, 5, 6
- [35] T. Zhang, B. Ghanem, and N. Ahuja. Robust multi-object tracking via cross-domain contextual information for sports video analysis. In International Conference on Acoustics, Speech and Signal Processing, 2012. 1
- [36] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Low-rank sparse learning for robust visual tracking. In ECCV, 2012. 1, 2, 3, 5
- [37] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via multi-task sparse learning. In CVPR, 2012. 1, 2, 3, 5, 6
- [38] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via structured multi-

task sparse learning.

International

Journal of Computer Vision,

101(2):367 - 383, 2013. 3

[39] T. Zhang, C. Jia, C. Xu,

Y. Ma, and N. Ahuja. Partial

occlusion

handling for visual tracking

via robust part matching. In

CVPR, 2014. 1

[40] T. Zhang, S. Liu, N.

Ahuja, M.-H. Yang, and B.

Ghanem. Robust

Visual Tracking via Consistent

Low-Rank Sparse Learning.

International Journal of

Computer Vision, 111(2):171 -

190, 2015. 3

[41] W. Zhong, H. Lu, and M.

Yang. Robust Object Tracking

via

Sparse Collaborative Appearance

Model. IEEE Transaction

on Image Processing,

23(5):2356 - 2368, 2014. 1

[42] B. Zhuang, H. Lu, Z. Xiao,

and D. Wang. Visual Tracking via

Discriminative Sparse

Similarity Map. IEEE

Transaction on

Image Processing, 23(4):1872 -

1881, 2014. 1