

A Survey on Object Tracking

Jialong Xu

Aviation Military Affairs Deputy Office of PLA Navy in Nanjing Zone, Nanjing Jiangsu
Email: pugongying_0532@163.com

Received: Aug. 1st, 2015; accepted: Aug. 17th, 2015; published: Aug. 20th, 2015

Copyright © 2015 by author 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 analyzes the characteristics of the algorithm. At last, some future directions are suggested.

Keywords

Object Tracking, Track Alignment, Object Detection

目标跟踪相关研究综述

徐佳龙

海军驻南京地区航空军事代表室, 江苏 南京
Email: pugongying_0532@163.com

收稿日期: 2015年8月1日; 录用日期: 2015年8月17日; 发布日期: 2015年8月20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。本文分类总结了目标跟踪的相关工作, 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。它在对运动目标图像的分析 and 应用中产生, 并成为计算机视觉领域中视频分析的基本内容之一, 实质上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、视频检索等领域均有广泛的应用[1]-[3]。

本文旨在总结目标跟踪的相关工作, 并进行分析和展望。本文主要内容安排如下: 第二节是对相关工作的总结, 主要包括运动目标的表示与建模、特征选取、目标检测和目标跟踪; 第三节是总结与展望。

2. 相关工作

根据目标跟踪的一般计算流程, 本文分以下几个方面对相关工作分别进行描述: 运动目标的表示与建模、特征选取、目标检测和目标跟踪, 下面进行具体描述。

2.1. 运动目标表示与建模

在目标跟踪场景中, 任何感兴趣的事物都可以作为跟踪目标, 如在大海上行驶的船只、水族馆中的游鱼、路上行驶的车辆、空中飞行的飞机、街道上的行人甚至水中的水泡都可以作为跟踪目标, 因为在一些特定的场景中这些目标的跟踪有着非常重要的意义。跟踪目标一般来说可以由目标的形状或外观来表示。图 1 给出了目标模型的分类。从大的方面, 目标表示方法所获得的模型可以分为形状模型和综合模型, 形状模型只考虑目标形状这一个因素, 而综合模型则考虑形状和外观两方面的因素对目标进行表示。

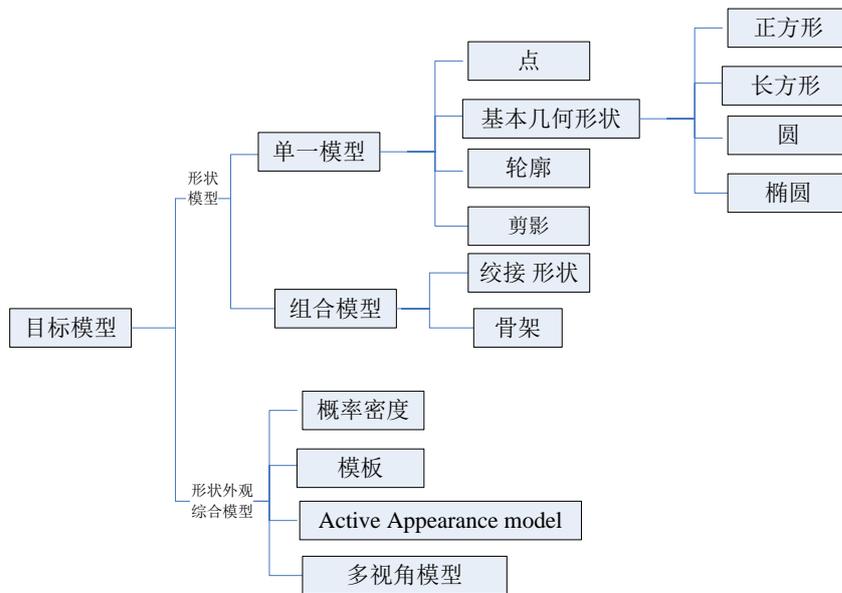


Figure 1. Classification of the object model

图 1. 目标模型的分类

形状模型中，单一模型包括：

- ◆ 点表示法：将目标由一点来表示，如由目标的中心点表示，或是由目标轮廓上的一系列关键点表示。
- ◆ 基本几何形状表示法：目标由一个框定目标区域的矩形或是椭圆形表示，在目标的运动过程中经常会发生目标运动的仿射变化、投影变化等形状上的改变。
- ◆ 目标的轮廓和剪影表示法：轮廓的定义是指目标的边界；目标轮廓的内部区域称为目标的剪影。组合模型则是在单一模型基础上，对复杂目标模型的组合，目前主要有两种方法：
- ◆ 铰接形状模型表示法：铰接目标指的是有多个部分通过关节铰接在一起的目标。
- ◆ 骨架模型表示方式：目标的骨架模型可通过中轴变换从目标轮廓提取。这种模型表示方式一般应用于目标识别。

综合模型是近年来兴起的一种目标建模方法，在实际跟踪中，目标的形状表示与目标的外形特征表示结合在一起。下面是一些目标跟踪中常用的表示方法。

- ◆ 概率密度模型表示法：目标外形特征的概率密度估计可以是有参数的如高斯模型、混合高斯模型，或是无参的密度估计比如 Parzen windows 和直方图方式[18]。
- ◆ 模板表示法：模板一般使用一些简单的结合形状或是轮廓表示。使用模板的目标表示方法的最大优点是，模板同时包含了目标的空间信息和外形信息。
- ◆ 主动外观模型法(Active appearance model)：主动外观模型是通过对目标的外观和形状同时建模得出的。一般来讲，目标的形状是由一系列的界标表示的。
- ◆ 多视点外形模型法：可以对目标的不同视角进行编码。其中的一种方法是在给定的视点中提取不同的子空间来表示不同视点的目标。

目标的表示是目标跟踪的基础，在目标跟踪过程中有着非常重要的意义。一个好的目标表示方法很大意义上能决定跟踪的成败与跟踪的性能。目标的表示到现在为止还没有一个统一的表示方法，一般都是基于不同的特定的跟踪场景选择目标的表示方式。

2.2. 特征选择

特征选取需考虑以下三方面问题：1) 鲁棒性，这主要是由于遮挡、光照变化及视角不同等造成的图像表象的变化，比如，边缘特征对场景中光线的变化不敏感，角点特征在图像中有很好的定位性；2) 简洁性，即用少量特征信息来描述目标，大大压缩目标的信息量，使整个算法简洁，也方便后续的处理；3) 可计算性，即特征的获得可能需要一定的图像处理方法，需要特征具有可检测性。实践证明，目标跟踪算法中选取什么样的特征，很多情况下并没有一个有章可循的方法。但是，特征选取的结果却会对跟踪的效率影响很大。

特征的选择与目标的表示方法密切相关的，比如说，颜色可以作为基于直方图的外观目标表示方法的特征，在很多情况下跟踪算法是使用多种特征的，一些主要的目标跟踪特征如下：

① 颜色特征。一个物体的颜色主要由两个物理特征因素决定。一个是光照环境；另外一个目标表面的反射特性。在图像处理中 RGB (Red, Green, Blue)颜色空间通常被用来表示颜色。

② 边缘特征。边缘定义为两个强度明显不同的区域之间的过渡，图像的梯度函数即图像灰度变化的速率将在这些过渡边界上存在最大值。

③ 光流特征。光流是一个密集场位移矢量，用来确定每一个地区的像素。光流计算使用亮度限制，假定相邻帧的亮度相素恒定。

④ 纹理特征。关于图像纹理并不存在严格的一般定义，我们把灰度分布性质或图像表面呈现出的方向信息称为纹理结构，它有助于区别不同的图像区域。

多数特征选取取决于其应用领域。然而，特征的自动选取问题已经引起了模式识别界的极大关注。自动特征选择方法可分为过滤器法和封装法。过滤器法是基于一般标准选取特征，比如，特征无关联性。

以上所有特征中，颜色是跟踪中使用的最广泛的特征。在使用的过程中，目标重心、边缘、面积、矩和颜色等一般作为整体特征；线段、曲线段和角点等则为局部特征提取结果；另外，还可将各种距离和特征间的几何关系组合成新的特征。

2.3. 目标检测

与目标的表示方法和所选取的特征相对应，目标检测方法分为点检测、图像分割、背景建模、监督分类等四种。其中，典型的点检测方法有角点检测算法[4]、尺度不变特征转换(Scale Invariant Feature Transform) [5]、仿射不变点检测法(Affine Invariant Point Detector) [6]；图像分割方法则以均值平移(Mean-shift) [7]、图分割(Graph-cut) [8]、主动轮廓模型(Active contours) [9]等为代表；背景建模法包括混合高斯模型[10]、特征背景模型[11]、Wall flower [12]、动态纹理背景模型(Dynamic texture background) [13]；监督分类的典型方法为支持矢量机(Support Vector Machines) [14]、神经网络(Neural Networks) [15]、Adaptive Boosting [16]。

2.4. 跟踪算法

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。

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，图 2 给出了各种运动约束形式的示意图，各图中， \circ 表示目标在第 $(t-1)$ 帧的位置， x 代表第 t 帧中目标的可能位置。

核跟踪方法可以选择模板等特征，典型算法有简单模板匹配、均值平移(Mean-shift)；也可以采用多视点外观模型，典型方法有 Eigentracking [17]、支持矢量机 [18]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[19]。

轮廓跟踪的方法以活动轮廓模型为典型代表[20]。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。

2.5. 形状匹配

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[21]。

3. 结束语

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

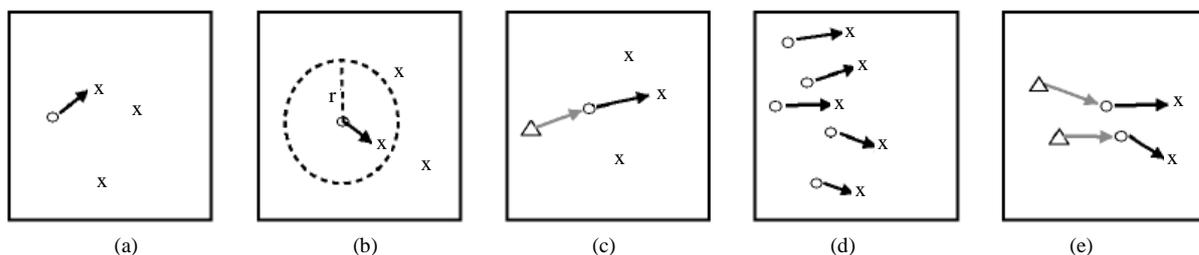


Figure 2. Diversified moving constraints. (a) Proximity; (b) Maximum velocity (r is the radius); (c) Minimum velocity; (d) The same velocity; (e) Rigid constraint

图 2. 不同的运动约束。(a) 邻近性; (b) 速度最大化(r 表示半径); (c) 速度变化最小化; (d) 共同运动性; (e) 刚性约束

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。

基于对现有工作的研究和分析，可以预测未来的研究方向：

1) 将来的总体趋势是借助信息融合的思想，充分利用不同角度对目标的描述信息、多种特征以及视频上下文关联信息，并对不同的算法也进行优化组合，最终目标是能够处理无任何约束条件的视频(尤其是新闻视频)中的运动目标，这些自然场景的视频图像往往是含噪的、被压缩的或非结构性的，而且很多经过了多个摄像机成像结果的剪辑，处理难度较大。

2) 遮挡情况的处理。目前采用的方法有通过预测跟踪跳过目标的完全遮挡的画面继续跟踪，通过选取一些特征使得在遮挡情况下也能通过特征检测定位目标，变换摄像机位置或多摄像机跟踪。

3) 目标特征的自动选取。目前 HMM 和 DBN (Dynamic Bayes Networks)等概率方法中，DBN 效果较好，只是 DBN 的有效求解方法仍然需要进一步研究。

4) 生物视觉系统的模拟。生物学结果表明，蛙眼有四种独立的图像动作使之对运动目标非常敏感：持续反差检测以检测尖锐边缘的存在性；本质突边检测用于检测暗运动目标是否具有弯曲的边界；运动边界检测；本质变暗检测则检测视野中的区域是如何变暗的。现在有人用自适应平滑滤波、Mean-shift 滤波、区域划分、轮廓提取等步骤，最后结合目标大小的先验信息获得期望的跟踪位置，基本上以 Mean-shift 方法的思路为主。如何更好地模拟蛙眼的生理特性也是一个值得深入研究的课题。

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] 陈宁强 (2010) 多目标跟踪方法研究综述. *科技信息*, 21-26.
- [3] 尹宏鹏 (2009) 基于计算机视觉的运动目标跟踪算法研究. 重庆大学博士论文, 重庆.
- [4] Harris, C.G. and Stephens, M. (1988) A combined corner and edge detector. *Proceedings of 4th Alvey Vision Conference*, 189-192. <http://dx.doi.org/10.5244/c.2.23>
- [5] Lowe, D. (2004) Distinctive image features from scale-invariant key points. *International Journal Computer Vision*, **60**, 91-110. <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [6] Mikolajczyk, K. and Schmid, C. (2002) An affine invariant interest point detector. *European Conference on Computer Vision*, 128-142.
- [7] 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>
- [8] 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>
- [9] 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>

- [10] Stauffer, C. and Grimson, W. (2000) Learning patterns of activity using real time tracking. *IEEE Transaction Pattern Analysis and Machine Intelligence*, **22**, 747-767. <http://dx.doi.org/10.1109/34.868677>
- [11] Oliver, N., Rosario, B. and Pentland, A. (2000) A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 831-843. <http://dx.doi.org/10.1109/34.868684>
- [12] 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>
- [13] Monnet, A., Mittal, A., Paragios, N. and Ramesh, V. (2003) Background modeling and subtraction of dynamic scenes. *Proceedings of the Ninth IEEE International Conference on Computer Vision*, Nice, 13-16 October 2003, 1305-1312. <http://dx.doi.org/10.1109/ICCV.2003.1238641>
- [14] Lukavsky, J. (2013) Eye movements in repeated multiple object tracking. *Journal of Vision*, **13**, 9-16. <http://dx.doi.org/10.1167/13.7.9>
- [15] Rowley, H., Baluja, S. and Kanade, T. (2014) Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, **40**, 159-171. <http://dx.doi.org/10.1037/a0033117>
- [16] Viola, P., Jones, M. and Snow, D. (2003) Detecting pedestrians using patterns of motion and appearance. *Proceedings of the Ninth IEEE International Conference on Computer Vision*, Nice, 13-16 October 2003, 734-741.
- [17] Black, M. and Jepson, A. (1998) Eigen-tracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computation Vision*, **26**, 63-84. <http://dx.doi.org/10.1023/A:1007939232436>
- [18] Avidan, S. (2001) Support vector tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, 8-14 December 2001, 184-191.
- [19] Rehman, A., Kihara, K., Matsumoto, A. and Ohtsuka, S. (2015) Attentive tracking of moving objects in real 3D space. *Vision Research*, **109**, 1-10. <http://dx.doi.org/10.1016/j.visres.2015.02.004>
- [20] Franconeri, S., Jonathan, S. and Scimeca, J. (2010) Tracking multiple objects is limited only by object spacing, not speed, time, or capacity. *Psychological Science*, **21**, 920-925. <http://dx.doi.org/10.1177/0956797610373935>
- [21] Zhang, S.L., Huang, Q.M., Jiang, S.Q., Gao, W. and Tian, Q. (2010) Affective visualization and retrieval for music video. *IEEE Transactions on Multimedia*, **12**, 510-522. <http://dx.doi.org/10.1109/TMM.2010.2059634>