

目标跟踪综述

王海涛¹, 王荣耀¹, 王文峰¹, 王海龙^{1,2}, 刘强^{1,3}

(1. 南京航空航天大学自动化学院, 南京 211106;

2. 山东长城计算机系统有限公司, 山东烟台 264003;

3. 江苏铭远轨道交通设备股份有限公司, 南京 210044)

摘要: 随着深度学习与人工智能技术的不断发展, 视频目标跟踪已经成为计算机视觉的重要研究内容, 在公安布控、人机交互、交通管制、军事等各个领域起到越来越重要的作用; 尽管现在国内外学者提出了多种目标跟踪算法, 也搭建了较为完善的目标跟踪系统, 但是算法的鲁棒性依然是一个比较大的挑战; 文章对运动目标跟踪系统结构进行了简要介绍, 并从特征提取及融合、外观模型、目标搜索等方面详细阐述了目前主流运动目标跟踪算法; 然后对目标跟踪算法在深度学习大环境下的新发展进行了分析, 从基于深度学习的目标跟踪及目标检测算法角度分析了深度学习在提高目标检测算法鲁棒性方面的有效性, 最后概述了深度学习在视频目标检测算法中的具体应用并对其未来发展进行了展望。

关键词: 目标跟踪; 特征提取; 外观模型; 深度学习; 神经网络

A Survey on Recent Advance and Trends in Object Tracking

Wang Haitao¹, Wang Rongyao¹, Wang Wenhao¹, Wang Hailong^{1,2}, Liu Qiang^{1,3}

(1. School of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. Shandong Great Wall Computer System Co., Ltd., Yantai 264003, China;

3. Jiangsu Mingyuan Rail Transit Co., Ltd., Nanjing 210044, China)

Abstract: With the development of deep learning and artificial intelligence technology, video object tracking has become an important research content of computer vision. Video object tracking plays a more and more important role in public security, human-computer interaction, traffic control, military and other fields. Although a variety of object tracking algorithms have been proposed by scholars, and a relatively perfect object tracking system has also been built, the robustness of the algorithm is still a big challenge. In this paper, the structure of moving object tracking system is briefly introduced. At the same time, the main moving object tracking algorithms are described in detail from feature extraction and fusion, appearance model, object search and so on. Then, a new development of object tracking algorithm in deep learning environment is analyzed. From the perspective of object tracking and object detection algorithm based on deep learning, the effectiveness of deep learning in improving the robustness of object detection algorithm is analyzed. Finally, the specific application of video object detection algorithm is summarized and its future development is prospected.

Keywords: object tracking; feature extraction; appearance model; deep learning; neural network

0 引言

视频目标跟踪是机器视觉领域的重要分支, 在人机交互、监控安防、无人驾驶及军事战争中起到了关键作用, 其主要功能是在获取的视频图像中遍历寻找感兴趣区域, 并在接下来的视频帧中对其进行跟踪^[1-2]。然而, 目标本身姿态的变化产生的形变和位置变化、摄像头被遮挡、背景差异较大及外部光照天气原因的影响, 对目标进行自适应的跟踪仍然具有比较大的困难^[3]。因此, 如何找到一种行之有效的方法, 使其能够具有足够的鲁棒性处理以上可能存在的各种复杂情况成为了当下亟待解决的问题。

依据跟踪目标数目的不同, 目标跟踪可分为单目标跟

踪和多目标跟踪, 鉴于目前的主流目标跟踪算法主要由外观模型和目标搜索两个部分组成。目标搜索主要是对视频帧中每一帧图像的信息来帮助目标跟踪; 外观模型主要是充分利用目标特征, 使跟踪目标尽可能和背景区分出来。本文将从目标的目标搜索和外观模型两个方面来介绍目标跟踪算法。

随着深度学习在图像分类和目标检测等计算机视觉领域中的成功应用, 深度学习也开始大量应用于目标跟踪算法中, 从深度学习中不断兴起的新技术如相关性滤波、卷积神经网络等也很快融入该领域, 并显著提高了领域的平均水平。随着深度学习技术的大量涌现, 计算机视觉领域又取得了飞速的发展, 基于深度学习的跟踪算法更是有着十分广泛的应用场景, 如: 智能视频监控控制、异常行为识别分析、人机交互和医疗图像、公共场所行人流量分析等。

深度学习的迅速兴起使目标跟踪算法又迎来了新的发展方向, 本文也将对基于深度学习的目标跟踪算法的进展进行介绍, 分析其应用场景并且对未来的发展进行展望。

收稿日期: 2020-02-19; 修回日期: 2020-03-05。

作者简介: 王海涛(1968-), 男, 江苏溧阳人, 教授, 主要从事计算机测控技术、图像处理技术、无损检测技术方向的研究。

通讯作者: 王荣耀(1995-), 男, 江苏南京人, 硕士, 主要从事图像处理与深度学习方向的研究。

1 运动目标跟踪系统结构

主流视频目标跟踪系统结构如图 1 所示,从整体上分为初始化、特征提取、外观模型、目标搜索 4 个部分。其中外观模型和目标搜索是整个系统的核心部分^[4]。



图 1 运动目标跟踪系统结构

跟踪算法的效果很大程度上取决于目标的外观模型的建立,建立合适的外观模型能够有效提高跟踪算法的跟踪性能。目标搜索用于对目标下一帧的运动状态进行预测,通过预测目标可能出现的位置对目标进行定位,减少目标搜索范围。

2 特征提取及融合

根据目前跟踪算法中所用到的外观特征,可以将外观特征大体分为边缘特征、灰度特征、颜色特征、纹理特征和梯度特征。

边缘特征是一种比较常用的特征,利用边缘检测的方法来提取目标的外观轮廓,边缘检测的基本思想是寻找图像中亮度发生剧烈变化的点的集合,而这个集合往往表现出来的就是目标的外观轮廓,边缘检测更像是特征提取的基本方式,在灰度、颜色、纹理、梯度这些特征的提取中或多或少都用到了边缘特征提取的基本方法。边缘检测已经有了很成熟的算法,如 Canny^[5]、Rothwell^[6]、Nalwa^[7]、Bergholm^[8]等。这些算法中,Canny 算子是最常用的一种算法也是最稳定效果较好的算法。

灰度特征由于其转化简单且容易计算,是特征描述中常用的表达方式,其中灰度值和区域灰度变化特征(Haar 特征)这两种灰度特征表征形式最为常用。采取灰度值作为目标外观模型的表达方式最为简单和直接,但是其对外观表征能力较弱以及容易受到环境变量干扰;Haar 特征反映了图像区域的灰度变化情况,最早由 Papageorgiou 等提出^[9],后来 Paul Viola 和 Michal Jones 最先将 Haar 特征用于检测图像中特定目标(比如人脸)^[10]。

颜色特征也是常用的描述特征,利用颜色直方图^[11-12]能够通过计算出图像中不同色彩所占的比例来大致描述外观,是一种全局特征。颜色特征比较难以准确地描述目标姿态的变化,本身的外观表征能力比较弱,因此,颜色特征通常会和其他特征来共同表征目标外观,来完善对目标的跟踪。

纹理特征通常用来描述目标或目标周围环境的表面变化也是一种全局特征。由于纹理特征只能描述目标的表面性质的局限性,因此仅仅用纹理特征无法描述目标的本质特征,无法获得目标的更深层次的图像。常用 LBP (local binary pattern)^[13]来描述目标的纹理特征,通对比于目标像素与其邻域像素对比得到目标像素的 LBP 值,目标像素的 LBP 值反映了该像素周围区域的纹理信息。

梯度特征通过描述目标图像的梯度分布来描述目标,梯度特征的最为原始的描述算子是 SIFT 算子^[14],在许多目标跟踪算法中都用 SIFT 算子来描述目标外观,但是 SIFT 算子的计算过程较为复杂且耗时,对于算法的实时性有一定的影响。因此,在 SIFT 的基础上 SURF 算子^[15],来减少计算时间,提高算法的实时性。另外在目标跟踪中使用更为广泛的是区域算子(HOG 算子)^[16],该算子对目标局部区域的梯度方向进行统计,能够很好地表征目标局部像素之间的关联。文献[17]中利用 SURF 算子来实现了对行人的跟踪,而在文献[18]中利用 HOG 特征进行外观建模又进一步提高了跟踪的实时性,但是梯度特征缺点在于无法准确描述外观的姿态、外观等信息。

在目标跟踪中目标有多种特征,每种特征都有其优点与缺陷,采用单一特征的方式往往不能很好地表达目标的外观变化,因此,要提高目标跟踪算法的跟踪性能与算法的鲁棒性,需要将多种特征进行融合,从而提高目标外观模型与背景的区别能力。目前,很多跟踪的算法都用到了特征融合,文献[19]中将颜色和纹理特征进行了直接融合达到了理想的效果,文献[20]将颜色和边缘特征进行了融合,文献[21]将颜色、纹理特征和目标运动特征进行融合使算法能够进行多目标的跟踪。

3 外观模型

近年来,跟踪算法中的外观模型的发展主要经历了以下两个阶段:在第一阶段,跟踪算法主要是在周围环境条件提前设定以及目标运动稳定的条件下被提出^[22-23],这些算法对于目标的短时跟踪和周围环境状况良好的条件下有不错的效果;第二阶段,目标跟踪算法将重心主要集中到设计出能够对周围环境以及目标运动状态改变进行自适应的外观模型,由此很多基于自适应外观模型的算法被提出。由于目标外观模型会在光照变化、运动姿态变化、背景遮挡影响下发生变化,因此需要对外观模型进行不断的学习训练来提高算法的自适应能力,外观模型的构建分为两种方法:生成式模型^[24-27]和判别式模型^[28-30]。

3.1 产生式外观模型

产生式模型首先对需要跟踪的目标进行学习,得到目标的外观模型,其后在需要跟踪的图像区域中选择和目标模型最为接近的区域作为目标区域。考虑到目标外观的不断改变,目标的外观模型需要进行不断跟新以保证跟踪的准确性。产生式模型总体上可以分为基于模板的模型、基于子空间的模型两类。

3.1.1 基于模板的学习模型

目前,基于模板的学习方法分为固定特征模板以及自适应特征模板两种。固定特征模板选取固定的特征分量来进行混合得到跟踪目标的模型,并在目标运动跟踪过程中对特征分量的参数进行调节来适应目标外观姿态的改变,以便更好地对目标进行跟踪,文献[31]中通过不同外观特征建立不同模板,通过特定的模板效果评价机制来选择跟踪效果最好的模板进行跟踪,并且不断跟新选择的模板

来达到最优的模板选择。文献 [32] 中提出了高斯混合模型来对目标进行建模,能够对目标的形态外观变化进行比较准确的逼近,但是固定特征模板的方法缺点在于无法提前改变参数来适应目标的外观变化,并且固定的特征参数很难完全描述目标的所有的外观变化情况。

针对固定特征模板的不足之处自适应特征分量的混合模型被提出。自适应高斯混合模型的特征分量能够根据目标外观变化来调整分量数量达到最好的逼近目标外观变化的效果。最具代表性的高斯混合模型^[32],该文章中提出的空间混合高斯模型(SMOG模型),将目标空间上的变化和时间上的变化相结合,得到了很好的外观模型。混合模型相对于固定特征模型在学习能力上有很好的提升,但是对于如何选择合适的特征分量去更好的逼近目标的外观还是需要更进一步的研究。

3.1.2 基于子空间的学习模型

基于子空间的模型实质是将目标映射到低维的子空间中,该子空间尽可能地包含了目标的外观模型数据,通过对子空间中对外观模型进行训练并不断更新,实现对目标的自适应跟踪。基于子空间的学习模型主要分为基于向量的子空间、基于非线性子空间以及基于稀疏子空间。

基于向量的子空间模型主要是将目标通过向量的形式映射到子空间中来建立目标的外观模型,通过对子空间的模型的学习来实现对目标外观的自适应性。文献 [33] 中首先运用向量子空间的方法来对目标进行建模以及跟踪,在此基础上 Ross 等人提出了增量子空间的学习方法^[34],减少了算法的计算量,并且实现了对目标的自适应跟踪。

基于向量的子空间学习方法是通过对线性映射把目标图像映射到子空间中,但是该方法的前提是目标的外观变化被默认为的线性变化,而现实中目标的外观变化往往是非线性的,因此研究者们提出了基于非线性子空间的学习模型。代表性的算法是 LLE^[35],该算法采用非线性降维方法将目标图像映射到非线性子空间中,使跟踪算法能够适应复杂的跟踪环境以及对目标突外观变化有很好的跟踪。

近年来,稀疏表示逐渐成为机器视觉中重要的建模方法。X. Mei 等人首先将稀疏表示的方式运用到运动目标跟踪中^[36],稀疏表示归根结底归结是求约束条件的 L1 范数最小化问题,该方法虽然是计算量很大,影响算法的实时性,但是对于目标被遮挡和目标姿态变化情况下有良好的跟踪效果。

3.2 判别式外观模型

由于机器学习在目标检测领域取得了巨大的成功,判别式模型通过借鉴机器学习的方法将跟踪目标和背景分成两类,这样目标跟踪就被转化成了一个分类的问题。判别式模型优点在于对跟踪目标外观有良好的区分能力,因此,在目前的目标跟踪中有着广泛的运用。

3.2.1 基于 SVM 的外观模型

基于 SVM 的跟踪是一种常用的跟踪方法,其原理是通过将跟踪目标设定为正样本,将背景目标设定为负样本,

通过 SVM 方法对正负样本进行训练得到训练好的模型,并结合其他的自适应的学习方法来实时更新目标外观模型。文献 [37] 将 SVM 引入目标跟踪中,并取得了良好的判别式模型。然而基于 SVM 的方法只是对单一的目标进行训练,这就需要准确选择训练的目标,一旦训练目标选择错误将会影响跟踪的效果。

文献 [38-39] 提出将需要训练的样本通过图像的形式进行分类训练,这使得训练出来的外观模型能够不叫好的适应环境的变化,提高了模型的质量。文献 [40] 提出将半监督学习方法与 SVM 算法相结合,使得目标外观模型在训练中受到一定程度的监督,能够保持分类的效果。

3.2.2 基于 Boosting 的外观模型

Boosting 方法能够将一堆弱分类器提升为一个强分类器, Kearns & Valiant 证明了弱的学习算法和强学习算法之间的等价关系^[41],只需要找到一个弱学习算法就能将其提升为强学习算法。Boosting 通过不断迭代的方法来优化分类器分类效果,每次迭代一个弱分类器,来克服当前存在的弱分类器组合的缺点,从而逐渐得到一个强的分类器。在 Boosting 的 Adaboost 算法^[42]中,分类器的缺点表征就是权重高的样本点。Gradient Boosting 算法^[43]中,分类器的弱点表征就是梯度。

文献 [44] 提出的 TLD 算法是对视频中位置物体的长时间跟踪的算法,该算法将跟踪器、检测器和机器学习相结合,利用学习模块对跟踪器和检测器产生的正负样本进行训练,完善分类器的效果,能够适应目标丢失、目标遮挡以及目标运行轨迹的漂移等情况。文献 [45] 提出的 MIL 算法在训练过程中将样本包来代替单个样本,同时样本包也有一个标签,若样本包中至少有一个正样本,那么该样本包就为正标签,反之为负标签,这些样本包就构成了训练的样本集合,通过这种方法能够提高标签的质量,能够很好地适应目标的姿态变化、目标运动过程中的遮挡、目标运动轨迹漂移等情况。TLD 算法和 MIL 算法都是基于本监督学习的跟踪算法,都提出了标签样本优化的思想,构成了半监督学习算法的基本框架。

3.3 模型更新

在对目标运动跟踪的过程中目标的运动状态会发生改变,目标的运动姿态也会随之发生变化,而目标的外观模型是在知道目标之前的样本信息的情况下建立的,无法反应目标当前的外观姿态,因此外观模型需要根据目标姿态的改变适当的进行更新。

4 目标搜索算法

建立好目标外观模型后需要对运动目标进行搜索,目标跟踪过程中,若对一帧图像中所有的目标进行检测与匹配将会增加计算量,降低跟踪算法的实时性^[4]。通常每一帧图像之间运动目标的运动距离不会间隔很远,为了减少计算量提高算法的实时性,需要采取一定的目标搜索算法来预测运动目标在下一帧图像中可能出现的位置,通常目标可能出现的位置在上一帧目标位置的周围,这样通过在

预测位置进行目标的检测和对检测到的目标进行匹配减少了搜索范围提高了计算速度。这种通过预测目标可能出现的位置的常用预测算法有卡尔曼滤波以及粒子滤波方法。

卡尔曼预测是对跟踪目标的运动轨迹进行最小方差估计从而对运动目标下一帧位置进行预测的算法^[46]。通过建立运动目标的观测方程和运动目标实际的运动方程,将两个方程进行卡尔曼预测,得到运动目标下一帧位置的估计,卡尔曼预测算法优点在于其计算量小以及不错的算法实时性,但是该算法只适用于符合高斯分布的运动系统。粒子滤波器则适用于非线性、非高斯系统,它是由卡尔曼滤波的基础上发展而来^[47]。粒子滤波通过在跟踪目标上一帧的位置附近进行粒子采样,得到每个粒子位置图像与目标的相似性分布,再对粒子位置进行重要性采样(相似程度高的位置多采样),这样就形成了对目标的实时跟踪,粒子滤波计算量较少,因此算法实时性能良好并被广泛应用于跟踪算法中。

均值漂移算法(Meanshift 算法)是另一种减少搜索范围的算法。Meanshift 是一种非参数估计的方法^[48],通过计算目标上一帧模板和当前帧目标位置的候选模板的相似性,选择相似程度最高的模板得到目标的 Meanshift 向量,该向量由目标上一帧位置指向目标当前帧位置,通过不断迭代均值漂移算法计算 Meanshift 向量,由于均值漂移算法的收敛性,最终算法将会定位到当前帧目标的最终位置,实现对目标的跟踪。

5 目标跟踪算法在深度学习大环境下的发展

随着卷积神经网络在图像领域的广泛应用以及 GPU 的迅猛发展,越来越多的注意力转移到了深度学习方面,时代的发展以及深度学习在模式识别、目标检测领域取得了巨大的成功将目标跟踪推向了一个新的高度。然而深度学习在目标跟踪领域中的运用不同于在模式识别和目标检测中那么顺风顺水,其根本原因在于训练样本的缺失,深度学习的运用是在大量样本训练基础上实现的,而在目标跟踪中只能提供第一帧的图像区域作为训练样本,很难针对当前目标训练一个深度模型。

5.1 基于深度学习的跟踪算法

随着深度卷积神经网络(CNN)的兴起,目标跟踪领域也出现了也出现了利用 CNN 来进行目标跟踪。该类方法直接使用 ImageNet 这类大规模的数据库来训练出 CNN 网络来提取跟踪目标的特征,再通过相应的观测模型来进行分类,实现对目标的跟踪^[49]。代表算法是 FCNT^[50],该算法利用在 ImageNet 中训练得到的 CNN 运用于实际目标跟踪后,对跟踪性能进行了进一步的分析并设计了后续神经网络,提高了跟踪算法性能。

目前,一类算法通过对大量的非跟踪辅助数据进行预训练,该辅助数据需要与跟踪目标具有相同的特征,这样经预训练得到了对需要跟踪目标的通用模型,在具体跟踪时只要根据目标样本对预训练的模型进行微调就可以得到对当前目标有很强的分类能力的模型,代表算法为 DLT^[51]

和 SO-DLT^[52],这些算法减小了深度学习对跟踪目标训练样本的依赖,并且提高了跟踪算法性能。文献[53]提出了一种对目标运动序列预训练的方法,通过对目标类在不同运动序列中进行预训练得到目标类的 CNN 网络,算法跟踪能力得到进一步提升。

5.2 目标检测算法在目标跟踪中的借鉴

鉴于深度学习在目标检测中取得的应用,可以把目标检测的算法借鉴到目标跟踪中来。目前基于深度学习的目标检测方法主要分为两类:一种是将检测问题分为两个阶段,显示产生候选区域,然后对选择的候选区域进行分类。这类方法的代表为 RCNN^[54],Fast RCNN^[55],Faster RCNN^[56]等;另一种是不需要产生候选区域,把目标判定和识别合二为一,大大提高了算法速度,代表的算法有 Yolo^[57]和 SSD^[58]。

RCNN 由 Girshick 在文献[54]中首先提出,RCNN 算法奠定了由 Region proposal 和 CNN 神经网络组成的目标检测的框架,在之后的这一类目标检测算法都以其为基本框架。RCNN 首先在检测图片中选候选区域,再对这些区域进行特征提取,将提取的特征在进行 SVM 分类,判断是否是属于一个特定的类别,最后使用相应的回归器对判断完成的候选区域进行位置修正。文献[55]中提出的 Fast RCNN 也是由 Girshick 在其 RCNN 算法上改进后提出的,Fast RCNN 优化了 RCNN 存在的测试速度慢、训练用时慢以及训练占用空间大的缺陷。同样 Faster RCNN 也是由 Girshick 在文献[56]中首先提出,在 Fast RCNN 的基础上用区域生产网络的方法来选取特征区域,减少了计算量,并且把整个目标检测步骤都集成到了一个深度学习的框架中,从而去掉了多余的重复计算量,使整个目标检测过程中没有多余的重复计算过程,极大提高了算法运行速度。

相对于第一类的算法,Yolo 和 SSD 运用回归的方法直接从图片中回归出目标的位置以及目标的种类,降低了准确度但是提高了处理速度。Yolo 算法在文献[57]中首先被提出,Yolo 算法直接在图片中划分网格并利用分类网络进行目标寻找,舍弃了寻找 Region proposal 的过程;SSD 则是将 Yolo 和 Faster RCNN 方法结合起来^[58],在保证检测速度的同时对目标定位精度和准确度较 Yolo 算法都有提高。目前最新的 SSD 算法 SSD300 可以达到每秒 46 帧,YoloV2^[59]可以达到每秒 40 帧,YoloV3 可以达到每秒 45 帧,对于目标跟踪来说,每秒 45 帧的处理速度已经达到了实时的标准。

目前卷积神经网络已经成了图像识别和视频跟踪领域的主要运用方法,随着目标检测速度与准确度的上升,今后的目标跟踪将与目标检测结合越来越密切。

6 目标跟踪的应用领域

由于目标跟踪技术的日渐完善,已越来越多地出现在智能安防监控、人机交互、异常行为识别分析等重要领域。作为计算机视觉领域的重要分支之一,智能安防监控技术进入了快速发展阶段,保障人民群众的生命财产安全和社

会和谐稳定,其中最具代表性的为美国 VSAM (video surveillance and monitoring) 项目和 HID (human identification at a distance) 项目^[60]以及日本的 CDVP (the cooperative distributed vision project) 视觉监控计划^[61],视频目标跟踪技术在这些项目中起到了至关重要的作用。在人机交互领域,2012年 Molyneux 等基于目标跟踪技术发明了一种基于红外摄像头和激光投影仪实现的手持空间可感知的交互系统,此系统可以在空间任意界面实现绘画,也可以对投影出来的物件进行触控操作,甚至可以通过手指的阴影对其进行相同的操作^[62]。在 CVPR2009 上, Kim. J 等提出了时空马尔科夫随机场异常行为检测模型,使用多个 CRF (control and related field) 模型对视频数据的多种特征各自进行建模,再组合形成检测模型,该模型具有较高的检测效果,使得异常行为分析技术向前迈了一步^[63]。

随着深度学习的不断发展,基于深度学习的目标跟踪算法在行人追踪、套牌车识别、无人机应用、无人驾驶及智能交通控制等领域发挥了至关重要的作用。胡超超等人以行人及骑行者为研究对象,在 YoLoV2 中加入残差网络构成 YoLo-R 网络,完成了行人及骑行者的分类,并运用 Kalman 滤波实现了多目标跟踪^[64]。无人驾驶技术的实现是对计算机视觉领域的重要补充,在自动驾驶过程中,可以通过对周围障碍物、信号灯及行人的位置判断及轨迹预测自动规划下一步运行路径,选择合适的路线为,而自动驾驶的核心技术就是目标跟踪,从斯坦福大学人工智能实验室分出来的 Drive.ai 公司利用深度学习算法解决了车辆的自动驾驶问题并获得了美国加利福尼亚州测试无人驾驶汽车的许可^[65]。在无人机应用方面,法国 Moon 等通过采用两个摄像头所拍摄图像的视差得到无人机距离地面高度的信息以及其与周围障碍物距离信息,并利用深度学习相关算法实现了基于双目视觉的无人机自主避障功能^[66]。此外,深度学习在城市交通控制系统中发挥了关键作用,其目的在于利用现有交通网络道路的基础设施,在没有增加大量成本的情况下有效缓解城市路口的交通拥堵, Lee 等人研发了一个基于机器学习预测的交通控制系统,并可以在减少拥塞率的情况下动态确定自主车辆系统路线^[67]。

7 结束语

本文在对视频目标跟踪方法进行了一定资料查阅的基础上,首先介绍了视频目标跟踪系统的大体流程,并对主流算法采用的外观模型和搜索方法进行了详细的介绍。然后结合深度学习在当下的迅速发展,介绍了基于深度学习的目标跟踪算法,并根据当前主流的目标检测算法能达到的处理速度提出了目标检测和跟踪相结合的方法,最后总结了深度学习在目标跟踪算法中的具体应用。在目标跟踪发展的这些年,如何在保证准确率的前提下提高算法的实时性认识是一个重大的挑战。

参考文献:

[1] Yilmaz A. Object tracking: a survey [J]. ACM Computing Sur-

veys, 2006, 38 (4): 1-45.

- [2] Liu M, Wu C D, Zhang Y Z. A review of traffic visual tracking technology [C]. International Conference on Audio, Language and Image Processing, 2008. ICALIP 2008. . IEEE, 2008.
- [3] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: A review [J]. Neurocomputing, 2011, 74 (18): 3823-3831.
- [4] 管 皓, 薛向阳, 安志勇. 在线单目标视频跟踪算法综述 [J]. 小型微型计算机系统, 2017, 38 (1): 147-153.
- [5] Ding L, Goshtasby A. On the canny edge detector [J]. Pattern Recognition, 2001, 34 (3): 721-725.
- [6] Yang Y. An improved edge detection method based on topology [J]. Proceedings of SPIE-The International Society for Optical Engineering, 2009: 74-94.
- [7] Nalwa V S, Binford T O. On detecting edges [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1986, PAMI-8 (6): 699-714.
- [8] Bergholm F. Edge focusing [J]. IEEE Trans Pattern Anal Mach. Intell., 1987, 9 (6): 726-741.
- [9] Papageorgiou C P, Oren M, Poggio T. A general framework for object detection [A]. International Conference on Computer Vision [C]. IEEE, 2002: 555-562.
- [10] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [A]. Computer Vision and Pattern Recognition [C]. 2001.
- [11] Pérez P, Hue C, Vermaak J, et al. Color-Based Probabilistic Tracking [A]. European Conference on Computer Vision [C]. 2002.
- [12] Possegger H, Mauthner T, Bischof H. In defense of color-based model-free tracking [A]. Computer Vision and Pattern Recognition [C]. IEEE, 2015: 2113-2120.
- [13] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions [A]. International Conference on Pattern Recognition [C]. 1994.
- [14] Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift [J]. Computer Vision & Image Understanding, 2009, 113 (3): 345-352.
- [15] Miao Q, Wang G, Shi C, et al. A new framework for on-line object tracking based on SURF [J]. Pattern Recognition Letters, 2011, 32 (13): 1564-1571.
- [16] Dalal N, Triggs B. Histograms of oriented gradients for human detection [A]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition [C]. 2005: 886-893.
- [17] Choi H C, Oh S Y. Robust segment-based object tracking using generalized hyperplane approximation [J]. Pattern Recognition, 2012, 45 (8): 2980-2991.
- [18] Henriques J F, Rui C, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (3): 583-596.
- [19] Wang Q, Chen F, Xu W. Adaptive multi-cue tracking by on-

- line appearance learning [J]. *Neurocomputing*, 2011, 74 (6): 1035–1045.
- [20] Morenonoguer F, Sanfeliu A, Samaras D. Dependent multiple cue integration for robust tracking [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2008, 30 (4): 670–685.
- [21] Liu C, Liu J, Huang J, et al. Human pose tracking based on partitioned sampling particle filter and multiple cues fusion [J]. *Journal of Computer Research & Development*, 2011, 48 (12): 2359–2368.
- [22] Haritaoglu I, Harwood D, David L S. W4: Real-time surveillance of people and their activities [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 22 (8): 809–830.
- [23] Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram [A]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*. 2006: 798–805.
- [24] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2003, 25 (5): 564–575.
- [25] Liang D, Huang Q, Jiang S, et al. Mean-shift blob tracking with adaptive feature selection and scale adaptation [A]. *IEEE International Conference on Image Processing [C]*. 2007: 369–372.
- [26] Yu T, Wu Y. Differential tracking based on spatial-appearance model (SAM) [A]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*. 2006: 720–727.
- [27] Han B, Davis L. On-line density-based appearance modeling for object tracking [A]. *Tenth IEEE International Conference on Computer Vision [C]*. 2005: 1492–1499.
- [28] Lim H, Camps O I, Sznaiar M, et al. Dynamic appearance modeling for human tracking [A]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*. 2006: 751–757.
- [29] Tian M, Zhang W, Liu F. On-line ensemble SVM for robust object tracking [A]. *Asian Conference on Computer Vision [C]*. 2007: 355–364.
- [30] Hare S, Saffari A, Torr P H S. Struck: structured output tracking with kernels [A]. *IEEE International Conference on Computer Vision [C]*. IEEE, 2012: 263–270.
- [31] Li X, Hu W, Zhang Z, et al. Robust Visual Tracking Based on an Effective Appearance Model [C]. *Computer Vision – EC-CV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part IV*. Springer-Verlag, 2008.
- [32] Wang H, Suter D, Schindler K, et al. Adaptive object tracking based on an effective appearance filter [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence [C]*. 2007, 29 (9): 1661.
- [33] Ho J, Lee K C, Yang M H, et al. Visual tracking using learned linear subspaces [A]. *Computer Vision and Pattern Recognition [C]*. 2004: 1–782–1–789.
- [34] Ross D, Lim J, Yang M H. Adaptive Probabilistic visual tracking with Incremental Subspace Update [A]. *European Conference on Computer Vision [C]*. 2004: 470–482.
- [35] Kwon J, Lee K M. Highly nonrigid object tracking via patch-based dynamic appearance modeling [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35 (10): 2427–2441.
- [36] Mei X, Ling H B. Robust visual tracking using L (1) Minimization [A]. *IEEE International Conference on Computer Vision [C]*. 2009: 1436–1443.
- [37] Avidan S. Support vector tracking [A]. *Computer Vision and Pattern Recognition [C]*. 2001.
- [38] Li X, Dick A, Wang H, et al. Graph mode-based contextual kernels for robust SVM tracking [A]. *IEEE International Conference on Computer Vision [C]*. IEEE, 2011: 1156–1163.
- [39] Bai Y. Robust tracking via weakly supervised ranking SVM [A]. *IEEE Conference on Computer Vision and Pattern Recognition [C]*. 2012: 1854–1861.
- [40] Tang F, Brennan S, Zhao Q, et al. Co-tracking using semi-supervised support vector machines [A]. *IEEE International Conference on Computer Vision [C]*. 2007: 1–8.
- [41] Kearns M J, Valiant L G. Cryptographic limitations on learning Boolean formulae and finite automata [J]. *Journal of the Acm*, 1994, 41 (1): 433–444.
- [42] Freund, Yoav, Schapire, et al. Using and combining predictors that specialize [J]. *Stoc.*, 1997: 334–343.
- [43] Li P, Burges C J C, Wu Q. McRank: learning to rank using multiple classification and gradient boosting [A]. *International Conference on Neural Information Processing Systems [C]*. 2007: 897–904.
- [44] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34 (7): 1409–1422.
- [45] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 33 (8): 1619–1632.
- [46] Park S B, Yoo D S. Quadratic Kalman filter object tracking with moving pictures [J]. *IEEE International Conference on Computer Vision*, 2016, 20 (1): 53–58.
- [47] Ristic B, Arulampalam S, Gordon N. Beyond the Kalman filter – particle filters for tracking applications [J]. *IEEE Trans of Aerospace & Electronic Systems*, 2003, 19 (7): 37–38.
- [48] Ukrainitz Y, Sarel B. Mean-shift object tracking [J]. *Iapr*, 2012, 21 (5): 767–771.
- [49] Ma C, Huang J B, Yang X, et al. Hierarchical convolutional features for visual tracking [A]. *IEEE International Conference on Computer Vision [C]*. IEEE Computer Society, 2015: 3074–3082.