

基于局部分块学习的在线视觉跟踪

余旺盛¹, 田孝华¹, 侯志强¹, 查宇飞²

(1. 空军工程大学信息与导航学院, 陕西西安 710077; 2. 空军工程大学航天航空工程学院, 陕西西安 710038)

摘 要: 视觉跟踪中, 如何构建一种能够适应目标表观特征变化的目标模型是增强算法跟踪精度和稳定性的关键之一. 本文提出利用跟踪区域内像素的初始分类标记来构建目标的局部分块模型, 并在贝叶斯理论框架下提出了基于局部分块学习的在线视觉跟踪算法. 首先, 利用标定的初始跟踪区域构建目标的局部分块模型; 然后, 在当前跟踪区域中通过局部分块学习和贝叶斯估计确定当前帧的跟踪结果; 最后, 利用特征聚类对局部分块模型进行更新. 实验结果表明: 所提算法对目标表观变化的适应性明显增强, 跟踪精度和稳定性较近年来的同类算法均有一定提高.

关键词: 视觉跟踪; 局部分块模型; 贝叶斯估计; 模型更新

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2015)01-0074-05

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2015.01.012

Online Visual Tracking Based on Local Patch Learning

YU Wang-sheng¹, TIAN Xiao-hua¹, HOU Zhi-qiang¹, ZHA Yu-fei²

(1. Information and Navigation College, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

2. Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: In visual tracking, how to construct an object model to cope with the appearance change is one of the key problems to improve tracking precision and stability. To resolve this problem, this paper proposes to construct a local patch model using the initial labels of the pixels in tracking area, and proposes an online visual tracking algorithm based on local patch learning under the framework of Bayesian theory. The detailed operation is as follows. Firstly, it constructs the local patch model according to the initialized tracking area. Then, it utilizes the object model to learn the local patches in current tracking area and estimates the current state via Bayes estimation. Finally, it updates the local patch model by feature clustering. The experiment results indicate that the proposed algorithm obtains a distinct improvement in coping with appearance change, and exceeds the recent local patch-based trackers in both tracking precision and stability.

Key words: visual tracking; local patch model; Bayes estimation; model update

1 引言

随着计算机视觉和多媒体技术的不断发展, 视觉跟踪技术逐渐应用于许多民用和军用领域. 由于在增强人工智能和提高生产效率等方面的潜在优势, 视觉跟踪理论研究已经成为学者们关注的热点问题. 在实际应用场合中, 跟踪系统对算法跟踪的精度和稳定性提出了较高的要求. 然而, 当被跟踪目标的表观特征发生较大变化时, 跟踪算法的某些性能往往达不到跟踪系统的要求. 此时, 如何构建能够适应复杂表观变化的目标模型是解决问题的关键. 早期的跟踪算法^[1]直接利用模板进行跟踪, 对目标的尺度、旋转和形变比较敏感. 文献[2]研究了模板跟踪过程中的模板更新问题. 在模板跟踪的基础

上, 文献[3]利用区域颜色直方图构建目标表观模型, 增强了算法对目标尺度、旋转以及形变的鲁棒性. 文献[4]也采用颜色直方图作为目标模型, 并结合粒子滤波完成目标跟踪. 文献[5]研究了基于特征子空间的模型构建方法, 增强了算法对目标表观特征变化的适应性. 近年来, 稀疏表示^[6]成功应用于视觉跟踪算法中, 其将视频图像投影到稀疏的特征空间中, 用尽可能少的系数来描述目标的主要信息, 增强了算法对目标遮挡、形变等的适应性. 局部分块模型^[7]通过目标的局部分块来增强算法对遮挡变化的适应性, 但并未改变跟踪算法对目标其他表观变化的敏感性. 文献[8]提出一种基于局部动态分块的表观模型, 增强了算法对目标尺度、旋转以及形变的鲁棒性.

局部分块模型对目标变化具有较强的适应性,然而,现有的局部分块跟踪算法^[7,8]对目标的光照以及复杂背景变化的适应性较差.对此,本文提出一种基于局部分块学习的在线视觉跟踪算法,通过在线学习的方法实时更新局部分块模型,进而增强算法对复杂表现变化的适应性.实验结果表明,本文所提算法对目标表现变化的适应性明显增强.

2 局部分块模型

局部分块模型是指将目标所在的图像区域从空间上分解成若干个子图像区域,然后分别进行特征描述并以此构建的目标模型.其对目标的部分遮挡具有更好的适应性.跟踪中,初始模板通常包含有部分背景信息,其在后续跟踪中会影响算法的跟踪性能并累积误差,最终导致跟踪失败.因此本文的出发点是最大限度地抑制背景信息,通过对目标模板及其周围区域进行局部分块,继而通过聚类分析来区分前景与背景.

定义以目标模板中心为中心,四倍于目标模板大小的图像区域为跟踪区域.建模步骤如下:首先将当前跟踪区域划分为 n^2 个子图像块,构建局部分块;然后对各分块进行特征描述,构建样本空间;接下来依据初始标记结果对样本进行聚类分析;最后依据聚类结果计算各分块为目标前景的概率.

建模过程中的关键步骤即样本的聚类分析.记 $hist_k$ 为第 k 个子图像块的特征向量,本文采用 Mean Shift 特征聚类算法^[9]对 n^2 个特征向量进行聚类,聚类过程表述为:

$$\{clust_m\}_{m=1}^C = MSCluster(\{hist_k\}_{k=1}^{n^2}, B) \quad (1)$$

式中, $MSCluster(\cdot)$ 为聚类算法, B 为聚类的带宽, C 为所得特征类的数目.

位于目标模板内部的所有像素的初始标记值为 1, 位于背景区域的所有像素的初始标记值为 0. 提取样本后,每个样本的标记值为样本内部所有像素初始标记的均值.特征聚类后,特征距离小于设定带宽的样本被聚为相同的类,类的标记值为类内所有像素初始标记的均值.标记确定后,反向修改类中样本标记,使类内样本与所属类具有相同的标记.改进目标模型记为:

$$Model = \{clust_m, label_m\}_{m=1}^C \quad (2)$$

式中, $clust$ 为包含样本及特征的聚类结果, $label$ 为类的标记值.

3 本文跟踪算法

基于提出的改进局部分块目标模型,在贝叶斯估计框架下提出了一种鲁棒的在线视觉跟踪算法.

3.1 跟踪原理

当输入新的视频图像时,以上一帧跟踪结果(或初

始标定结果)为中心,标定当前跟踪区域,对当前跟踪区域进行局部分块抽取样本,分别计算各样本到模型中各类的特征距离,依据最近邻分类原则,利用模型内聚类的标记值标记当前样本.根据当前样本标记构建跟踪区域的似然空间,最后利用贝叶斯估计得到目标状态的最大后验概率估计.

3.2 算法步骤

本文跟踪算法的主要步骤如下:

第 1 步:模型初始化.

在初始帧中标定目标并构建局部分块模型.

第 2 步:样本抽取.

输入新视频帧时,依据上一帧的跟踪结果确定当前跟踪区域 R_t 并提取样本.

第 3 步:样本学习.

计算每个样本到目标模型中所有类的距离,以距离最小的类标记对该样本进行标记.第 k 个样本到模型中第 m 个类的距离计算公式如下:

$$d_{km} = \arg \min_i [d(hist_k, hist_{mi})], \forall hist_{mi} \in clust_m \quad (3)$$

式中, $d(hist_k, hist_{mi})$ 为向量的距离度量公式:

$$d(hist_k, hist_{mi}) = 1 - \left[\sum_{u=1}^{bin} \sqrt{hist_k(u) \cdot hist_{mi}(u)} \right]^2 \quad (4)$$

式中, bin 为向量的维度,也即直方图的柱数.

样本标记完毕后,样本内像素与所属样本具有相同的标记值,记为 $p(i, j), (i, j) \in R_t$.

第 4 步:贝叶斯估计.

依据上一时刻的目标状态 X_{t-1} 和状态转移模型 $p(X_t | X_{t-1}) \sim N(X_t; X_{t-1}, \Phi)$ 进行粒子采样,采样结果记为 $\{X_t^{(k)}\}_{k=1}^N$. 计算粒子 $\{X_t^{(k)}\}_{k=1}^N$ 后验概率的公式如下:

$$p(X_t^{(k)} | Y_{1:t}) = \alpha_0 \cdot p(X_t^{(k)} | X_{t-1}) \cdot \sum_{(i,j) \in X_t^{(k)}} p(i, j) \quad (5)$$

式中, $Y_{1:t}$ 为目标在 t 时刻的所有观测. α_0 为归一化参数,保证后验概率的取值范围在 $[0, 1]$ 之间.

目标状态的最大后验概率估计结果为

$$\hat{X}_t = \arg \max_k p(X_t^{(k)} | Y_{1:t}), \forall k = 1, 2, \dots, N \quad (6)$$

状态转移模型中, Φ 为描述高斯分布标准差的对角矩阵,表征各维度内参数变化幅度的标准差.若令 $X_t = (x_t, y_t, h_t, w_t)$, 则 $\Phi = \text{diag}[\delta_x, \delta_y, \delta_h, \delta_w]$.

在式(5)中增加“尺度适应因子”得到如下公式:

$$p(X_t^{(k)} | Y_{1:t}) = \alpha_0 \cdot p(X_t^{(k)} | X_{t-1}) \frac{S(X_t^{(k)})}{S(\hat{X}_{t-1})} \sum_{(i,j) \in X_t^{(k)}} p(i, j) \quad (7)$$

式中, $S(X_t^{(k)})$ 为粒子 $X_t^{(k)}$ 的面积.当粒子包含的前景像素越多时,粒子面积越大,则粒子对应的后验概率值

也越大;当粒子包含的背景像素越多时,粒子面积越大,则粒子对应的后验概率值反而越小。“尺度适应因子”能够增大真实反映目标状态粒子的后验概率,从而达到尺度自适应的效果。

结合式(6)和(7)即可得到目标当前状态的最大后验概率估计 \hat{X}_t 。

第5步:模型更新。

更新的原理为利用一定数量帧(W 帧)所包含的局部分块构建样本空间,通过样本空间的更新来更新目标模型。当满足样本空间的更新条件时,利用当前跟踪结果对样本空间进行更新;当满足目标模型更新条件时,对样本空间进行聚类分析更新目标模型。该方法在保留目标历史特征的同时不断学习新的目标特征,从而抑制模型漂移。

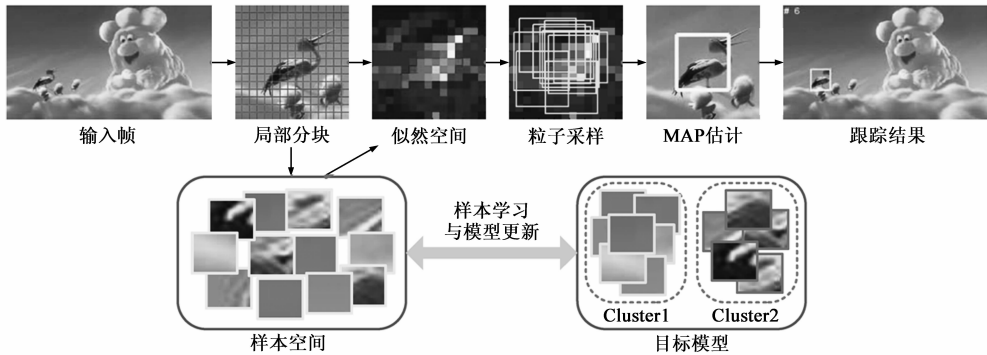


图1 本文算法跟踪过程示意图

4 仿真实验

为验证本文算法的有效性,进行了一系列的跟踪仿真实验.实验中,本文算法参数设置为:局部分块数目 $n^2 = 225$;样本特征为 HSI 空间的 $8 \times 8 \times 4$ 颜色直方图;Mean Shift 聚类的带宽参数 $B = 0.15$;贝叶斯估计时粒子数目为 $N = 300$;遮挡阈值参数 $\theta_0 = 0.45$,模型更新条件参数 $W = 5, U = 2, V = 8$.上述参数在本文的所有跟踪实验中均未作任何调整.需要根据跟踪任务进行设置的参数仅为动态模型参数 Φ ,本文实验在 3.29GHz, 3.24GB 的个人计算机上,用 Matlab2011 编程实现。

4.1 实验结果

为充分说明本文算法在处理目标遮挡、复杂背景干扰以及目标尺度变化等方面的优势,选择了3组具有代表性的测试视频与近年来利用局部分块模型进行目标跟踪的算法进行对比测试.参考算法为 Frag^[7]和 OHAB^[8].Frag 算法的代码来自文献[10],OHAB 算法的代码由文献[8]提供。

“Basketball”序列的跟踪难点为目标运动速度较快,且存在明显遮挡和背景干扰因素.由图 2 可知,Frag 算

法虽然完成了对目标的跟踪,但跟踪的误差较大;OHAB 算法开始时跟踪结果较为理想,但在第 620 帧中出现严重的偏差.本文算法克服了目标遮挡对跟踪的影响,跟踪结果的位置误差相对较小。

“Bolt”序列的跟踪难点为目标快速运动,分辨率较低且存在复杂的背景干扰.图 2 显示,Frag 算法跟踪效果较差,原因主要是目标的非刚性形变以及复杂背景干扰.OHAB 算法对目标的非刚性形变具有一定的鲁棒性.本文算法通过样本聚类分析增强了前景与背景的可分性,在抑制背景因素的同时增强了对目标非刚性形变的适应性。

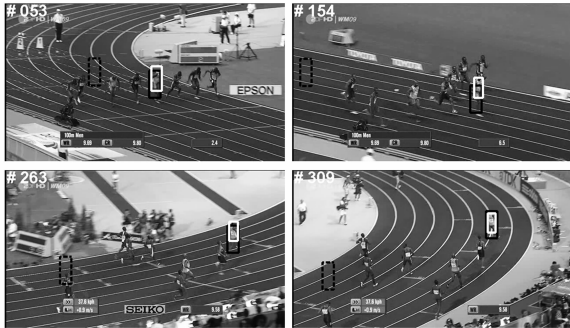
为验证各算法对噪声环境下运动目标的跟踪效果,在测试视频“MountainBike”中添加了强度为 5% 的高斯噪声.被跟踪目标为山地摩托车,由于目标的颜色特征不明显,且存在目标旋转和背景干扰因素,因此跟踪难度较大.跟踪过程中先对视频帧进行噪声抑制预处理,然后利用各算法完成对视频的跟踪.由图 2 所列跟踪结果可见,在噪声预处理后的视频中,本文算法能够有效地对目标进行定位,受噪声的影响较小.参考算法中,Frag 算法在所列的跟踪结果中均出现不同程度的偏差,这是由于目标区域内包含的背景信息所造成的;O-

跟踪的整体过程如图 1 所示。

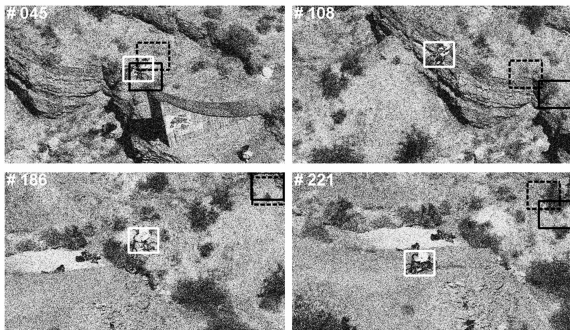
HAB算法仅在第45帧中完成了跟踪,在后续帧中算法受背景因素的影响出现了严重的跟踪漂移.



(a) “Basketball”序列,参数 $\phi=[10,10,0,0]$



(b) “Bolt”序列,参数 $\phi=[10,10,2,2]$



(c) “MountainBike”序列,参数 $\phi=[10,10,0,0]$

图2 更多跟踪实验结果比较,“MountainBike”中添加了强度为5%的高斯噪声.白色实线为本文算法,黑色实线和虚线分别为Frag和OHAB

4.2 性能分析

衡量算法跟踪性能的客观评价指标主要有中心位置误差(Center Location Error, CLE)和跟踪成功率(Tracking Success Rate, TSR)等.

(1) 算法跟踪精度分析

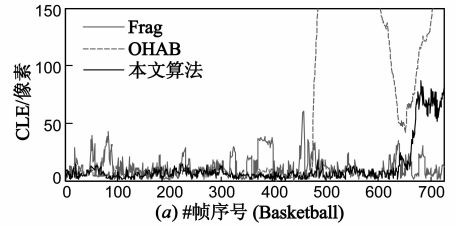
中心位置误差 CLE 用于衡量算法的跟踪精度,第 k 帧 CLE 的计算公式为:

$$\varepsilon(k) = \sqrt{[x(k) - x'(k)]^2 + [y(k) - y'(k)]^2} \quad (9)$$

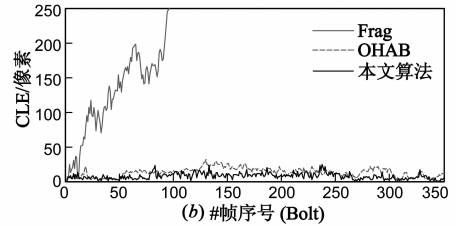
式中, x 和 y 为跟踪结果中心位置的横纵坐标, x' 和 y' 为 ground truth 标定的横纵坐标. 所得 CLE 的单位为像

素,数值越小,表明算法跟踪的精度越高.

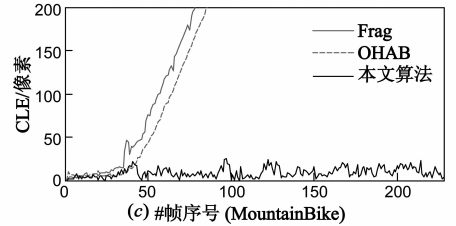
图3所示为本文算法与参考算法对3段测试视频跟踪所得 CLE 的详细比较,由图可以看出,参考算法中,Frag算法仅在“Basketball”序列中获得了较小的中心位置误差,OHAB算法在“Bolt”和“David3”序列中跟踪较为理想,在其余两段视频的后半部分均出现了较大的误差.相较而言,本文算法在所有测试视频中所得的 CLE 曲线比参考算法均要低.



(a) #帧序号 (Basketball)



(b) #帧序号 (Bolt)



(c) #帧序号 (MountainBike)

图3 跟踪所得的中心位置误差曲线比较

表1给出了CLE的统计数据.分析表中数据可知,4段视频中,本文算法均得到了最小的平均中心位置误差.Frag算法在“Basketball”序列中取得了次优结果.OHAB算法则在剩余的2组视频中取得了次优结果.整体而言,OHAB算法跟踪所得的CLE较Frag算法有较大幅度的降低,而本文算法的CLE最小,表明本文算法的跟踪精度更高.

表1 平均中心位置误差比较(单位:像素)

	Frag ^[7]	OHAB ^[8]	Proposed
Basketball	11.8	69.2	6.9
Bolt	333.8	12.9	8.5
MountainBike	205.4	203.8	9.1

(2) 跟踪结果的稳定性分析

TSR定义为跟踪成功的帧数占视频总帧数的百分比,而跟踪成功则定义为跟踪结果与ground truth的覆盖率大于0.5.TSR的数值越大,表明跟踪结果的稳定性越好.表2给出了本文算法与参考算法的TSR比较.Frag算法仅在“Basketball”序列中超过90%,OHAB算法

在“Bolt”序列中超过 99%。本文算法除了对“Bolt”序列为 99.4% 之外,对其他序列跟踪均为 100%。由此可见,本文算法跟踪所得的结果更加稳定。

表 2 跟踪成功率比较

	Frag ^[7]	OHAB ^[8]	Proposed
Basketball	96.1%	73.2%	100%
Bolt	2.9%	99.1%	99.4%
MountainBike	20.1%	22.3%	100%

4.3 实验总结

实验发现,本文算法的跟踪结果在一定程度上取决于初始跟踪区域的确定。当上一帧跟踪结果的位置误差或者目标的运动速度超过目标的尺寸大小时,跟踪算法的性能会受到影响。因此,本文算法需要建立在上一帧具有一定的跟踪精度且目标运动在空间上连续的基础上。当目标发生遮挡或出现与目标特征相似的背景区域时,跟踪的结果会出现一定的偏差,但当遮挡消失或背景干扰因素消失时,只要目标在图像特征上与背景具有一定的可分性,算法依然能够继续实现对目标的跟踪。

本文通过样本空间的更新和聚类分析来更新目标模型,既保留了目标的历史特征,又不断学习新的目标特征,因此能够较好地适应光照的变化。由于局部分块学习能够有效地提取目标的前景区域,因此模型不受目标非刚性形变影响,这也可以从实验结果中得到印证。

5 结论

本文提出了一种基于局部分块学习的在线视觉跟踪算法。在所提出的基于局部分块学习的目标模型基础上,算法通过在背景中增强目标特征,进而在贝叶斯框架内实现最大后验概率跟踪。算法采用一种有效的在线模型更新方法对目标模型进行更新,增强了算法对目标复杂表观特征变化的适应性。实验结果表明,本文算法能够有效抑制复杂背景变化的干扰和处理目标的遮挡问题,得到更小的跟踪误差和更高的跟踪成功率。

下一步的研究将利用目标的运动信息对目标的运动进行预测,研究基于目标运动信息与图像信息融合的目标跟踪算法,增强算法对目标快速运动情况下跟踪的鲁棒性。

参考文献

[1] G D Hager, et al. Efficient region tracking with parametric models of geometry and illumination[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(10): 1025 - 1039.

- [2] I Matthews, et al. The template update problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 810 - 815.
- [3] D Comaniciu, et al. Kernel-based object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564 - 577.
- [4] P P'erez, et al. Color-based probabilistic tracking [A]. Proceedings of the 7th European Conference on Computer Vision [C]. Berlin, Germany: Springer-Verlag, 2002. 661 - 675.
- [5] D Ross, et al. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1): 125 - 141.
- [6] D Wang, et al. Online object tracking with sparse prototypes [J]. IEEE Transactions on Image Processing, 2013, 22(1): 314 - 325.
- [7] A Adam, et al. Robust fragment-based tracking using integral histogram [A]. Proceedings of the 21st IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE Press, 2008. 798 - 805.
- [8] S M S Nejhun, et al. Online visual tracking with histograms and articulating blocks [J]. Computer Vision and Image Understanding, 2010, 114(8): 901 - 914.
- [9] D Comaniciu, et al. Mean shift: A robust approach toward feature space analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603 - 619.
- [10] Y Wu, et al. Online object tracking: a benchmark [A]. Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland, USA: IEEE Press, 2013. 2411 - 2418.

作者简介



余旺盛 男, 1985 年 1 月出生, 湖南平江人, 分别于 2008 年和 2010 年获空军工程大学电讯工程学院信号与信息处理专业学士学位和硕士学位, 现于空军工程大学信息与导航学院攻读博士学位, 主要研究方向为数字图像处理与模式识别。

E-mail: xing_fu_yu@sina.com



田孝华 男, 1965 年 10 月出生, 湖南石门人, 2003 年获西安电子科技大学信号与信息处理专业博士学位, 现为空军工程大学信息与导航学院教授、博士研究生导师, 主要研究领域为无线电定位、阵列信号处理、非平稳信号处理、无线电导航等。

E-mail: xht65@sina.com