

DOI: 10.7652/xjtuxb201412011

利用全局约束霍夫模型的目标跟踪算法

贺文骅, 刘志镜, 屈鉴铭

(西安电子科技大学计算机学院, 710071, 西安)

摘要: 针对目标跟踪中局部特征模型容易受到形变和误匹配等影响造成漂移的问题,提出了一种利用全局特征对霍夫局部模型进行约束的目标跟踪算法。该算法通过检测和提取目标的局部特征,构建了一个基于局部特征投票的概率模型。该模型以局部特征为基本元素,利用广义霍夫变换对局部特征的稳定性进行加权,通过对局部特征的增减维护实现模型的在线更新;将稳定的局部特征作为前景,计算目标全局颜色特征的概率分布,实现简单的柔性目标分割,用于调整局部特征模型并对跟踪进行优化,从而减少形变等问题产生的跟踪误差;局部特征和全局特征相互约束补充,共同完成在线自适应更新。实验结果表明,该算法有效提高了局部特征模型的准确率和稳定性,对部分遮挡和复杂形变的情况表现良好。

关键词: 目标跟踪;局部特征;霍夫变换;目标分割

中图分类号: TP392 **文献标志码:** A **文章编号:** 0253-987X(2014)12-0069-05

A Visual Tracking Method Based on Global Constrained Hough Model

HE Wenhua, LIU Zhijing, QU Jianming

(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

Abstract: A visual tracking method is proposed to solve the problem that the local feature model is prone to be influenced by the deformation and feature mismatch and leads to drifting. The method constrains the Hough local model by using global features. The Hough model is constructed with a set of local features of an object and estimates the stability of each local feature based on the generalized Hough Transform. Then, the model is updated flexibly through modifying the local feature set to adapt to the object appearance variations. Color cues of the local features are used to calculate the global color probability of being foreground or background. Then the global probability is used to adjust the tracking results and constrains the local features of the model in return. Experimental results on several public video sequences show the robustness of the proposed method in tracking objects with deformation and partial occlusion.

Keywords: visual tracking; local feature; Hough transform; segmentation

视频序列中的目标跟踪是计算机视觉领域一个长期的研究热点,被广泛地应用于人机交互、安防监控、模式识别等领域,同时也是行为分析和场景理解等进一步研究的关键基础。跟踪过程中会遇到各种各样的问题,例如目标自身发生的尺度变化、旋转、

遮挡以及摄像头的运动等。建立可靠的特征模型对目标进行描述,同时采用合理有效的在线学习策略对目标模型进行更新是保证跟踪稳定性和准确性的关键。经典的在线目标跟踪算法^[1-3]将跟踪结果作为训练样本,对目标模型进行在线更新,使其能够较

收稿日期: 2014-06-06。 作者简介: 贺文骅(1986—),男,博士生;刘志镜(通信作者),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(61173091)。

网络出版时间: 2014-11-27 网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20141127.1818.003.html>

好地适应目标的外观变化。这些方法以目标整体作为模型,当目标发生剧烈形变或者被遮挡时,整体模型难以处理,更新过程会引入大量的背景噪声,最终导致跟踪失败。

一些学者提出了使用较低层次特征来建立目标模型的方法。文献[4-5]采用像素信息作为特征,通过霍夫投票估计目标的位置,最后通过前后景分割获取目标的最终状态。这两种方法主要采用了像素的颜色特征,对于非刚性目标的跟踪和部分遮挡有较好的表现,但在背景复杂或像素特征不明显时不够稳定。仿射不变的局部特征^[6-7]的相继提出被广泛应用于图像配准以及目标的检测和跟踪^[8-11]。文献[8]提出利用局部特征点霍夫投票实现目标跟踪来处理遮挡。文献[9]利用运动显著性对特征点进行约束,提高了跟踪的稳定性。文献[10-11]采用了尺度不变特征转换(scale-invariant feature transform, SIFT)特征来实现跟踪。文献[10]通过对特征点的大量仿射变换采样和增量式主成分分析建立局部特征模型集合,提高了特征点的匹配准确率,但是这些方法没有考虑目标结构形变的问题,而且运算量较大。文献[12]提出利用超像素分割对图像进行处理,以超像素为单位通过聚类对颜色进行统计,生成判别式模型。该方法由于仅采用了颜色特征,因此对于形变的鲁棒性非常强,但也容易受到近似颜色和遮挡的干扰,而且反复的聚类运算也大大增加了运算量。

本文提出一种综合考虑前后景分割和局部特征投票的在线跟踪算法。算法检测和提取目标的局部特征点,并结合局部特征与目标的相对关系构建基于局部特征的目标模型。利用广义霍夫变换作为投票机制,通过局部特征的投票估计出目标的中心位置,同时获得每个局部特征的稳定性权值。利用全局颜色特征计算前景和背景的概率分布,优化跟踪结果并对局部特征进行约束。局部特征投票和全局概率分布在跟踪上相互补充,在更新过程中相互约束,对于目标的严重形变和部分遮挡有很强的适应性。

1 算法介绍

设 t 时刻目标的状态由其位置坐标 x 和区域 R 表示,记作 $X_t = (x_t, R_t)$ 。设 t 时刻之前观测状态为 $Z_{1:t}$,由贝叶斯概率理论得到目标状态的后验概率为

$$p(X_t | Z_{1:t}) = p(Z_t | X_t) \cdot$$

$$\int p(X_t | X_{t-1}) p(X_{t-1} | Z_{1:t-1}) dX_{t-1} \quad (1)$$

式中: $p(Z_t | X_t)$ 是目标的观测概率模型; $p(X_t | X_{t-1})$ 是运动概率模型。本文的观测模型由两部分组成,基于局部特征的霍夫投票模型和基于全局特征的颜色概率模型。对于新一帧的图像,检测目标所在区域的局部特征,并与局部特征模型进行匹配。利用匹配到的局部特征对目标的中心进行投票,所有投票结果的线性加权作为目标当前位置的概率,投票空间的反向投影作为局部特征的稳定性权值。同时,利用全局颜色特征计算每个像素属于前景目标的概率。对目标区域内所有像素属于前景的概率求和,即为目标的全局颜色特征概率。融合两种模型得出的概率分布作为最终的观测概率模型,结合运动概率模型,计算出目标状态的后验概率,再通过最大化后验概率估计(maximum a posteriori, MAP)求出目标的当前位置。更新过程中,以跟踪结果为训练样本对两种模型进行更新。通过局部特征区域的颜色统计对颜色概率模型进行更新,而颜色概率模型又反过来对局部特征集合的更新和维护进行约束。算法的基本流程如图1所示。

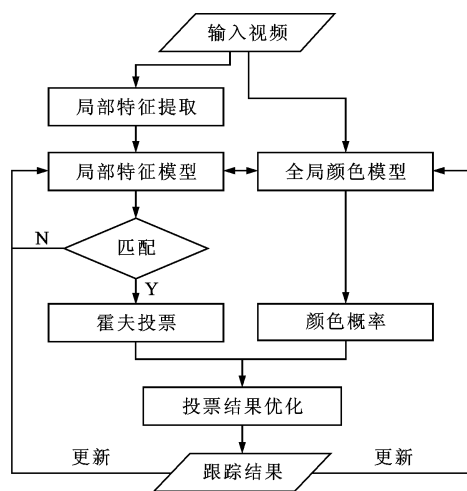


图1 算法流程示意图

2 目标模型

2.1 霍夫投票

霍夫投票是利用局部特征对目标的中心进行投票,通过寻找投票空间的最大值,估计目标中心的位置。通过投票空间的反向投影计算局部特征的稳定性作为权值。采用类似文献[9]的方法构建局部特征模版的集合 S 作为霍夫投票的模型

$$S = \{s_i | s_i = (d_i, w_i, r_i, c_i)\}_{i=1}^M$$

式中: d_i 是特征描述子; w_i 是稳定性权值; $r_i = (dx_i, dy_i)$ 是局部特征相对于目标中心的位移; c_i 是

局部特征的类型,表示其属于前景的概率。对于新一帧的图像,检测目标及其附近区域的局部特征。本文采用 Harris 角点和 SIFT 特征来检测特征点,采用 SIFT 特征描述检测到的特征点。设当前帧检测到的局部特征集合为 $F = \{f_j\}$,将 F 与模版集合 S 中的局部特征进行匹配,利用匹配成功的局部特征对目标中心进行投票。局部特征的匹配通过计算特征相似度和特征点跟踪两种方式来实现匹配。计算两个局部特征 SIFT 特征描述子的距离,距离最小并且小于设定阈值则认为两者匹配。对于未能匹配的局部特征,使用 FB 算法^[13]对相邻帧的特征点进行跟踪,得到的结果也作为匹配。

设 s_i 和 f_j 匹配成功,记作 $m_{i,j}$, $l(\cdot)$ 表示局部特征在当前帧的坐标,则该匹配对跟踪目标中心的投票为 $l(m_{i,j}) = l(f_j) + r_i$ 。为了平滑输出,我们用高斯函数近似表示采样 X 在 $m_{i,j}$ 投票下的概率

$$p(X|m_{i,j}) = \exp(-\|x - l(m_{i,j})\|_2 / 2\sigma^2) \quad (2)$$

式中: σ 是方差,用于控制生成的置信度图像的平滑程度; x 是采样的中心坐标; $\|x - l(m_{i,j})\|_2$ 是采样中心与投票位置的 2 范数距离。定义 t 时刻检测到的局部特征 F 与霍夫投票模型 S 的所有成功匹配为集合 M ,则最终的投票结果为

$$p_h(X|M) = \epsilon \sum_{c_j > 0} V_{i,j} + (1 - \epsilon) \sum_{c_j < 0} V_{i,j}, \quad \forall V_{i,j} = w_i p(X|m_{i,j}), m_{i,j} \in M \quad (3)$$

式中: ϵ 是特征类型权重,用来调整目标特征和背景特征投票的重要性比例,突出属于目标特征点的投票,减少背景的干扰。

2.2 全局特征约束

目标发生持续形变时,特征点的检测和匹配都会受到影响,并且会改变特征点与目标的相对关系,导致投票出现偏差。对于细节特征不明显的目标,局部特征也难以检测和匹配。采用全局颜色特征计算前后景的概率分布,对投票结果进行优化,可以避免因为形变以及有效局部特征过少而产生漂移,降低背景局部特征对投票的干扰。

在 HSV 颜色空间下,分别统计前景的颜色直方图 h_f 和背景的颜色直方图 h_b 。设 $\text{bin}(x)$ 为像素 x 在颜色直方图内的区间索引,则该像素对应前景和背景直方图的值分别为 $h_f(\text{bin}(x))$ 、 $h_b(\text{bin}(x))$ 。计算像素 x 属于前景目标的概率

$$p(c_t^x = 1 | Z_{1:t}) = \frac{h_f(\text{bin}(x))}{h_f(\text{bin}(x)) + h_b(\text{bin}(x))} - \delta \quad (4)$$

式中: δ 是惩罚因子常数,作为区分前后景概率的阈值。当概率为正时,认为像素 x 属于前景,反之则属于背景。对于任意区域 R ,其属于前景目标的概率为其包含所有像素前景概率的均值

$$p_c(R) = \frac{1}{N_R} \sum_{x \in R} p(c_t^x = 1 | Z_{1:t}) \quad (5)$$

前景和背景颜色直方图在跟踪过程结束后进行在线更新。以所有属于前景的局部特征的矩形块所覆盖的区域为前景,统计当前帧的前景颜色直方图 h^f 。以目标矩形周围的区域为背景,统计当前帧的背景颜色直方图 h^b 。通过自回归的方式对全局前景和背景颜色直方图进行更新

$$\left. \begin{aligned} h_t^f &= \rho_h h_{t-1}^f + (1 - \rho_h) h^f \\ h_t^b &= \rho_h h_{t-1}^b + (1 - \rho_h) h^b \end{aligned} \right\} \quad (6)$$

式中: ρ_h 是更新速率常数。颜色概率通过对局部特征区域的统计进行计算,同时对局部特征进行约束。利用式(5)计算每个局部特征属于前景目标的概率,从而调整局部特征的投票贡献。

3 目标跟踪

在新一帧图像中,以上一帧目标位置为中心,设定搜寻窗口进行特征点检测,利用霍夫投票模型求出目标中心的概率分布,并将投票空间最大概率所在位置的坐标记作 $x_{\max}$,同时利用全局颜色概率模型求出窗口内所有像素属于目标的概率。对于任一采样 $X = \{x, R\}$,计算其似然概率为

$$p(Z|X) = \alpha p_c(R) + (1 - \alpha) \exp\left(-\frac{(x - x_{\max})^2}{2\sigma^2}\right) \quad (7)$$

式中: α 是线性加权系数,用于平衡两种概率模型的关系。当投票空间上 $x_{\max}$ 的概率小于设定阈值时,认为投票结果无效,令 $\alpha = 1$ 。由式(1)根据最大化后验概率估计目标当前的位置

$$\hat{X}_t = \arg \max_{x_t^{(l)}} p(X_t^{(l)} | Z_{1:t}), l = 1, \dots, N \quad (8)$$

图 2 是算法关键步骤的示意图。图 2a 是投票空间得出最大概率的位置;图 2b 是全局颜色特征概率分布;图 2c 是在投票概率的基础上根据颜色概率分布进行优化,虚线矩形是投票得出的位置,实线矩形是优化后的位置;图 2d 是根据两种模型的概率分布求出的目标当前位置。

4 模型更新

对于匹配成功的局部特征模版,利用其对应的局部特征信息进行更新,更新方式为

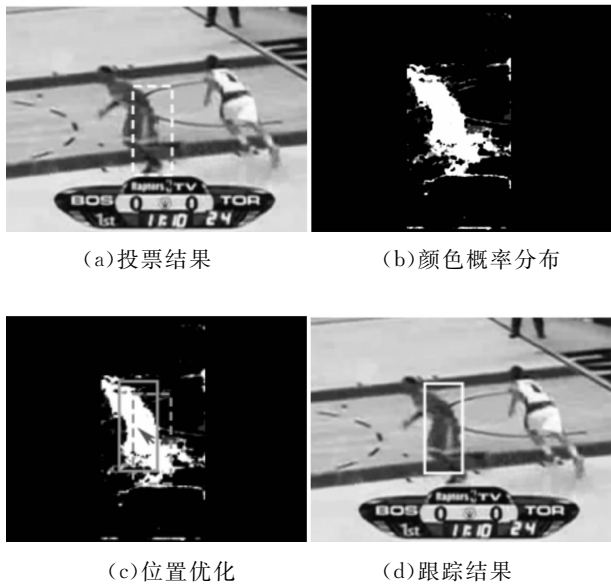


图2 算法关键步骤示意图

$$d_t^i = \rho d_{t-1}^i + (1 - \rho) d^j \quad (9)$$

$$r_t^i = \rho r_{t-1}^i + (1 - \rho) (x_i - x^*) \quad (10)$$

$$c_t^i = \rho c_{t-1}^i + (1 - \rho) p_c(f_i) \quad (11)$$

式中: ρ 是特征模型的更新速率; d^j 是当前帧匹配的局部特征的 SIFT 特征描述子; x^* 是当前帧跟踪结果目标中心的坐标; $p_c(f_i)$ 是特征 f_i 属于前景目标的概率。由于颜色概率模型对投票结果得出的目标位置进行了改变,不能直接使用投票空间的反向投影更新局部特征的权值,需要根据最终跟踪结果的目标中心位置,计算局部特征当前帧的稳定性权值 w^j ,再进行更新,更新方式同式(9)。

$$w^j = \exp(-\|l(m_{i,j}) - x^*\|_2 / 2\sigma^2) \quad (12)$$

对于未能成功匹配的局部特征,将其添加到投票模型的模版集合。为了提高运算效率,保证霍夫投票模型中局部特征模板的有效性,一段时间内连续不被匹配的局部特征模板会被移除。同时,为局部特征模板的数量设置上限,当模型中的局部特征模板的数量超过设定的上限阈值时,移除其中权值较低的模板。

5 实验结果

采用 C++ 和 OpenCV 编写程序对本文算法进行了测试实验。实验在处理器为酷睿 2.93 GHz 的个人计算机上执行,平均速度可达到 13 帧/s。实验采用了 9 个常用的公共视频对算法进行了测试,测试视频来自于文献[14],其内容包括了光照变化、尺度变化、旋转、形变和遮挡等情况。实验结果与经典的跟踪算法 MIL(multiple instances learning)^[1]、

TLD(tracking-learning-detection)^[3]、SPT(super pixel tracking)^[12]进行了对比,图 3 给出了实验结果的部分截图。采用正确跟踪的帧数和平均中心误差对跟踪结果进行评价,评价标准的具体计算方式同文献[12]。表 1、表 2 分别给出了两种评价标准下各算法的对比结果,粗体表示最优的两项。实验结果表明,本文算法在选取的测试视频上表现良好,能够稳定准确地完成跟踪。



图3 部分实验和对比结果截图

在视频 lemming 和 liquor 中,目标经历了大量的尺度变化、遮挡和旋转形变。MIL 在 liquor 中由于尺度变化和遮挡同时发生导致目标整体变化过大,从而丢失了目标。本文算法利用 SIFT 特征的尺度旋转不变性,结合颜色特征的约束,有效地适应了尺度、旋转和遮挡的复合变化。TLD 算法由于错误的尺度变换导致跟踪效果不理想。

视频 basketball 和 david_out_door 是两个相对复杂的场景。目标人物在跟踪过程中经历了遮挡、快速移动以及转身等严重的非刚性形变,这些问题使得局部特征与目标的相对关系发生了持续变化,也造成特征点大量替换,难以检测和匹配。本文算法利用全局颜色特征分布对投票结果进行了约束修正,成功完成了对目标的跟踪。TLD 算法由于目标

的快速变化和相互遮挡,在这两个视频中表现不佳。

表 1 跟踪结果成功帧数

视频 序列	成功帧数			
	MIL	TLD	SPT	本文算法
lemming	1 105	361	1 277	1 225
liquor	353	1 398	1 689	1 446
basketball	204	46	695	696
woman	39	36	298	301
tiger	272	45	142	286
faceocc	793	824	592	817
david	351	502	114	502
david_in_door	396	537	409	527
david_out_door	243	71	249	241

表 2 跟踪结果平均中心误差

视频 序列	平均误差			
	MIL	TLD	SPT	本文算法
lemming	14	17	7	13
liquor	165	28	9	17
basketball	104	170	6	9
woman	120	95	11	14
tiger	11	39	24	13
faceocc	19	17	28	12
david	16	5	51	6
david_in_door	19	11	21	14
david_out_door	13	53	11	15

在视频 faceocc 中,目标发生了严重的遮挡,并且维持了较长时间。以整体为模型的 TLD 和 MIL 算法在目标被遮挡后,在线更新时将遮挡物也视为目标的一部分,导致了跟踪误差的产生。TLD 算法产生了向上的轻微漂移,MIL 算法产生了向下的漂移。SPT 算法也在遮挡最严重时完全丢失了目标。本文算法通过未被遮挡部分的局部特征对目标中心投票,获得了更小的误差。

在视频 woman 和 tiger 中,目标经历了多次遮挡。不同的是,woman 中遮挡方式相对简单,但目标在 89 帧附近发生了大幅形变,这使得 MIL 和 TLD 都丢失了目标。而在 tiger 视频中,遮挡方式相对复杂,导致了颜色分布相对复杂,使得 SPT 表现不佳。本文算法采用颜色特征分布和局部特征投票相互约束,在这两个视频中都取得了很好的表现。

在视频 david 和 david_in_door 中,目标经历了光照变化、尺度变化和旋转。本文算法通过霍夫变换在更新过程中对局部特征权值的调整,有效适应

了目标的变化,取得了良好的跟踪效果。SPT 算法因为只采用了颜色特征,在背景近似颜色的干扰下丢失了目标。

6 结 论

针对视频序列目标跟踪过程中的形变和遮挡问题,提出了一种全局颜色特征概率约束下局部特征霍夫投票的目标跟踪算法。算法通过分析局部特征及其与目标的相对关系,构建了一个基于局部特征的霍夫投票模型。利用全局颜色特征计算前后景的概率分布,优化投票结果并对局部特征进行约束。由于采用了局部特征投票和全局概率分布相结合的方式,算法能够有效地处理部分遮挡的情况,对目标形变有较强鲁棒性,但是本文算法依赖于目标局部特征的丰富和稳定,且更侧重于短期连续的跟踪,如何处理目标局部特征不明显以及长期跟踪中目标丢失找回的问题,是下一步的研究方向。

参考文献:

[1] BABENKO B, YANG M H, BELONGIE S. Visual tracking with online multiple instances learning [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ, USA: IEEE, 2009: 983-990.

[2] ROSS D A, LIM J, LIN R S, et al. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1/2/3): 125-141.

[3] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking learning detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.

[4] GODEC M, ROTH P M, BISCHOF H. Hough-based tracking of non-rigid objects [J]. Computer Vision and Image Understanding, 2013, 117(10): 1245-1256.

[5] DUFFNER S, GARCIA C. Pixeltrack: a fast adaptive algorithm for tracking non-rigid objects [C]// Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos, CA, USA: IEEE Computer Society, 2013: 2480-2487.

[6] LOWE D G. Distinctive image features from scale invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[7] BAY H, TUYTELAARS T, VAN G L. Speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

(下转第 86 页)