

Application of Tensor Voting in Object Tracking

Xiaofang Shao, Dalong Li

Qingdao Branch of Naval Aeronautical Engineering Institute, Qingdao Shandong
Email: pugongying_0532@163.com

Received: Sep. 2nd, 2015; accepted: Sep. 20th, 2015; published: Sep. 24th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Object tracking is a process to locate an interested object in a series of image, so as to reconstruct the moving object's track. This paper presents a summary of related works and introduces how to apply the tensor voting method in object tracking. The algorithm's flowchart and typical experimental result are demonstrated. At last, we analyze the characteristics of the algorithm and suggest some future directions.

Keywords

Object Tracking, Track Alignment, Object Detection

张量投票在目标跟踪中的应用

邵晓芳, 李大龙

海军航空工程学院青岛校区, 山东 青岛
Email: pugongying_0532@163.com

收稿日期: 2015年9月2日; 录用日期: 2015年9月20日; 发布日期: 2015年9月24日

摘要

目标跟踪就是在视频序列的每幅图像中找到所感兴趣的运动目标的位置, 建立起运动目标在各幅图像中

的联系。在分类总结相关工作的基础上, 介绍了张量投票方法在目标跟踪中的应用, 给出了算法流程和实验结果并进行了分析和展望。

关键词

目标跟踪, 轨迹校正, 目标检测

1. 引言

目标跟踪是在一段视频序列中定位感兴趣的运动目标, 并形成目标运动的路径或轨迹。作为计算机视觉领域中视频分析的基本内容之一, 目标跟踪起到承上启下的作用, 它在目标检测的基础之上, 又是目标行为分析的基础, 是在捕获到的目标初始状态和通过特征提取得到的目标特征基础上, 进行一种时空结合的目标状态估计。目标跟踪在智能监控、视频检索、人机交互、智能交通和基于运动的目标识别等领域均有广泛的应用[1]。

尽管目标跟踪已经获得很多研究成果, 包括一些实时性和鲁棒性都很好的算法, 但是很多算法都是在一定约束条件下进行的, 因而与目标跟踪有关的特征选择、目标表示、运动估计等问题都值得进一步深入研究, 尤其是特征的选择、遮挡的处理和轨迹误差的校正。张量投票方法应用于目标跟踪正是针对目标跟踪面临的难题。

本文旨在总结目标跟踪的相关算法的基础上, 具体描述张量投票方法在目标跟踪中的应用。主要内容安排如下: 第二节是相关工作; 第三节介绍张量投票在目标跟踪中的应用, 并给出实验结果示例; 最后是总结与展望。

2. 相关工作

为了克服由噪音、遮挡和复杂环境下引起的目标或背景的变化, 已有许多目标跟踪算法被提出。不同算法对目标(外观、形状、数目)、相机数目及运动形式、目标运动形式、光照条件等都有不同要求。根据在目标检测所选择的特征不同, 可以将跟踪算法分为点跟踪、核跟踪、轮廓跟踪和形状匹配。下面进行分别描述。

点跟踪选取点特征作为目标检测的依据, 以预测目标在下一帧的位置再根据实际图像进行修正为主要解决思路, 一般情况下施加的运动约束为邻近性、最大速度、最小速度改变、共同运动、三维结构约束等。这类方法可以进一步分为确定性方法和统计性方法, 确定性方法以 MGE tracker [2]、GOA tracker [3] 为典型; 而统计性方法则以 Kalman 滤波[4]、粒子滤波[5]、联合概率数据关联(JPDAF) [6] (建立多目标的联合概率分布)、PMHT [7] (在跟踪过程中保留多种匹配假设 MHT, 通过概率估计减小 MHT 的计算量)、条件随机场模型[8]为代表。上述方法中, 只有 Kalman 滤波是针对单一目标, 而 MGE、PMHT 是其中可以处理进入摄像机镜头的新目标和消失的目标, 并可以处理一些遮挡。不过其中应用较广的还是 Kalman 滤波和粒子滤波, 在实际应用中应该根据实际需要来选择合适的滤波器: 若无须考虑系统的非线性强度和 非高斯环境, 则优先采用 EKF (Extended Kalman Filter, 扩展卡尔曼滤波); 对于一般的非线性高斯模型, 宜采用(UKF, Unscented Kalman Filter, 不敏卡尔曼滤波); 实时性要求不高, 状态维数小的非线性非高斯问题中, 可选用粒子滤波(Particle Filter) [9]。

核跟踪方法可以选择模板、颜色直方图、光流场或混合模型等特征, 典型算法有简单模板匹配、均值平移(Mean-shift) [10]、外观模型的 KLT 变换[11]、分层算法[12]; 也可以采用多视点外观模型, 典型方法有 Eigentracking [13]、支持矢量机[14]等。这类方法的主要目的是通过匹配的方式估计目标的运动,

一般对目标采用模板表示法或概率密度模型表示法, 施加的约束条件有运动模型约束(匀速或常加速等)和光流约束条件。这类方法的评价标准为目标数目、遮挡处理、是否需要训练、是否需要人工初始化等。从这些角度来看, 分层算法的性能最优, 不但能够处理多目标和完全遮挡, 而且可以处理旋转、平移和尺度等类型的运动, 不需要训练和初始化; 其余算法都是针对单目标, 能够处理平移或仿射变换型的运动以及部分遮挡。

轮廓跟踪的方法以活动轮廓模型为典型代表, 基于活动轮廓的跟踪思想是利用封闭的曲线轮廓来表达运动目标, 并且该轮廓能够自动连续的更新。这些算法的目的是直接提取目标的形状, 相对于基于区域的算法提供针对目标更有效的描述。相对基于区域的跟踪, 此类方法对于描述物体较为简单而有效率, 可以减少运算的复杂程度, 由于轮廓是一种封闭曲线, 即使存在干扰或是部分遮挡, 该算法也可以实现对目标的连续跟踪。但是, 跟踪精度只限在轮廓级水平, 并且仅是从轮廓上对三维目标进行形态描述也比较困难, 再加上跟踪效果与初步轮廓的检测或选取密切相关, 较难于全自动的检测跟踪系统。比较突出的计算方法有状态空间模型(State space models) [15]、变分法(Variational methods) [16]、启发式方法(Heuristic methods) [17]等。

像人体目标这样的复杂目标一般不能用简单几何形状来表示, 一般根据前几帧的检测结果生成复杂的形状模型来表示。比如, 人体形状除了可以用线图法和二维轮廓简单近似外, 还有利用广义锥台、椭圆柱、球等三维模型来描述人体的结构细节的立体模型和层次模型(包括四个层次: 骨架、椭圆球体模拟组织和脂肪、多边形表面代表皮肤、阴影渲染)。形状匹配法通过将目标模型投影到图像数据进行匹配来完成目标跟踪, 而形状模型则可以根据先验知识获取, 通常利用离线的人工测量、计算机辅助设计(CAD)和计算机视觉技术来构造模型。在车辆跟踪方面, 目前主要使用雷丁大学(University of Reading)提出的三维线框车辆模型[18]。模式识别国家重点实验室(NLPR)、德国卡尔斯鲁厄大学(University of Karlsruhe)的研究小组也基于该模型的车辆定位和跟踪方面做出了重要贡献。这类方法的匹配策略以 Hausdorff 形状匹配法[19]、Hough 变换[20]、直方图法等为代表[21]。

3. 张量投票方法在目标跟踪中的应用

在目标跟踪方面, 张量投票方法主要针对静止相机拍摄到的一系列运动目标图像的跟踪问题, 通过投票施加运动轨迹的平滑性约束, 从而进行不确定区域的合并以及提取出特征显著度、滤除噪声点、处理部分遮挡等[22] [23], 其计算框架如图 1 所示, 算法首先对输入图像进行运动区域检测, 然后生成各运动区域的包围框集合; 关键步骤在于生成输入数据的 $(2D + t)$ 维的张量表示, 进而通过张量投票过程进行运动速度估计, 主要特点在于可以通过张量投票施加运动曲线的平滑性约束优化目标运动速度的计算。

图 1 中, 运动区域检测采用的是文献[23]中的基于偏微分方程的方法, 输出为检测到的运动区域及其包围框。

张量投票算法的输入即为前面检测到的运动区域的包围框集合, 其 $2D + t$ 的表示方法中包含了包围盒中心在 t 时刻的坐标 (x_i, y_i, t) 、包围盒中包含的运动元素和包围盒本身, 如式(1)所示:

$$B = \{C_i^t, B_i^t, R_i^t\}, t = 0, \dots, n, i = 0, \dots, N(t) \quad (1)$$

式中, n 为运动图像序列的帧数, $N(t)$ 表示 t 时刻包围盒中包含的相连元素的数目, $C_i^t = (x_i, y_i, t)$ 代表包围盒中心的 $2D + t$ 坐标, R_i^t 代表运动元素的集合, B_i^t 表示包围盒。

包围盒的位置和速度是用 $3D$ 张量表示的, 设根据帧间对应点的位置计算出的包围盒速度为 $\mathbf{v} = (v_x, v_y, 1)$, 则其 $3D$ 棒形张量表示为:

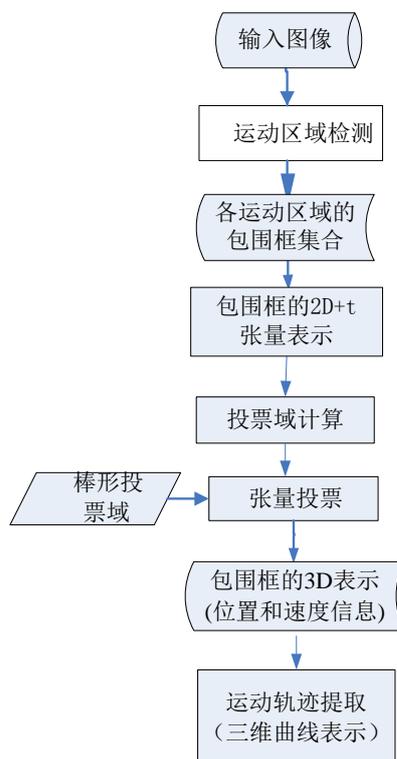


Figure 1. The computation flowchart of object tracking

图 1. 目标跟踪的计算流程

$$\mathbf{v}\mathbf{v}^T = \begin{pmatrix} v_x^2 & v_x v_y & v_x \\ v_x v_y & v_y^2 & v_y \\ v_x & v_y & 1 \end{pmatrix} \quad (2)$$

式中, \mathbf{v} 即为该张量的非零特征值对应的特征矢量, 该张量的显著度为 $(\lambda_1 - \lambda_2)$ 。

投票过程是在帧间进行的, 也就是说, 各数据点不会对同一帧的数据点进行投票。投票域的取向根据目标是匀速运动还是匀加速运动分别建立了线性模型和抛物线模型, 线性模型的投票域取向即为两点连线方向, 抛物线模型的取向则为曲线的切线方向; 投票的权重(幅度大小)综合了四个方面的因素, 即投票点和接收点之间的瞬时距离、投票者的显著度、抛物线模型中体现的加速度和两点邻域的遮挡问题。

具体来说, 投票点 $A:(x, y, t)$ 对接收点 $B:(x', y', t')$ 的速度估计为:

$$\mathbf{v}_{AB} = \left(\frac{x' - x}{t' - t}, \frac{y' - y}{t' - t}, 1 \right) \quad (3)$$

为防止对速度过大者或相关程度低的帧间投票, 这里还对投票域作了范围限定, 令 d_{\max} 为各帧间速度点的最大位置偏差, Δt 为投票点之间帧数的限制, 则要求两点之间满足以下条件方可进行投票:

$$|x' - x| \leq d_{\max}, |y' - y| \leq d_{\max}, |t' - t| \leq \Delta t \quad (4)$$

假设运动区域存在遮挡和分裂, 可通过不确定区域检测来解决这一问题。不确定区域定义为相关的包围盒 $(C_i' + B_i')$ 和 $(C_i'' + B_i'')$ 之间的且属于 $(C_i' + B_i')$ 的匹配候选点集。如计算得出包围盒 $(C_i' + B_i')$ 与所有相关包围盒的不确定区域总和占包围盒 $(C_i' + B_i')$ 的比例为 χ , 则投票的权重计算公式为:

$$w(\cdot) = \chi \cdot e^{-\alpha_i |t-t'|} \tag{5}$$

式中, α_i 为预先设定的常数, 与张量投票域的衰减常数功能类似。

计算出投票域的速度取向和权重之后, 即可开始投票过程。初始输入的数据点会对不确定区域中的数据点进行松散投票, 投票之后保留显著度最大点为包围盒 $(C_i' + B_i')$ 和 $(C_i' - B_i')$ 之间的公共区域的最佳匹配点。在此基础上, 再进行稠密投票, 根据运动目标点在各帧中的位置获得目标的运动轨迹。在此, 可发挥张量投票算法的优势来对运动轨迹施加平滑性约束并修复部分缺口。

示例应用如图 2 [22] 所示, 一般过程是先在运动图像序列中检测运动区域; 然后应用张量投票进行处理, 目标的三维空间位置数据构成三维张量, 再加上不同时刻的位置变化就构成了融合运动轨迹的四维数据, 但是对运动估计的滤波处理以及运动估计都是通过三维张量投票来实现的; 最后获得的就是选出跟踪区域的图像序列以及优化的运动轨迹曲线。

4. 结束语

总体说来, 目前跟踪的目标主要是人和车辆, 点跟踪是对运动目标作了最高程度的抽象, 其中基于 Kalman 滤波及预测的目标跟踪算法是目标跟踪的主流算法, 研究比较广泛; 核跟踪综合了目标的一些外观特征, 与点跟踪方法可以结合使用; 轮廓跟踪法和形状匹配法考虑目标的完整形状, 但是一般没有考虑遮挡情况的处理, 只是可以处理目标的分离和合并, 比如: 一个人拿了一件东西走了一段时间又将东西放下 [24]。尽管目标跟踪已经获得很多研究成果, 包括一些实时性和鲁棒性都很好的算法, 但是很多算法都是在一定约束条件下进行的, 因而与目标跟踪有关的特征选择、目标表示、运动估计等问题都值得进一步深入研究。时至今日, 实现复杂场景下的稳健跟踪依然是一个具挑战意义的研究课题。这种挑战绝大多数是来自视频序列中的图像变化和多个运动目标存在。目前所有的流行方法都不足以做到对光照、多运动目标的鲁棒性。所以说, 张量投票方法针对的正是目标跟踪中的难点问题。

张量投票方法除了可以应用于目标跟踪之外, 其思想还可应用于图像序列的插值, 现有的方法如灰度平均法、选择粘贴法、遗传算法和高维数值计算中的样条插值等方法均忽略了图像序列内在的几何构造。由于图像序列可以将不同的背景和运动的前景区分开来, 因而可在一定范围内仅对运动目标进行插值。于是可利用张量投票将运动目标的可能轨迹找出, 再结合一些限制条件选取目标合适的过渡位置,

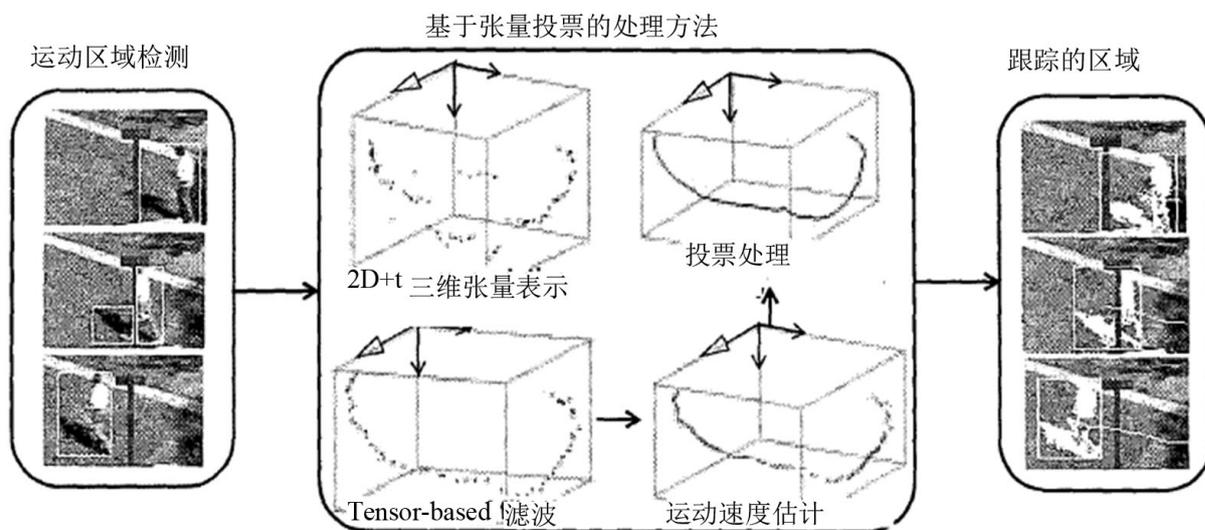


Figure 2. The process of object tracking
图 2. 目标跟踪过程

最后在保持原有几何元素不变的情况下重建过渡图像, 对于直线运动, 确定端点后即可找出运动方式; 对于曲线运动, 用若干较短的线段逼近曲线即可确定轨迹上的一些插值点。

参考文献 (References)

- [1] Yilmaz, A., Javed, O. and Shah, M. (2006) Object tracking: A survey. *ACM Computing Surveys*, **38**, 1-45. <http://dx.doi.org/10.1145/1177352.1177355>
- [2] Salari, V. and Sethi, I.K. (1990) Feature point correspondence in the presence of occlusion. *IEEE Transaction Pattern Analysis Machine Intelligence*, **12**, 87-91. <http://dx.doi.org/10.1109/34.41387>
- [3] Veenman, C., Reinders, M. and Backer, E. (2001) Resolving motion correspondence for densely moving points. *IEEE Transaction Pattern Analysis Machine Intelligence*, **23**, 54-72. <http://dx.doi.org/10.1109/34.899946>
- [4] Lochner, M. and Trick, L. (2014) Multiple-object tracking while driving: The multiple-vehicle tracking task. *Attention, Perception, & Psychophysics*, **76**, 2326-2345. <http://dx.doi.org/10.3758/s13414-014-0694-3>
- [5] Meyerhoff, H., Papenmeier, F. and Huff, M. (2013) Object-based integration of motion information during attentive tracking. *Perception*, **42**, 119-121. <http://dx.doi.org/10.1068/p7273>
- [6] Chevalier, F., Dragicevic, P. and Franconeri, S. (2014) The not-so-staggering effect of staggered animated transitions on visual tracking. *IEEE Transactions on Visualization and Computer Graphics*, **20**, 2241-2250. <http://dx.doi.org/10.1109/TVCG.2014.2346424>
- [7] Feria, C. (2013) Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. *Attention, Perception, & Psychophysics*, **75**, 53-67. <http://dx.doi.org/10.3758/s13414-012-0369-x>
- [8] 高琳, 唐鹏, 盛鹏, 等 (2010) 复杂场景下基于条件随机场的视觉目标跟踪. *光学学报*, **6**, 1721-1728.
- [9] 郭晓松, 李奕芃, 郭君斌 (2009) 贝叶斯目标跟踪方法的研究. *计算机工程*, **12**, 138-141.
- [10] Comaniciu, D., Ramesh, V. and Andmeer, P. (2003) Kernel-based object tracking. *IEEE Transaction Pattern Analysis Machine Intelligence*, **25**, 564-575. <http://dx.doi.org/10.1109/TPAMI.2003.1195991>
- [11] Shi, J. and Tomasi, C. (1994) Good features to track. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 21-23 Jun 1994, 593-600.
- [12] Tao, H., Sawhney, H. and Kumar, R. (2002) Object tracking with Bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis Machine Intelligence*, **24**, 75-89. <http://dx.doi.org/10.1109/34.982885>
- [13] Lukavsky, J. (2013) Eye movements in repeated multiple object tracking. *Journal of Vision*, **13**, 1-16. <http://dx.doi.org/10.1167/13.7.9>
- [14] Avidan, S. (2001) Support vector tracking. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, 8-14 December 2001, 184-191. <http://dx.doi.org/10.1109/cvpr.2001.990474>
- [15] Isard, M. and Blake, A. (1998) Condensation—Conditional density propagation for visual tracking. *International Journal of Computer Vision*, **29**, 5-28. <http://dx.doi.org/10.1023/A:1008078328650>
- [16] Bertalmio, M., Sapiro, G. and Randall, G. (2000) Morphing active contours. *IEEE Transactions on Pattern Analysis Machine Intelligence*, **22**, 733-737. <http://dx.doi.org/10.1109/34.865191>
- [17] Ronfard, R. (1994) Region based strategies for active contour models. *International Journal of Computer Vision*, **13**, 229-251. <http://dx.doi.org/10.1007/BF01427153>
- [18] Gardner, W.F. and Lawton, D.T. (1996) Interactive model-based vehicle tracking. *IEEE Transactions on Pattern Analysis Machine Intelligence*, **18**, 1115-1121. <http://dx.doi.org/10.1109/34.544082>
- [19] Rowley, H., Baluja, S. and Kanade, T. (2014) Tracking by location and features: Object correspondence across spatio-temporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, **40**, 159-171. <http://dx.doi.org/10.1037/a0033117>
- [20] Sato, K. and Aggarwal, J. (2004) Temporal spatio-velocity transform and its application to tracking and interaction. *Computation Vision Image Understanding*, **96**, 100-128. <http://dx.doi.org/10.1016/j.cviu.2004.02.003>
- [21] Kang, J., Cohen, I. and Medioni, G. (2004) Object reacquisition using geometric invariant appearance model. *Proceedings of the International Conference on Pattern Recognition*, Cambridge, 23-26 August 2004, 759-762. <http://dx.doi.org/10.1109/ICPR.2004.1333883>
- [22] Kornprobst, P. and Medioni, G. (2000) A 2D+t tensor voting based approach for tracking. *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona, 3-8 September 2000, 1092-1095.
- [23] Kornprobst, P., Deriche, R. and Aubert, G. (1999) Image sequence analysis via partial differential equations. *Journal of*

Mathematical Imaging and Vision, **11**, 5-26. <http://dx.doi.org/10.1023/A:1008318126505>

- [24] 张宇, 韩振军, 焦建彬 (2010) 一种基于综合特征评估的运动目标跟踪算法. *中国科学技术大学学报*, **5**, 491-495. <http://dx.doi.org/10.1109/ICPR.2000.903736>