

硕 士 研 究 生 读 书 报 告



题目 单视觉未校准相机实现多目标检测

作者姓名 刘溢辉

作者学号 NB15074

指导教师 李启雷

学科专业 移动互联网与游戏开发

所在学院 软件学院

提交日期 二○一五 年12月

Multiperson Tracking-by-Detection from a Single Uncalibrated Camera

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: LiQilei

By

LiuYihui

Zhejiang University, P.R. China

2015.12

摘要

本文主要研究解决在一个复杂的场景中使用单视觉且未标定相机来自动检测和追踪人或物的活动。本文主要讨论在这一过程中不可靠的信息来源，并通过建立基于粒子滤波的网络流模型将它应用到对多人的跟踪过程中，所涉及算法在避免建模和或地面校准的同时可以在复杂环境中检测并跟踪大量的移动物体。因此，强加限制非常少，适合于在线应用，该方法产生良好的跟踪性能应用在多种的高度动态的场景中，如典型的监控视频,摄像头录像等。这个方法结合了局 部策略和全局策略,有效地克服了基于检测的跟踪中存在的问题。这个跟踪方法 可以分为两个阶段:局部阶段和全局阶段。在局部阶段,完全离散的检测结果被 连成一段一段可靠的tracklets。在全局阶段,用网络流模型来表示tracklets间的数 据连接问题并转为求解最小代价流的问题。最后用一种两步优化的算法来处理长 时间遮挡的问题。

**关键词：多目标检测，运动分析，粒子滤波**

Abstract

In this research to solve in a complex scene using monocular mobile and to calibrate the camera to automatically detect and track the activities of people. This article mainly discuss in this process is not reliable sources of information, and applying it to the people in the process of tracking, the algorithm is involved in can avoid modeling and calibration or ground at the same time in detection and tracking of moving objects in complex environment. Therefore, impose restrictions were very small, suitable for online applications. Application of the method to create a good tracking performance in a variety of highly dynamic scenarios, such as the typical monitoring video, camera, video and so on .This method fuses the local and global strategies to effectively overcome the problems of tracking-by-detection and it can be divided into two stages: local stage and global stage. In the local stage, detection responses obtained from the high-quality object detector are combined by particle filter to generate reliable tracklets . In the global stage, the general network flow model is employed to solve the data association between tracklets. Afterwards, a double-step optimization is proposed to address long term occlusion.

**Keywords：Multi-object tracking, sports analysis, pedestrian detection**

1引言

摄像机越来越多的出现在世界各地，主要应用于监控或者其他方面。随着其使用程度的广泛增加，人们对于它自动实时处理大量数据的需求也逐渐提高。而在大部分的使用过程中，我们更感兴趣的是对人的行为的观察分析，例如在视频监控、自动辅助驾驶以及运动过程分析中，我们主要工作的是关于运动状态中人的行为，最常见的主要用于：

（1）安全性。在某些特殊的场合，如银行，商场等，跟踪系统可以找出从跟踪信息的异常信息，并预测潜在的危险，以保障人民群众的生命和财产安全。

（2）智能交通。跟踪系统可以统计和分析行人的号码将出现在该地区，如路口，步行街。有了这个统计数据，我们可以做出更好的运输计划，以消除交通堵塞。

（3）自驾驶汽车。每天都有很多人死于交通事故。有时候，人不能做一些动作来防止这种情况发生在极短时间——通常是小于0.5秒。代但是，机器可以代替人在足够短的时间内做出反应以防止交通事故。

然而，大多数的任务在处理这些信息的时候需要依靠一个轨道信息来用于探测和跟踪真实场景。要实现这种做法需要多个摄像头输入，详细的场景信息以及较为较为全面的静态场景及背景深度信息。这种做法并不适合在线处理。

在本文中，我们主要解决在一个复杂的动态场景中利用为标定相机来自动检测跟踪活动中的人的行为的问题。这是一个非常具有挑战性的问题，因为在实际操作过程中会存在很多“意外”因素像杂声，噪音，遮挡以及背景的切换等，不像基于3D信息的行人跟踪（通常需要多个摄像头的协作），目标被遮挡的情况在单视觉跟踪中几乎是不可避免的。该方法对多个对象跟踪大致可以分为两类。一类是基于本地策略，这解决了数据关联帧一帧和对象逐个对象。这一类包括均值漂移[1]，卡尔曼滤波器[2]，粒子滤波器[3]等。对于单目标跟踪，这些方法都有不错的表现，因为他们考虑了几个以前的帧进行数据关联的决定。不幸的是，这些方法不能很好地处理该对象的相互作用。另一类是基于使用的所有对象的信息，使全局最优协会决定全球战略。此类别包括多假设跟踪（MHT）[4]和联合概率数据关联滤波器（JPDAF）[5]。同时，大多数全局最优的方法也跟随跟踪——通过检测策略，如[6，7，8]中描述。这些方法通常使用原始检测在其输入中，构造一个模型来表示数据关联，制定这个模型的目标函数和通过优化算法解决该目标函数。这些方法的不足是跟踪器的性能在很大程度上依赖于检测的可靠性。网络流量模型是在全球战略中一个流行的图形为基础的模式。全局数据关联问题通常映射到网络模型，并通过优化算法解决的，如线性规划[9]，K-最短路径算法[10]和最小成本流算法[11，12]。

2 检测对象

因为物体检测是任何跟踪方法的重要阶段，本章节整理了在现代的用于跟踪目标的比较流行的方法。在自动跟踪系统中，尤其是在多个对象跟踪的过程中，检测器需要定位的每个对象中的每个帧。在这种情况下，物体检测也称对象定位。如果追踪对象是已经指定在第一帧中，检测器只需要提供在下面的帧中的特征或外观模型即可。

检测的目的是确定待检测对象的位置或者仅仅需要显示外观，以便跟踪器能够完成发现两个或几个连续的帧之间的数据关联的内容。最理想的情况是，检测器可以准确地提供在每个帧中的位置和每个对象的外观。错误的检测结果将直接导致跟踪出现误差甚至直接导致检测追踪结果差之千里。例如，如果检测器射门在一帧的对象时，跟踪器决不会找出对象的对应，这将导致未命中（假阴性）。因此它以一个合适的检测器组合以跟踪系统是重要的。马兹[13]列的检测方法和分类它们分为四类：点式探测器，分割，背景建模和监督分类。现代国家的最先进的和目标检测方法的典型框架是，他们通常使用的一些方法，如分割和兴趣点萃取，生成的外观模型或特征描述符，然后学习特定的分类。最流行的物体检测器，这是在这个框架内，是HOG检测，ISM探测器和基于部分模型检测。

2.1. HOG检测

 方向梯度直方图（Histogram of Oriented Gradient, HOG）是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子，是目前最流行的检测。它通过计算和统计图像局部区域的梯度方向直方图来构成特征。Hog特征结合SVM分类器已经被广泛应用于图像识别中，尤其在行人检测中获得了极大的成功。需要提醒的是，HOG+SVM进行行人检测的方法是法国研究人员Dalal在2005的CVPR上提出的，而如今虽然有很多行人检测算法不断提出，但基本都是以HOG+SVM的思路为主。

**（1）主要思想：**

       在一副图像中，局部目标的表象和形状（appearance and shape）能够被梯度或边缘的方向密度分布很好地描述。其本质是一个梯度的统计信息，而梯度主要存在于边缘的地方（如图2-1）。

**（2）具体的实现方法是：**

       首先将图像分成小的连通区域，我们把它叫细胞单元。然后采集细胞单元中各像素点的梯度的或边缘的方向直方图。最后把这些直方图组合起来就可以构成特征描述器。

**（3）提高性能：**

       把这些局部直方图在图像的更大的范围内（我们把它叫区间或block）进行对比度归一化（contrast-normalized），所采用的方法是：先计算各直方图在这个区间（block）中的密度，然后根据这个密度对区间中的各个细胞单元做归一化。通过这个归一化后，能对光照变化和阴影获得更好的效果。

**（4）优点：**

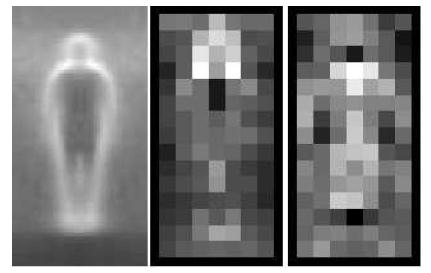
       与其他的特征描述方法相比，HOG有很多优点。首先，由于HOG是在图像的局部方格单元上操作，所以它对图像几何的和光学的形变都能保持很好的不变性，这两种形变只会出现在更大的空间领域上。其次，在粗的空域抽样、精细的方向抽样以及较强的局部光学归一化等条件下，只要行人大体上能够保持直立的姿势，可以容许行人有一些细微的肢体动作，这些细微的动作可以被忽略而不影响检测效果。因此HOG特征是特别适合于做图像中的人体检测的。

图2-1 HOG描述

2.2 ISM探测器

ISM探测器分为物体检测和分割两个交错的过程。在检测过程中可被描述为一个广义类，有紧密合作的过程中受益，相互借鉴，提高了综合性能。在开始前，局部出现，这是通过找到兴趣点和提取补丁围绕兴趣点，为特定对象产生的码本需要构造，明确了自下而上的方式，通过投匹配概率票，地方特色创建。图2-2显示的兴趣点（左）和围绕它们（右）的补丁。码本由中心的每个群集的，它由类似补丁的和它们周围的兴趣点的补丁映射到最接近码本词条，然后基于所述码本，所述隐形状模型的ISM（C）（C，Pc），它由所述码本（C）和一个空间概率C分类的PC，它指定，其中每个码本条目可以在物体上发现。接下来的程序是认可，该探测器适用于狗的兴趣点检测并提取补丁，并比较码本项投票对象中心可能的位置。随着观察e在位置i，概率投票对象n和位置x来进行计算：

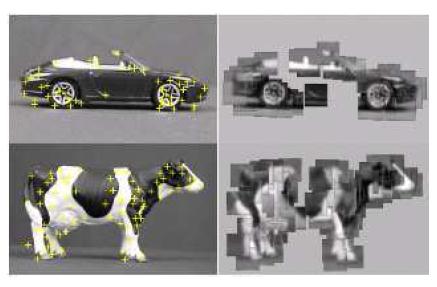


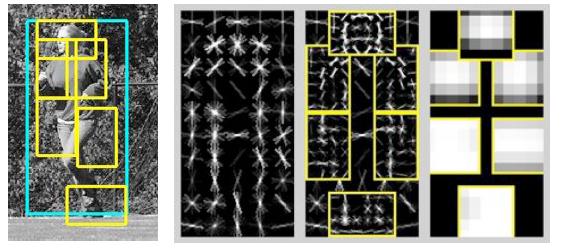
图2-2ISM探测光斑

p(on,x|e,l) ∑p(on,x|e,Ci,l)p(Ci |e,l)

∑p(x|on,Ci,l)p(on |Ci,l)p(Ci |e)

2.3 基于部分模型探测器

与上面的方法不同的是，这是一种通过有代表性的部分来替代整体的一种探测方法。由对象分别代替的单一实体的不同部分进行分类，这种方法对处理部分阻塞问题良好的性能如图2-3。针对整个部分中最灵活、维度最多的部分来建模这个模型包括了全局模板（根滤波器）为整个对象和更高的分辨率部分模板（部分过滤器）为对象的小零件。对于歧视性的探测，一些支持向量机作为移交潜变量，如位置，泛化和数据挖掘“硬负面”的例子在一个新的探测方法。

图2-3基于部分探测

3粒子滤波网络流模型

在最近几年中，跟踪——通过检测策略（依靠来自检测器（见图1）独立的检测追踪）变越来越流行，因为它使跟踪处理是自动的。无检测的帮助下，就很难使初始跟踪和处理的对象的出口和入口。跟踪逐检测的约束是检测的准确率，必须是合理的。随着国家的最先进的探测器[14，15]，大量的跟踪——通过检测策略的发展已经被提出。然而，跟踪——通过检测策略的一些问题依然存在。其中一个问题是，检测结果将误导跟踪器，如果它是不够准确，甚至它是否正确。第二个问题是在检测器不能给包藏对象的检测响应，如果一个对象被另一个或对象组被拥挤的阻塞。这个问题可能会导致跟踪不能保持跟踪的遮挡物。最后一个问题是，该检测器可以产生一些误报。这是可能的跟踪器都把误报作为一个真正的对象，如果假阳性出现在一个长的连续帧序列。

通过 [13]的方法的启发，我们基于带动下粒子滤波（BPF），有效地克服了跟踪逐检测策略的问题构造一个新的网络流模型。在我们的方法，全球数据关联问题转化为一个MAP估计问题，并根据最佳的最小费用流算法求解。

然而，对于网络图的最佳解决方案不一定是正确的，因为输入的检测可能已经有许多错误。一个改善的方法在网络流量模型的跟踪器的性能，让输入检测变得更加精确。不像在原始检测的检测反应，这是彼此独立的，由粒子滤波器的跟踪是一个连续的处理，这是利用先前帧的信息来处理当前帧。这可以纠正在原始检测许多错误，使得由颗粒过滤器产生的tracklet可以比原始检测更可靠。的双步骤优化设计通过连接类似tracklets，这可能代表了相同的目标，通过一个长的时间间隔，以解决长期闭塞问题。在这种双重一步的优化，我们首先找到所有tracklets的最佳解决方案，然后选择曲目，这可能是不完整的轨迹，在此解决方案做下一个最小代价优化。

3.1解决方法

不同于以往的全局最优跟踪逐检测方法，我们优选支离破碎轨道的检测，这是所谓的tracklet，是在我们的全局最优框架的基本单位。这些tracklets应该是可靠的，并将它们中的任何一个被认为是阳性。我们的方法可以分为两个阶段：阶段局部和全局的阶段。在本地台，我们结合从高质量对象检测器和微粒过滤器，以产生可靠tracklets获得的检测响应。然后，在全球阶段，我们采用的一般网络流模型来解决tracklets之间的数据关联。为了处理长期闭塞，我们构造一个出口/进口车型来选择需要的轨迹进行再加工，并将其添加到剩余的轨道集。然后，将残留的轨道之间的数据关联将由一般网络流模型再次求解。我们的方法的框架示如图3-1。

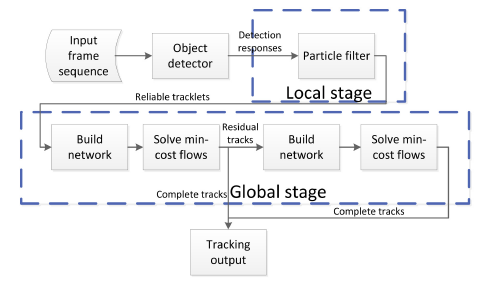


图3-1 网络流框架

3.1.1检测粒子滤波

使用基于部分检测的模型[18]，我们可以直接得到检测响应，即几个矩形框帧，并且得分检测假说是在每帧正的概率他们的信心。但也有大量的假阳性和假阴性的原始检测集。使用基于局部策略的跟踪方法中，我们可以消除大部分的假假说和修改的正确假设的位置和大小，使它们更精确。

3.2.2适用微粒过滤器

在这个阶段，我们得到基于BPF的tracklets。然后通过下面的公式来计算帧在t时间的取值：

cont i con1t i (1)con2t i

其中是加权因子取决于跟踪的情况下，包括由检测器和闭塞提供的检测。 cont i是tracklet帧t时间的取值。有时，我们更新通过从检测器获得的检测结果的观测模型。然而，检测结果如果不足够精确这可能会导致漂移。幸运的是，cont i可以降低漂移的概率。

3.2优化方案

在传统的成本流网络，节点n的数目可以比帧的数目N，因为每一个检测被添加到网络中大十倍以上。并且如果一个节点具有一个以上的候选连续节点，边缘米数量甚至可以是几百倍大于N。流行的最佳最小代价流算法，如按重新标签，连续最短路径（SSP）[14]和K-最短路径[12]，有过高的计算复杂性，以满足实际需求 ——时间处理。具体地讲，推重新标签的计算复杂算法，SSP和K-最短路径算法是O（mn2logn），O（knlogn）和O（K（mnlogn）），分别。 Pirsiavash等人[14]中描述的高效贪婪算法产生具有计算复杂度为O（kn）的近似解，其中k是磁道的数目。

对比基于所述检测的成本流网络，我们的网络要小得多，这通常是节点和边的量磁道只会是几十的数目少于10次。因此，完全至多数百个节点和边的我们的网络中存在。即使使用上述最优的最小代价流算法，根据他们的计算复杂度，运行时间来解决我们的网络很短，甚至可以不理。我们使用所有tracklets建立网络图，并应用SSP算法来找到最小成本的流动。

4 小结

随着计算机技术的帮助下，多个对象跟踪已成为计算机视觉越来越流行。本文概括目标检测与跟踪的历史，然后介绍网络流量模型的目标跟踪问题。本文着重于对网络流模型框架和算法基础的改善。从Online Multiperson Tracking-by-Detection from a Single Uncalibrated Camera这篇文章中我们得出的主要结论有以下几点：

（1）网络流模型的多目标跟踪的全局优化方法。此方法扩展最近的工作而制定的跟踪逐检测到最大后验（MAP）数据关联的问题。观测可能性和意见之间的亲和力被重新定义分别办理假阳性和长期闭塞。

（2）一种改进的贪婪算法被设计解决了原来贪婪算法的最小费用流代替，减少的ID的量显然切换。

（3）一种新的网络流量模型的基础上的简洁可靠的检测序列，能够对多个对象进行跟踪。这个模型融合了本地和全球数据关联的战略合作，弥补各自的不足，可以分为两个阶段：局部舞台和国际舞台。在当地的阶段，我们按照跟踪，通过检测框架的基础上提高了粒子滤波产生信心tracklets。在全球舞台上，tracklets之间的数据关联问题是制定一个最大-后验（MAP）的问题，并通过典型MIN-费用流算法求解。

（4）在本地和全球框架，基于tracklet网络流模型，双步优化是为了解决长期闭塞。实验结果表明，该方法优于国家的最先进的几种方法多目标跟踪。

当然，文中的方案也存在一些不足之处，主要有：

（1）系统仍然不适合实际应用。它花费的时间来检测行人大量（每帧20秒）。真实的场景将比我们的评估数据集更复杂，因此，跟踪结果也不能满足精度要求。

（2）在本文中提出的方法依赖于被设置为根据实验的一些因素。因此，如果我们改变了数据集的因素必须手动复位。他们应该通过培训过程中自动获得。

参考文献

[1] D. Comaniciu, V. Ramesh, and P. Meer, Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(5): 564-577, 2003.

[2] T. J. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(1): 90- 99, 1986.

[3] R. Hess and A. Fern. Discriminatively trained particle filters for complex multi- object tracking. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 240-247, 2009.

[4] R. Streit and T. Luginbuhl. Maximum likelihood method for probabilistic multihypothesis tracking. International Symposium on Optical Engineering and Photonics in Aerospace Sensing, 394-405, 1994.

[5]  Y. Bar-Shalom. Tracking and data association. Academic Press Professional, Inc., 1987.

[6]  A. Andriyenko and K. Schindler. Globally optimal multitarget tracking on a hexagonal lattice. 2010 European Conference on Computer vision, 466-479, 2010.

[7] A. Andriyenko, K. Schindler, and S. Roth. Discrete continuous optimization for multi-target tracking. IEEE Conference on Computer Vision and Pattern Recognition, 1926-1933, 2012.

[8] J. Berclaz, F. Fleuret, and P. Fua. Multiple object tracking using flow linear programming. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 1-8, 2009.

[9] J. Berclaz, F. Fleuret, and E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(9): 1806-1819, 2011.

[10] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. IEEE Conference on Computer Vision and Pattern Recognition, 1-8, 2008.

[11] H. Pirsiavash, D. Ramanan, and C. Fowlkes. Globally optimal greedy algorithms for tracking a variable number of objects. 2011 IEEE Conference on Computer Vision and Pattern Recognition, 1201-1208, 2011.

[12] A. Yilmaz, O. javed, and M. Shah, Object tracking: A survey. ACM Computing Surveys, 2005.

[13] A. Yilmaz, O. javed, and M. Shah, Object tracking: A survey. ACM Computing Surveys, 2005.

[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. IEEE Computer Conference on Computer Vision and Pattern Recognition, 886-893, 2005.

[15] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. International Journal of Computer Vision, 77(1): 259-289, 2008.

[16] N. Gordon, D. Salmond, and A. Smith, “Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation,” IEE Proc. F Radar Signal Processing, vol. 14, no. 2, pp. 107-113, Apr. 1993.

[17] H. Grabner and H. Bischof, “On-Line Boosting and Vision,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006.

[18] [Y Zheng](http://xueshu.baidu.com/s?wd=author%3A%28Yu%20Zheng%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson" \t "_blank)，[A Mahabalagiri](http://xueshu.baidu.com/s?wd=author%3A%28Mahabalagiri%2C%20A.%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson" \t "_blank)，[S Velipasalar](http://xueshu.baidu.com/s?wd=author%3A%28Velipasalar%2C%20S.%29%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight%3Dperson" \t "_blank) [Detection of moving people with mobile cameras by fast motion segmentation](http://xueshu.baidu.com/s?wd=paperuri%3A%285eb319af86055acadb2ee605ec0dd897%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6778215&ie=utf-8" \t "_blank) International Conference on Distributed Smar Cameras2013:1-6