文章编号:1003-6199(2018)01-0073-10 DOI:10.16339/j.cnki.jsjsyzdh.201801016

在线多目标视频跟踪算法综述

李月峰[1,2†],周书仁[1,2]

(1.综合交通运输大数据智能处理湖南省重点实验室(长沙理工大学),湖南 长沙 410114;
2.长沙理工大学 计算机与通信工程学院,湖南 长沙 410114)

摘 要:视频多目标跟踪是计算机视觉领域重要的研究课题之一,不论是在军用还是民用都有广泛应用。目前对单目标的跟踪算法研究已经相当成熟,但对于多目标跟踪的研究还处于发展阶段。重点研究了多目标跟踪过程中的四个重要阶段:特征提取、检测器、数据关联、跟踪器。特征提取阶段详细介绍了目前主流的特征提取方法以及各个方法之间的优缺点;检测器阶段首先详细介绍了目标外观模型在具体应用场景中的跟踪效果,接着对基于检测跟踪的多目标跟踪算法和基于深度学习的多跟踪算法进行了分析;跟踪器阶段分别介绍了目标运动模型的建立和利用不同跟踪器混合的多目标跟踪算法;数据关联阶段分别介绍了基于能量最小化的多目标跟踪以及常用的数据关联算法。接着,介绍了目前主流的数据集以及评测方法;最后对多目标跟踪未来的发展进行了思考和展望。

关键词:视频分析;计算机视觉;多目标跟踪;深度学习

中图分类号:TP391 **文献标志码**:A

Survey of Online Multi-object Video Tracking Algorithms

LI Yue-feng[1,2†], ZHOU Shu-ren[1,2]

(1. Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, (Changsha University of Science and Technology), Changsha, Hunan 410114, China;
2. School of Computer & Communication Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114, China)

Abstract: Video multi-object tracking is one of the important research topics in the field of computer vision, which is widely used in military and civil areas. At present, the research of single object tracking algorithm has quite mature, but for multi-object tracking of the research is still ongoing. This paper focuses on four important stages in the multi-object tracking process: feature extraction, detector, data association and the tracker. The feature extraction part introduces the current methods of feature extraction, as well as the merits and demerits of each method; In the stage of detection, the tracking effect of the object appearance model in specific applications is described, and then we analyze the multi-object tracking algorithm based on detection and tracking as well as the multi-object tracking algorithm based on deep learning; In the tracking phase, the establishment of object motion model and multi-object tracking with different tracker hybrid algorithm are introduced; During the stage of data correlation, we introduce the multi-object tracking based on energy minimization and commonly used data association algorithm, respectively. Then we introduce the current mainstream datasets and evaluation methods. Finally, the future development of the multi-object tracking is discussed and forecasted.

Key words: video analysis; computer vision; multi-object tracking; deep learning

---

收稿日期:2017-09-05
基金项目:国家自然科学基金资助项目(61402053,61602059);湖南省教育厅科学研究资助项目(16C0046,16A008,17A007)
作者简介:李月峰(1992—),男,湖北荆门人,硕士研究生,研究方向:图像处理,目标检测与跟踪。
† **通讯联系人**,E-mail:530481021@qq.com

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1 引 言

目标跟踪作为计算机视觉领域的一个分支，多年来一直是研究的热门方向[1]。特别是在复杂背景下，当目标所处环境发生变化或者目标被遮挡时，使得目标跟踪更具挑战性。按照跟踪目标的数量来分，目标跟踪可以划分为单目标和多目标两大类，其中对单目标跟踪的研究，国内外方法已经相当成熟，但对多目标跟踪的研究还处于发展阶段。视频目标跟踪广泛应用于人机智能交互、视频监控以及军事武器应用等领域[2]。正是因为目标跟踪的应用范围之广，所以一直以来都备受关注。

多目标跟踪算法研究至今，目前主流的研究框架有两种：一种是基于检测跟踪的框架（tracking-by-detection）；另一种是最近几年热门的基于深度学习（deep learning）的框架。在基于检测跟踪的框架算法中，常使用两种模型来建模，分别是外观模型和运动模型。外观模型，主要是对目标的整体外观特征进行建模，尽最大可能将目标与背景分离；运动模型，主要是对目标的运动特性进行建模，预测目标的位置，挖掘帧间的相关信息，然后通过事件分析来获得目标的运动轨迹。

随着深度学习和大数据时代的到来，在新环境下多目标跟踪领域出现了许多新的研究方法。目标检测和目标识别是目标跟踪的前提，而最近几年深度学习方法在目标检测和目标识别领域取得了巨大的成功，极大地促进了深度学习在目标跟踪领域的发展。可以说深度学习给计算机视觉领域带来了一场新的革命。

2 研究难点

目前，多目标跟踪方法的研究成果在一些简单场景已经有不错的跟踪效果，但仍存在许多问题。背景的复杂性、目标的多样性、背景与目标的相似性以及多个目标之间的相互影响，都将给多目标跟踪方法的研究带来困难。当前多目标跟踪的难点主要存在以下几个方面：

1）遮挡问题：遮挡是目标跟踪过程中的常见现象。遮挡可分为自身遮挡、被背景中静止的实物遮挡、被其他目标遮挡这几种情形，另外遮挡的程度也有不同。目标遮挡过程分为两个阶段：第一阶段是目标慢慢被遮挡阶段，在这个过程中目标可检测到的信息越来越少，因此常出现目标丢失的现象；第二个阶段是目标遮挡后重新被跟踪，也叫目标重现。重现的过程目标信息逐渐恢复，如何处理重现后目标再次被准确跟踪也是研究难点。遮挡会导致目标有效信息丢失，尤其是在目标长时间被遮挡的情形下。目前，多数系统都不能很好的解决长时间遮挡的问题，如何寻找一个有效地解决方法将是学者们不断研究的目的。

2）背景复杂度问题：目标背景的复杂度和稳定性将决定最后跟踪的难度和效果。背景中的干扰主要包括：光照的亮度变化、背景中物体的变动、场景的变换以及阴影等。背景的颜色容易随着光照条件的变化而改变，而目标又在实时的运动，所以需要及时对目标进行更新处理。由于背景中不止目标一个，其他类似于目标的物体会导致误检，这将大大增加对目标跟踪的难度。此外，阴影会导致背景颜色的差异，给目标检测带来困难。

3）数据关联：数据关联也是多目标跟踪的难点，关联算法的效率和复杂性将直接影响跟踪效率和准确率。数据关联是将追踪器跟踪的目标轨迹与目标预测的运动轨迹进行比较配对后，来最后确定目标正确运动轨迹的过程。单个目标不需要建立数据关联，当存在多个目标且目标之间很相近时，数据关联就显得尤为重要，同时关联的过程也将更复杂。目前，虽然已经有很多数据关联算法，但如何建立更高效的数据关联算法也是研究的难点之一。

3 视频多目标跟踪框架

目前研究领域内多目标跟踪的框架有很多，由于不断发展出现了许多改进的框架。一般多目标跟踪系统框架如图 1 所示。

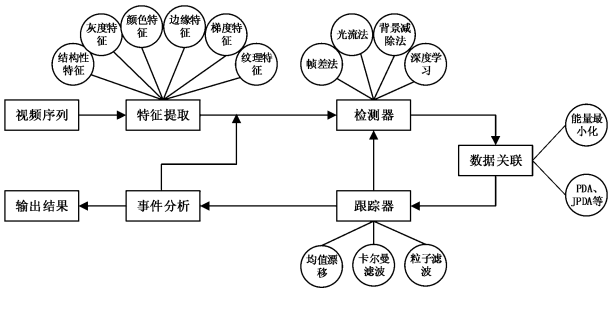


图 1 多目标跟踪系统框架

检测器中包括对目标外观模型的处理，跟踪器包含对目标运动模型的处理。外观模型，包含目标特征提取、分类匹配以及更新三个部分，接着通过数据关联，挖掘前后帧之间的关联信息，在运动模

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

型中通过搜索和采样目标潜在的空间位置信息,为后面分类匹配提供可靠的样本,然后计算比较这些候选样本的可信度分数来进行预测,得分最高的样本即为预测结果。另外,通过运动模型预测得到的目标位置,有一部分会再输入到检测器中,用于处理经过遮挡后目标重现的情况。最后通过事件分析消除噪声干扰,得到目标连续的运动轨迹。

4　特征提取

特征提取是目标检测的第一步,能否提取合适的特征直接影响到后面对目标检测和跟踪的效果。随着目标形状在检测和跟踪过程中的变化,理想的特征要具有很强的区分性,另外对目标的变化要具有较强的鲁棒性。特征的选取往往依赖于目标的表示方法。目前对目标的表示方法主要有以下 6 种,分别是:灰度特征、颜色特征、边缘特征、纹理特征、梯度特征和结构性特征。但在实际的研究中,由于场景的复杂性,往往需要将几种方法混合使用。

灰度特征是最简单和直观的特征表达方式,其计算速度快,为行人检测的可实用化点燃了希望之火。该特征有三种表现形式,分别为:原始灰度特征、灰度直方图和 Haar-like 特征。前两种特征方法都是首先将视频图像灰度化,进而利用灰度图中的像素灰度信息,代表性算法分别是 CSK 算法和 LSHT 算法。Haar-like 特征系列是一个典型的标量特征,对于图像的边缘、垂直、水平敏感等优点被广泛应用于目标检测当中,代表性算法是 CT 算法等。

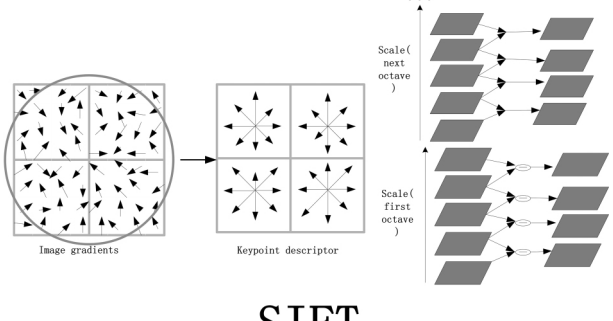
颜色特征是最基本的特征表示方式,它通过不同的颜色对目标外观进行标记。颜色直方图是颜色特征的一种常用表现形式,它通过计算每种颜色在图像颜色空间中的比例来代替计算每种颜色在图像空间中的位置。近几年,出现了一种新的颜色特征方法—Color Name 特征[3],该特征具有更好的表征能力。颜色特征优点是对图像本身的特征属性依赖性小,具有很好的鲁棒性;缺点是受光照的变化影响较大,且容易被其他背景颜色干扰。

边缘特征是物体与背景边缘产生剧烈地变化得到的,在跟踪过程中可以通过边缘检测的方法来检测出边的变化。与颜色特征相对比,边缘特征的产生依靠目标与背景产生运动,而不依赖人对颜色的主观认识。但是边缘特征在检测过程中容易出现边缘曲线断开的现象,导致图像边缘模糊的问题,因此在实际运用中要考虑边缘检测算法的准确性。

纹理特征是把图像看成一个平面,通过判断平面密度的变化,来对图像的平滑度和规律性进行量化,从而描述图像上的空间颜色和光强分布。纹理特征并不是直接获取得到的,而是通过图像预处理产生。纹理特征可分为以下四种类型,分别是:统计型纹理特征、模型型纹理特征、信号处理型纹理特征和结构型纹理特征。常用来描述纹理特征的算子有 LBP(Local Binary Patterns,局部二值模式),它具有很好的旋转不变性和灰度不变性等优点,所以应用广泛。

梯度特征是通过统计图像局部区域的梯度方向来构成全局特征。常用的梯度特征有 SIFT 特征和 HOG 特征,如图 2 所示。SIFT 特征和 HOG 特征相比,相同点都是利用的局部信息特征,但后者的实时性好一些。SIFT 特征具有尺度不变性,即使图像发生旋转、尺度缩放或亮度变化等也能保持很好的检测效果。HOG 特征更多的是利用局部分块单元的梯度信息,用局部统计信息来表示整个图像信息。梯度特征因环境变化产生的影响相对要小,性能稳定,但它不能直接反映出物体的尺寸和姿态等信息。

结构性特征是通过神经网络训练得到的,从网络底层越往高层特征就越复杂。高层的特征由底层特征逐层组合表示,随着层数的增加,可用的特征信息就越多,参考信息也就越丰富,准确性也会提高。利用神经网络得到的结构性特征具有很强的区分性,用一些简单的分类器就可以实现高精度的分类效果。但是,该方法也存在缺点,特征增多带来的是计算复杂和搜索空间增大,所需要训练的时间也会变长,所以并不是训练层数越多就越好。目前对结构性特征的研究主要集中于在满足网络结构功能的前提下,改进网络结构,微调网络来缩短训练的时间,进而得到目标更好的表现特征。



SIFT

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

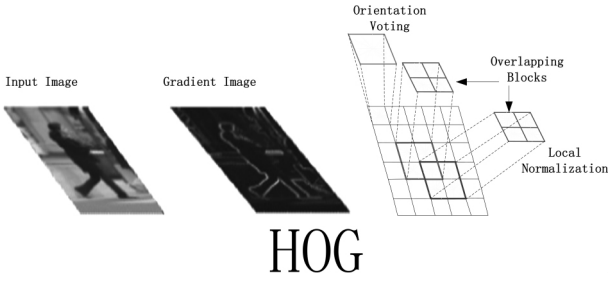


图 2 SIFT 与 HOG 梯度特征图示

5 检测器

5.1 外观模型

在完成对目标的特征提取后，我们需要对目标进行检测，而在检测的过程中就要利用到外观模型。外观模型是视频目标跟踪中的重要一步，好的外观模型可以提高跟踪的性能。目前在目标检测领域的检测算法有很多，常用的有光流法、背景减除法和帧差法等。近几年来，随着深度学习方法应用于目标检测领域，检测器外观模型得到了极大的发展，出现了很多基于深度学习的检测算法，如 RCNN 算法[4]，Fast-RCNN 算法[5] 以及 Faster-RCNN 算法[6]。

按照目标与摄像机之间的运动关系，目标检测可以分为两类，即：基于静态背景下的目标检测和基于动态背景下的目标检测。基于静态背景下目标检测摄像机在整个监控过程中是不动的，只有被监控的对象在动；基于动态背景下目标检测摄像机和被监视的目标都在运动，两者之间产生了复杂的相对运动。

5.1.1 静态背景

1）光流法

光流法是最早应用于目标检测的方法，它利用图像像素随时间变化的光流特性，来计算运动物体相邻帧之间的运动信息。光流法既可以运用于动态场景也可以运用于静态场景，而且无需预先知道场景中的任何信息。但是光流法计算复杂，容易受环境干扰，实时性也很差，容易产生噪声。另外，当把真实场景中的三维对象投射到二维平面上时，光流的信息会存在损失，这样很容易引起遮挡问题和孔径问题。

2）背景减除法

背景减除法从字面意义来讲就是一种能区分视频序列中目标背景和环境背景的方法，算法通过将当前帧与背景图像的像素值进行差分来实现检测。随着目标的运动，所处的环境一直都在变化，因此背景减除法的关键就在于对背景的建模和更新。鉴于上面的原因，所以背景减除法容易受环境干扰，当环境出现光照变化、阴影等其他干扰时，都将影响到算法的检测效果。但是背景减除法拥有很好的实时性，同时计算复杂程度小，所以应用比较广泛。

3）帧差法

帧差法是通过将图像序列中相邻两帧或多帧信息作差分运算，来获得运动目标轮廓的方法，对存在有多个运动目标的场景检测效果很好。帧差法适应动态环境能力较强，受光照变化干扰影响小，算法复杂度也小，稳定性较好。算法通过判断检测物体的速度快慢来调节帧间时间间隔，因此在检测时由于要不停的调整帧间时间间隔，会出现不连续的空洞，导致有时会检测出不完整的目标轮廓。

4）神经网络

基于深度学习的检测方法，是指通过神经网络将视频第一帧的信息作为样本，训练出一个能追踪目标运动的深度模型。但大部分物体在运动过程中会发生形变，导致目标识别和追踪难度较大，易丢失目标。为了解决这个问题，人们利用迁移学习的思想，采用数据扩充（Data Augmentation）等方法扩充训练样本，预先对非跟踪目标进行网络训练，以学习到不同物体在不同环境（如光照变化、背景变换等因素）下的通用表征特征，在跟踪时利用前几帧信息微调网络，进而使网络适应不同场景下的目标跟踪。

上面一小节介绍了目前常用的检测方法，每一种方法都各有所长，在实践应用时要根据具体的环境和条件采用不同的检测方法，同时为了达到最好的实验效果，一般会将上述方法中某几个联合使用。

5.1.2 动态背景

由于目标与摄像机之间的相对运动，基于动态背景下的目标检测要比基于静态背景下的目标检测复杂得多。一般来说，摄像机的运动可以分为两种类型：

1）摄像机的支架是不动的，但摄像机可以进行左右或者上下偏转；

2）将摄像机安装在移动的物体上有规律的移动。

鉴于以上两种运动情况，目标前景和背景都在做全局运动，所以首先要对图像进行全局运动估

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

计。虽然图像上每个点的运动矢量不同,但是他们是在同一台摄像机上采集到的,因此可以用同一个参数模型。这样全局运动估计问题就简化为全局运动模型参数的估计问题,通常采用块匹配法或光流估计法来估计运动参数。块匹配法有以下三项关键技术:

1)匹配法则,如最大相关、最小误差等;

2)搜索方法,如三步搜索法、交叉搜索法等;

3)块大小的确定,如分级、自适应等。

光流估计法是利用光流方程求解图像像素点运动速度的一种方法。它能够应用于摄像头与运动目标发生相对运动的情形,但是该方法在处理复杂场景时效果差,容易受环境干扰。

5.2　基于检测跟踪的多目标跟踪算法

上述两小节介绍了目前比较常用的检测方法,检测的结果最终将用于目标跟踪的输入。由于现实场景的复杂性,很难用单一的特征来表示目标,因此出现了很多基于多特征融合和自适应模板的多目标跟踪方法。

针对单一特征的不稳定性,文献[7]提出了基于片段和多特征自适应融合的目标跟踪。具体实现时,通过导入颜色、HOG 特征和角特征的片段来区分不同类型的遮挡,然后对不同的遮挡采用不同组合方法进行前后信息的匹配,来重建目标遮挡前后轨迹信息,进而可靠的跟踪目标。文献[8]将卡尔曼滤波算法和均值漂移算法联合起来应用,通过将追踪器分为不同的层次,来对不同的检测模板进行追踪,文中以追踪器所拥有检测模板的数量为基准将追踪器进行分层,通过一个固定阈值来进行层次间的转换,达到一个在线自适应的跟踪。针对长时间遮挡问题,文献[9]采用粒子滤波对目标进行跟踪,处理遮挡时通过比较遮挡前后粒子的位置,线性估计粒子在被遮挡期间的运动方向,然后筛选匹配这些预测位置从而得到目标遮挡前后完整的运动轨迹。文献[10]在文献[9]的基础上提出了基于多特征融合的分层粒子滤波器(HPF)框架。所提出的 HPF 框架采用不同的特征信息,不仅可以处理长时间遮挡,而且对复杂环境中针对单个特征跟踪失败的情况处理效果很好。文献[11]同样采用了分层的思想,与文献[8]不一样的是,文中将检测器也进行了分层。具体实现时,第一层负责生成新的模板和处理上一帧生成的轨迹;第二层重点处理轨迹漂移现象;第三层处理剩余的候选模板;最后一层处理目标经过遮挡而产生的轨迹碎片问题,通过分层关联信息,达到了最好的分类处理效果,因而能够更好地跟踪目标。

经过十几年的发展,基于检测跟踪的多目标跟踪方法已经有很多,目前研究方向更趋于结合特定的应用去寻求新的突破,在未来的发展中,目标检测和目标跟踪之间的关系将会更加密切,同时目标检测的方法也可以辅助目标跟踪,从而进一步提高跟踪的准确率和鲁棒性。

5.3　基于深度学习框架的多目标跟踪算法

深度学习成为当下最热门的研究学科,各行各业都在将深度学习的方法融入本行领域中。目前深度学习已经在语音识别、图像识别、自然语言处理等领域取得巨大突破,并且还在不断发展变化。但是,在目标跟踪领域,深度学习取得的成果不是很突出,其中的主要原因有以下几点:

1)在视频多目标跟踪系统中,一般只对视频第一帧的图像数据进行标注,之后数据信息更新的数量级也不大,这使得深度学习无法发挥其在处理大数据方面的优势。

2)因为视频是实时在线的,因此视频多目标跟踪系统对算法的实时性要求高。目前,应用深度学习方法的框架,一般都拥有庞大的网络结构,因而很难满足实时性和速度的要求。

虽然存在以上几点问题,但是已经有学者不断探索和改进网络结构,来适应具体的跟踪环境。TLD(Tracking-Learning-Detection)算法是最早将深度学习理论运用到目标跟踪的算法,它在原始检测和跟踪的环节中增加了学习训练阶段,使算法能够快速学习目标特征,并用少量先验知识进行跟踪,因其高效率和高准确率,算法刚问世就广受学者追捧。但算法本身也有缺点,首先算法开始只能对单目标进行跟踪,同时需要离线训练。但随着不断改进,衍生出一些新的 TLD 算法。文献[12]通过改进的 TLD 算法,提出了 ATLD (Accelerated TLD)算法,新算法不仅提高了运行速度,而且实现了对多目标的跟踪。文献[13]基于 TLD 算法,提出了一种新的 P—N 实时学习的方法,实现了在线训练。同时,新算法在存在遮挡和杂波的情况下跟踪效果很好。

通过比较先前的 TLD 算法,有学者提出了新的 TLP(Tracking-Learning-Parsing)算法,文献[14]将 TLP 算法与 AOG(And-Or Graph,与或图)相结合,通过 AOG 网构建树,采用或、与、非三种门来提取不同的特征信息,用 TLP 算法去动态分析跟踪目标,相比先前的 TLD 算法更具有表现特征、更灵活。通过学习跟踪的方法逐渐成为了多

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

目标跟踪研究的重点，相对于检测跟踪方法，学习跟踪方法的精度要高很多。

随着 CNN(Convolutional Neural Networks，卷积神经网络)在目标检测领域取得的巨大突破，其成果也间接推动了目标跟踪研究的发展。一般卷积神经网络基本结构如图 3 所示。

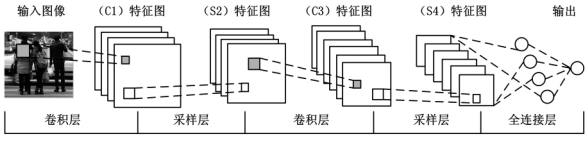


图 3 CNN 网络结构图

从图中可以看出，其基本结构包括：输入层、卷积层、采样层、全连接层和输出层。其中特征提取是核心，它是由卷积层和采样层交替进行的，从多个特征图中得到特征向量，最后接入全连接层和输出层。

通过网络结构训练分类得到的目标特征，更具有区分性，大大提高了对目标的识别能力，进而提高了目标跟踪的精度。文献[15]介绍了基于 CNN 的数据关联方法。文中首先训练 Siamese 卷积神经网络，来获得输入前后帧图像的结构性特征向量，接着分别比较 CNN 输出的特征图像和经过梯度增强分类器处理后的图像之间的匹配概率，通过概率大小来确定目标的位置，从而规划出目标的运动轨迹。有学者提出，一个高性能的检测外观特征可以产生高精度的跟踪效果，文献[16]先通过 CNN 做检测，得到良好的外观模型特征，再通过结合了 POI(Person of Interest)区域的卡尔曼滤波器去跟踪，得到了很好的实验结果。文献[17]采用对冲算法，提出了一种新颖的基于 CNN 的跟踪框架，算法将不同的 CNN 层中获取的特征融入 KCF 跟踪器中，得到许多弱分类器，然后使用 Hedge 算法把这些弱分类器训练得到一个强大的跟踪器去追踪目标，算法流程如图 4 所示。文献[18]用 RNN(Recurrent Neural Networks，循环神经网络)建模，相比之前用 CNN 训练学习的方法，RNN 更关注的是记忆功能，通过不断地记忆和学习，进而形成一条链式序列，再通过 LSTM(Long Short-term Memory)来进行数据关联，最终实现了端到端的跟踪。

基于深度学习的目标跟踪方法是目前最为火热的研究方向。其主要的研究方法有以下两种思路：

1)先通过离线训练大规模的图像和视频数据，

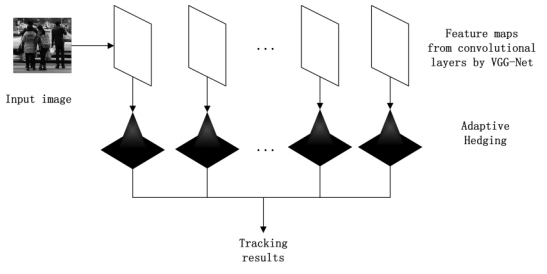


图 4 Hedge 算法流程图

来构造某一类别特征的深度模型网络。然后具体在线跟踪时，用之前训练好的深度模型网络去提取目标的特征，再利用在线检测的数据去微调网络，以达到适应目标外观的变化。这是一种“离线＋在线”的设计方案，目前多数系统采用的是此设计方法。

2)将深度神经网络模型的内部结构进行一定的改变，微调网络来满足在线跟踪的要求。同时在满足网络性能要求的条件下，降低网络的层数来缩短训练的时间，简化训练过程。这是一种直接“在线”设计方案，目前还处于不成熟的阶段，有待进一步研究。

从起先的离线训练到如今的实时在线训练，学习跟踪的方法在不断地进步。与离线训练相比，在线训练在采样、训练和更新等操作上都要做一定改进。通过学者们不断努力，深度学习在多目标跟踪领域已经取得了一些成果，但也仅仅是小范围的应用，可探索空间仍然巨大。

6 跟踪器

6.1 运动模型

通过外观模型产生了有效的模板信息，而运动模型是在目标被跟踪时，预测出目标下一帧的位置。产生的预测位置是多选的，这就要求寻找一种最优的方法去进行选取分配，经典的分配算法有匈牙利算法(Hungarian Algorithm)。在选择最优的位置时，需要关联上一帧甚至好几帧间的信息，选择效率高的关联算法将会提高跟踪的效率。

在运动模型中常用到马尔可夫运动模型，我们假设一个目标的状态表示为 $Z_t=(x_t,y_t,u_t,v_t)$，其中 (x_t,y_t) 为目标的位置坐标，(u_t,v_t) 为速度的水平和垂直分量，则有：

$$(x_t,y_t)=(x_{t-1},y_{t-1})+(u_s,v_s)\cdot\triangle t+\varepsilon(x,y) \tag{1}$$

$$(u_t,v_t)=(u_s,v_s)+\varepsilon(u,v) \tag{2}$$

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

公式(1)、(2)中 (u_s, v_s) 为目标的瞬时速度,$\varepsilon(x, y)$、$\varepsilon(u, v)$ 分别为目标位置和速度的高斯噪声,Δt 为视频的帧速率。通过马尔可夫运动模型,即使当目标发生快速变化,也可以很容易捕捉到目标。目前运动模型主要分为四种:

1)均值漂移(Meanshift)

均值漂移是一种基于密度梯度上升的非参数方法,通过迭代运算找到目标位置,实现目标跟踪。利用均值漂移算法进行物体跟踪的实质是求解最优化的 Bhattacharyya 系数函数。在跟踪中,为了寻找到相似度值最大的候选目标位置,Meanshift 方法沿着概率密度的梯度方向进行迭代移动,最终达到密度分布最值位置。因此该算法实时性好,计算量小、运行速度快。

2)滑动窗口(Slide Window)

滑动窗口是最基本的运动搜索模型,一般搜索框会设置成方形或圆形,通过不同的尺寸来扫描整个图像,它是一种穷举的搜索采样方式。很显然只要目标所在的区域附近都有可能被搜索,所以计算量就大。

3)卡尔曼滤波(Kalman Filter)

卡尔曼滤波是目标跟踪算法中经典算法,能够在目标跟踪的过程中预测目标的位置和速度,并且效率很高。但是这种预测是有偏差、有噪声的。经典的卡尔曼滤波算法是线性且符合高斯分布的,通过学者们不断研究也出现了非线性的,如扩展卡尔曼滤波器(Extended Kalman Filter,EKF)和无损卡尔曼滤波(Unscented Kalman Filter,UKF)。文献[19]中根据 EKF 和 UKF 的优缺点,分别介绍了两种算法在目标跟踪问题中的应用。

4)粒子滤波(Particle Filter)

粒子滤波是由经典的卡尔曼滤波演变而来的。针对卡尔曼滤波的不足,粒子滤波在非线性和非高斯的条件下表现出很好的优越性。具体算法是从带噪声的数据中估计运动状态,在状态空间中通过随机传播大量带权离散变量,来近似概率分布并递归,每次递归通过比较每个粒子不同的权值进行重采样,直到完成整个目标样本采集。粒子滤波算法效率较高,因此在目标跟踪研究中应用很广泛。

6.2　混合跟踪算法

上一节介绍了几种基于运动模型建模的跟踪算法,虽然不同的跟踪算法之间优缺点不一样,但仍然存在着互补性。因此,目前将几种算法结合起来的混合跟踪算法成为研究热点之一。文献[20]提出了基于 ACF(Aggregate Channel Features)和粒子滤波的多目标跟踪算法,首先采用两种特定的 ACF 检测器对视频序列中目标进行检测,再用 Adaboost 分类器进行训练分类,然后通过粒子滤波器对目标进行追踪,达到对目标的实时监控。

不同领域、不同专业技术也存在知识互补性,有学者将其他专业技术运用到多目标跟踪研究中,使得多目标跟踪方法呈现出多种多样的趋势。文献[21]将图像分割技术运用到多目标跟踪,通过改进的 CRF(Conditional Random Field)模型,首先将图像进行分割,再用差样本去评估目标被遮挡后重现的轨迹,完成轨迹碎片的拼接。实验显示采用多目标分割后,能够提高 10%平均召回率,同时减少 ID 开关的数量。文献[22]提出了一个添加身份信息的多目标跟踪解决方案,通过给检测到的目标添加身份信息,避免了目标之间因相似性引起的错误,同时还可以自动确定潜在的追踪错误,进一步提高了跟踪的精度。

混合跟踪算法极大地解决了各种算法之间因差异性所造成的问题,但不论是基于检测跟踪还是学习跟踪的方法,都需要进行数据关联,数据关联是连接跟踪前后的关键。

7　数据关联

数据关联是多目标跟踪过程中的一个重要阶段,国内外已经有很多学者将多目标跟踪问题看成数据关联问题,从数据关联过程中去寻求多目标跟踪研究方法。数据关联的目的是,计算在当前帧从检测器中检测到的每一个观测值与前一帧跟踪器中可能的各种跟踪目标之间的关联概率,通过概率的匹配度去关联前后帧之间的信息,从而形成一段连续的轨迹。对数据关联的研究有多个角度,常用的有基于能量最小化和基于概率两种方法。

7.1　基于能量最小化的数据关联方法

在目标跟踪中,由于场景的复杂性,使得目标跟踪难以实现,因此可以将多目标跟踪问题转化为能量最小化问题。所谓的能量最小化就是在解空间中,每个解都对应一个损失,整个求解过程要做的是把这个损失函数表示出来,并找到一个合适的方法求最优解。在做求解的过程时,往往已知所有的检测反馈信息以及这些反馈信息所组合的所有可能轨迹,每一个组合都有一个损失,求解的结果就是要寻求最优的组合。

文献[23]就是一个典型的基于能量最小化来进行多目标跟踪,文中清晰地阐述了损失函数的构

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

成以及最小化，通过构造连续的损失函数，跳出局部最优来寻找全局最优，从而求得最优解。文献[24]在之前文献[23]的基础上进行了改进，提出了离散连续能量函数，函数中包含单个目标的数据关联和轨迹估计信息，而轨迹的信息是通过全局的标签损失获得，标签损失描述了各个轨道的物理特性，最后通过基于梯度的连续能量函数来更新各个轨迹的形状。

基于能量最小化的数据关联方法，需要扎实的数学功底，因其计算量大，容易出错，在实践应用中不是很广泛。

7.2 基于概率的数据关联方法

基于概率的数据关联算法也有很多，常用的有最近邻数据关联算法（Near Neighbor Data Associa-tion，NNDA）、多假设跟踪（Multiple Hypothesis Tracking，MHT）以及概念数据关联（Probability Data Association，PDA）和它的升级版联合概率数据关联算法（Joint Probabilistic Data Association，JPDA）。上面列举的方法，都是经典的算法。通过学者们不断的创新，已经有很多改进后的算法。文献[25]改进了之前的 JPDA 算法，使关联信息处理时间减少，同时在应用场景混乱的情形下效果显著。文献[26]针对以往算法通常会忽略目标大小导致跟踪降低的情况，提出了改进的 JPDA 算法，降低了传统 JPDA 算法的复杂度，解决了由目标大小而引发的遮挡问题。

基于最近邻域法的数据关联算法（NNDA）虽然方法简单、计算量小，但抗干扰能力弱；而基于 PDA 和 JPDA 的数据关联算法虽然能很好地适应杂波和其他复杂环境，但是需要计算所有可能的量测概率，导致计算量大大增加；基于 MHT 的数据关联算法很好地综合了 NNDA 算法和 JPDA 算法的优点，但算法本身很依赖于目标和杂波的先验知识。针对算法各自的优缺点，已经有学者将几种方法混合起来使用，衍生出新的数据关联算法。

采用单一的概率数据关联方法的时代即将过去，目前主要的研究方向倾向于结合了多特征技术、多运动模型、多传感器等综合方法来跟踪目标。同时，随着深度学习方法的进一步发展，数据平台的逐步完善，将会有更多基于神经网络结构的数据关联方法出现。深度学习正引领着一股科技创新的浪潮。

8 视频目标跟踪数据集

8.1 通用数据集

一个好的、权威的数据集对实验结果的可信度至关重要。随着深度学习的发展，在图像识别和目标检测领域出现了很多大型的数据集平台，如：ImageNet、Pascal VOC、COCO 等。同时，这些数据集也间接性推动了目标跟踪数据集的发展。其中最有影响力的数据集是：VTB 数据集和 VOT 数据集。目前在多目标跟踪领域比较大的一个数据集是 MOT Challenge[27]，在这个数据集中有很多公开视频序列，常用的视频序列如下：PETS 2009、Town-Centre、PETS 2016 Challenge。PETS 2009 视频序列中，行人会频繁的被其他目标或者障碍物（如交通标志）遮挡，因此行人在前进的过程中会突然改变运动方向，这个视频序列可以用来跟踪目标出现轨迹漂移的情形。Town-Centre 视频序列中，路上的行人会更加的密集，被跟踪的目标会长时间被其他目标或者障碍物（如长凳）遮挡，这个视频序列可用于跟踪目标出现闭塞的情形。PETS 2016 Challenge 视频序列中，相对于前两个视频序列，拥有更多的数据，同时视频序列中对行人的姿势改变、尺度变化以及闭塞等处理方法的挑战性更大，对学者所提出的算法综合性要求要更高。

针对各种复杂的环境，多目标视频跟踪的数据集每年都在更新中，为了适应现在热门的基于深度学习的研究方法，数据集的容量也在不断扩大，数据内容也更加丰富，出现很多新数据集，具体的信息内容可以参考文献[28]。

8.2 评测方法

好的算法要具有：高准确性、高鲁棒性以及高效性。目前对单目标跟踪的评测指标已经有较为明确的判断标准，但对于多目标跟踪一直以来都未统一。CLEAR MOT 指标[29]是目前公认度较大的评价方法，分别从以下两个方面来评估算法的优劣：多目标跟踪的精确度（Multiple Object Tracking Precision，MOTP），该指标是评测算法在确定目标位置上的精确度；多目标跟踪的准确度（Multiple Object Tracking Accuracy，MOTA），该指标是评测算法在确定目标的个数以及目标其他相关属性的准确度。虽然两种指标的评测点不一样，但两者共同评测算法对目标进行连续跟踪的能力。除了比较上面两个指标外，通常还会比较召回

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

率(Recall)、跟踪成功率(Mostly Tracked,MT)、丢失率(Mostly Lost,ML)、漏检率(Missed Detections,MS)、误检率(False Positives,FP)、每秒处理的帧数(Frames Per Second,FPS)以及所用到的 ID 开关总数(Id Switches,IDS)等。

9　发展趋势展望

视频多目标跟踪技术结合了计算机科学、模式识别、图像处理等多个学科的知识,是一个复杂且多变的课题,目前还处于发展阶段,还有很多问题有待解决。经过学者们的研究,虽然有些简单的场景已经能够很好地处理,但是面对更复杂的环境,效果仍然不够理想。在大数据的时代下,基于深度学习的研究方法给多目标跟踪研究添加了新的动力,为了寻找更好的、更稳定、使用场景更多的多目标跟踪算法,仍需要付出大量的努力,目前的研究重点和发展趋势主要集中于以下几点:

1)研究重点的转移。基于检测跟踪的研究方法虽然在很多方面已经取得很好的鲁棒性,但未来的研究重点将会侧重于目标长时间被遮挡的情形,同时对目标再现后重新被检测跟踪也是后面研究的重点。

2)深度学习与在线学习融合。视频跟踪的本质就是一个在线学习的问题,一味地在线学习可能导致进入死胡同,如何压缩网络结构,缩短训练时间,提高样本质量,避免学习过程中数据冗余,都是值得深入研究的问题。

3)多传感器的应用。多传感器的使用可以扩宽监督的范围,同时也可以全方位的观察目标,不同的视角效果有助于处理目标被遮挡的问题。另外,目前的跟踪方法大多是 2D 建模,如何将 3D 建模运用到目标跟踪上,实现目标全息投影跟踪也是后面研究的重点。

4)数据平台的创建。随着大数据时代的到来,给多目标跟踪带来了海量的数据信息,因此出现了很多基于深度学习的训练数据库。但是,目前大多数目标跟踪算法采用的是目标识别的数据集,如何针对目标跟踪的特性来构建专属性的数据平台也是今后研究的重点。

5)研究环境的转移。随着计算机芯片 GPU 的发展,越来越多的实验环境已经移植到 GPU 上去进行,同时应用深度学习的方法也需要学者采用更快的、更高的处理器。如何将实验环境快速移植到 GPU 上而尽可能少改动代码,也是值得研究的问题。

6)与生物仿生学相结合。目前,谷歌研究的智能机器人已经在很多方面战胜了人类。而人眼是最容易去观察跟踪目标的,同时通过人眼与大脑神经的处理,可以忽视任何的复杂场景。所以未来是否可以考虑模拟人眼的观察追踪能力,通过生物仿生技术,让机器具备人类的视觉思维,去追踪目标呢?如何实现这种技术,也是值得研究的方向。

10　结束语

本文分析总结了近年来多目标视频跟踪的研究方法,给读者提供了丰富的多目标跟踪相关知识。通过描述多目标跟踪的整个过程,分别介绍了特征提取、检测器、数据关联和跟踪器几个重要的阶段。在检测器和跟踪器中,分别对目标的外观模型和运动模型进行了详细地叙述,同时也列举出了一些目前热门的多目标跟踪算法,着重介绍了基于深度学习框架的算法以及各种数据关联的算法。另外也介绍了多目标跟踪相关的数据集和评测方法。文章最后分析了大数据和深度学习给多目标跟踪带来的新机遇以及在新环境下多目标跟踪的未来发展趋势。我相信,通过国内外学者不断研究、大胆尝试,多目标跟踪的研究将拥有更广阔的前景。

参考文献

[1] WU Y, LIM J, YANG M H. Online object tracking: A benchm-ark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2411－2418.

[2] 黄凯奇, 陈晓棠, 康运锋,等. 智能视频监控技术综述[J]. 计算机学报, 2015, 20(6):1093－1118.

[3] DANELLJAN M, SHAHBAZ Khan F, FELSBERG M, et al. Adaptive color attributes for real-time visualtracking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1090－1097.

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarc-hies for accurate object detection and semantic segmentati-on[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580－587.

[5] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE interna-tional conference on computer vision. 2015: 1440－1448.

[6] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelli-

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

gence, 2017, 39(6): 1137－1149.

[7] LI W, YAO J, DONG T, et al. Object Tracking Based on Fragment Template and Multi-Feature Adaptive Fusion [C]//Computational Intelligence and Design (ISCID), 2015 8th International Symposium on. IEEE, 2015, 2: 481－484.

[8] ZHANG J, PRESTI L L, SCLAROFF S. Online multi-person tracking by tracker hierarchy[C]//Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. IEEE, 2012: 379－385.

[9] GUAN Y, CHEN X, YANG D, et al. Multi-person tracking-by-detection with local particle filtering and global occlusion handling[C]//Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE, 2014: 1－6.

[10] GAN M, CHENG Y, WANG Y, et al. Hierarchical particle filter tracking algorithm based on multi-feature fusion[J]. Journal of Systems Engineering and Electronics, 2016, 27(1): 51－62.

[11] JU J, KIM D, KU B, et al. Online Multi-object Tracking Based on Hierarchical Association Framework[C]//Proceed-ings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 34－42.

[12] **金哲, 刘传才. 加速的 TLD 算法及其在多目标跟踪中的应用[J]. 计算机系统应用**, 2016, 25(6):196－201.

[13] SHARMA S, KHACHANE A, MOTWANI D. Real time multi-object tracking using TLD framework [C]// Inventive Computation Technologies (ICICT), International Conference on. IEEE, 2016, 2: 1－6.

[14] LU Y, WU T, CHUN S. Online object tracking, learning and parsing with and-or graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3462－3469.

[15] LEAL-TAIXÉ L, CANTON-Ferrer C, SCHINDLER K. Learning by tracking: Siamese CNN for robust target association[C]//Pr-oceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 33－40.

[16] YU F, LI W, LI Q, et al. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature [J]. 2016:36－42.

[17] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking [C]//Proc-eedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4303－4311.

[18] MILAN A, REZATOFIGHI S H, DICK A R, et al. Online Multi-Target Tracking Using Recurrent Neural Networks [C]//AAAI. 2017: 4225－4232.

[19] YANG S, LI H. Application of EKF and UKF in Target Tracking Problem[C]//Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2016 8th International Conference on. IEEE, 2016, 1: 116－120.

[20] KOKUL T, RAMANAN A, PINIDIYAARACHCHI U A J. Online multi-person tracking-by-detection method using ACF and particle filter[C]//Intelligent Computing and Information Systems (ICICIS), 2015 IEEE Seventh International Conference on. IEEE, 2015: 529－536.

[21] MILAN A, LEAL-TAIXÉ L, SCHINDLER K, et al. Joint tracking and segmentation of multiple targets[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5397－5406.

[22] YU S I, MENG D, ZUO W, et al. The solution path algorithm for identity-aware multi-object tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3871－3879.

[23] MILAN A, ROTH S, SCHINDLER K. Continuous energy minimization for multitarget tracking [J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(1): 58－72.

[24] MILAN A, SCHINDLER K, ROTH S. Multi-target tracking by discrete-continuous energy minimization[J]. IEEE transact-ions on pattern analysis and machine intelligence, 2016, 38(10): 2054－2068.

[25] REZATOFIGHI S H, MILAN A, ZHANG Z, et al. Joint Probabilistic Data Association Revisited[C]// IEEE International Confer-ence on Computer Vision. IEEE, 2016: 3047－3055.

[26] SHI X, SONG Y Q, YANG Z, et al. Multiple target tracking under occlusions using modified Joint Probabilistic Data Association [C]//Communications (ICC), 2015 IEEE International Conference on. IEEE, 2015: 6615－6620.

[27] LEAL-TAIXÉ L, MILAN A, REID I, et al. Motchallenge 2015: Towards a benchmark for multi-target tracking[J]. arXiv preprint arXiv:1504.01942, 2015.

[28] MILAN A, LEAL-TAIXÉ L, REID I, et al. MOT16: A benchmark for multi-object tracking[J]. arXiv preprint arXiv:1603.00831, 2016.https://motchallenge.net/.

[29] KENI B, RAINER S. Evaluating multiple object tracking performance: the CLEAR MOT metrics[J]. Eurasip Journal on Image & Video Processing, 2008, 2008 (1):246309.

?1994-2018 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net