DOI:10.13878/j.cnki.jnuist.2017.06.010

刘忠耿[1]　练智超[1]　冯长驹[1]

多种信息融合的实时在线多目标跟踪

摘要

多目标跟踪算法在目标发生遮挡、目标快速运动时容易跟踪失败,而且无法从失败中恢复跟踪.针对该问题,首先利用目标的外观信息、运动信息和形状信息多种信息融合的目标特征表示,准确地计算目标间的相似性,使同一目标之间相似性距离尽量小,不同目标间的相似性距离尽量大;其次,基于判别能力强大的相关滤波器和卡尔曼预估器结合的单目标跟踪器可以在目标遮挡、快速运动中准确地跟踪目标.实验结果表明,多目标跟踪算法能够实时准确地跟踪被遮挡的目标和快速运动的目标.

关键词

多目标追踪;多信息融合;相关滤波器;卡尔曼滤波器

中图分类号 TP391.41

文献标志码 A

收稿日期 2017-08-28

资助项目 江苏省自然科学青年基金(BK20150784);中国博士后基金面上项目(2015M581800);中央高校基本科研业务费专项资金(30917011324);江苏省社会安全图像与视频理解重点实验室创新基金(30920140122007)

作者简介

刘忠耿,男,硕士生,主要研究方向为目标追踪.lzg_njust@163.com

练智超(通信作者),男,博士,副教授,主要研究方向为图像处理、模式识别与医学图像分析.lzcts@163.com

1 南京理工大学 计算机科学与工程学院,南京,210094

0 引言

多目标追踪在计算机视觉领域有着重要的应用,例如视频监控、运动分析、机器人导航以及自动驾驶等.多目标追踪的任务可以分解为2个部分,每个部分处理各自独立的问题.第1部分是目标检测,在每一帧中独立地进行检测目标数目和目标位置;第2部分是依赖目标外观信息和运动信息匹配检测和跟踪.文献[1-2]等说明目标检测在多目标追踪算法中起着重要的作用.传统的目标检测算法[3-4]使用滑动窗口在不同尺度的输入图像中选定某一个候选区域,然后提取haar-like、hog等特征,使用传统的机器学习方法训练分类器,比如Adaboost、SVM等,最后对候选区域进行分类,区分目标和背景.传统方法使用滑动窗口采样,样本数目巨大,为了平衡检测速率,一般采用计算简单的图像特征,这些特征的表现能力非常有限,比如模板匹配特征(hog特征)在目标发生形变时,将无法检测到目标.近年来,随着深度学习技术的发展,深度卷积网络进行分类时能够学习到描述能力强大的特征.文献[5]中,RCNN通过selective search算法从输入图像中提取可能存在目标的候选区域,提取候选区域的深度卷积特征,最后将特征送入分类器进行分类,区分目标和背景.之后,众多文献[6-8]对其进行改进:文献[6,8]提出了ROI pooling层来解决重复的卷积计算问题,文献[7]引入候选区域生成网络(region proposal network)替代selective search算法实现了端到端的深度学习训练算法.文献[5-8]均是采用对候选区域分类的方法解决目标检测问题,速度较慢.最近,文献[9-10]通过深度学习端到端的回归方法检测目标.YOLO[9]和YOLO9000[10]速度快、效果好,因此,本文采用YOLO9000作为目标检测算法.

多目标追踪算法主要基于通过检测实现跟踪,然后将每一帧的检测目标位置连接匹配成各个目标的轨迹.2015年之前,很多工作[11-13]关注于设计一个健壮的、完美的、全局最优的算法来求解数据匹配问题.Pirsiavash等[11]将匹配检测到目标轨迹集合的任务视为一个图模型,并用K最短路径优化算法求解.Leal-Taixé等[12]将数据匹配问题视为一个图模型,图的节点为目标检测的位置,图的边为两两检测的相似性度量,采用线性规划来快速求解优化问题.Dicle等[13]发现相似目标外观信息容易混淆,他们提出了一种更加关注目标运动信息的模型,匹配检测的相似性仅仅使用简单的距离度量.文献[14-

18]研究发现,借助强大的目标特征表示可以很大地提高目标跟踪的性能,因此,近年来,多目标跟踪算法更加关注于设计一个强有力的相似性度量方法,它们大部分基于一些强大表示能力的特征.Fagot-Bouquet 等[14]使用目标的外观稀疏表示;Kim 等[15]采用了一种在线的自适应的更新机制来保证模型的鲁棒性;Kieritz 等[16]采用了积分通道特征.深度学习方法同样对目标追踪有着重要的影响:Sadeghian 等[17]使用递归神经网络来编码目标的外观信息、运动信息以及相互作用信息;Tang 等[18]采用深度度量学习,学习到一种相似性度量计算方法.

本文提出一种新的多目标追踪算法,如图 1 所示,该算法由先进的目标检测模块和强大的特征表示模块组成.在目标检测模块,使用 YOLO9000 深度网络在人脸和车辆等数据集上重新训练,得到了一种专门检测人脸和车辆的深度网络.在特征表示模型,结合了目标的运动信息、形状信息以及外观信息,然后计算特征表示的欧氏距离来度量目标间的相似性.实验证明,本文算法效果显著,而且在使用 GPU 的条件下(NVIDIA TITAN X),速度可以达到实时性要求.

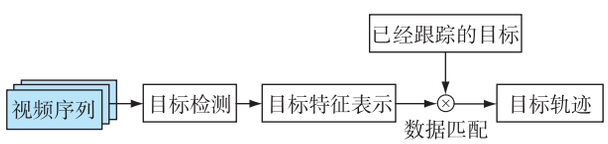


图 1 多目标追踪流程

Fig. 1 Flowchart of multiple object tracking

本文首先介绍目标检测算法;随后,描述如何结合目标的运动信息、形状信息以及外观信息;最后为实验结果总结.

1 目标检测

基于深度学习方法的一个特点就是实现端到端的检测.相对于其他目标检测与识别方法[5]将目标识别任务分类目标区域预测和类别预测等多个流程,YOLO[9]将目标区域预测和目标类别预测整合于单个神经网络模型中,实现在准确率较高的情况下快速目标检测与识别的目的,更加适合现场应用环境.后续研究中进一步优化 YOLO 网络结构,提高了 YOLO 准确率[10].

传统目标检测方法[3-4]一般采用滑动窗口法提取目标候选区域,然后采用分类器分类.最近 RCNN[5]采用候选区域生成算法产生候选区域,输入深度卷积网络提取特征,最后采用分类器进行分类.这种方法流程复杂,速度慢且训练困难.YOLO 采用一整个卷积神经网络来回归预测目标的位置已经目标的类别.YOLO 算法流程如下:首先将输入图像划分为 $S\times S$ 个网格.如果某个目标的中心落入该网格中,则该网格就负责检测该目标.每个网格预测存在某个类别的物体的概率以及目标的 B 个包围框.每个包围框预测物体的位置坐标参数以及目标的置信度.YOLO9000 针对 YOLO 召回率低、定位不准确等缺点进行改进,而且将网络进一步精简,使检测速度进一步提升.在本文实验中,在人脸数据集 FDDB[19]和车辆数据集 KITTI[20]上重新训练 YOLO 网络得到人脸检测模型和车辆检测模型,为多目标跟踪算法提供检测服务.

2 多种信息融合的在线多目标追踪算法

强健的目标之间的相似性可以提高多目标跟踪算法的性能.为此,本文结合了目标外观信息、运动信息和形状信息.其中外观信息相似性采用深度卷积特征的欧氏距离作为度量,运动信息和形状信息由卡尔曼预估器和相关滤波器结合的单目标跟踪算法获得.

2.1 外观信息

计算外观相似性应该满足:同一目标之间的外观应该尽量相似,不同目标之间的外观应该尽量不同.近年来,深度学习在计算机视觉领域得到迅速发展,在图像分类、目标检测和目标识别等领域表现出色.不同于传统的手动设计的图像特征,深度卷积特征具有强大的表示能力.借助深度网络的强大迁移能力,本文在图像分类任务 ImageNet 上预训练 GoogleNet,然后提取 pool5 层的深度卷积特征应用到目标跟踪问题上.为了计算目标之间的外观相似性,首先归一化特征向量,最后使用欧式距离来度量.

$$S_{\text{appearance}}(\boldsymbol{d}_i,\boldsymbol{d}_j)=\|\boldsymbol{d}_i-\boldsymbol{d}_j\|, \tag{1}$$

其中,$S_{\text{appearance}}$表示计算第 i 个目标从 GoogleNet 上提取深度卷积特征 $\boldsymbol{d}_i$ 和第 j 个目标特征 $\boldsymbol{d}_j$ 之间的外观相似性,$\|\cdot\|^2$ 表示向量二范数.

2.2 运动信息和形状信息

同类目标之间的外观信息判别能力弱,不容易区分同类目标,为此,本文进一步结合了目标的运动信息和形状信息.利用 Kalman 预估器和相关滤波器结合的单目标跟踪算法获得目标的运动信息和形状信息.

近年来,基于相关滤波的跟踪方法因为速度快、效果好吸引了众多研究者的目光.相关滤波器为了训练一个最优的滤波器,算法采用岭回归机器学习方法,在特征空间中回归到目标的二维高斯分布.然后在后续跟踪序列中寻找相关输出中的响应峰值来定位目标的位置.相关滤波器在运算中巧妙应用快速傅立叶变换获得了大幅度速度提升.利用循环矩阵来模拟采样,可以做到密集采样,增加了模型的判别能力.目前基于相关滤波的拓展方法也有很多,包括核化相关滤波器[21]以及加尺度估计的相关滤波[22]等.卡尔曼预估器是一个在误差协方差最小准则下的最优估计方法,计算量小、实时性高,能利用实际的运动参数不断修正未来运动状态的估计值,提高估计精度,兼顾实时性和稳健性[23].场景中的目标在每一帧图像中的位置构成了目标运动的轨迹,引入卡尔曼预估器的目的就是根据当前帧中目标位置点的信息预测下一帧中目标的可能位置.由于相关滤波器跟踪算法在发生目标遮挡、快速运动时容易跟踪失败,因此采用了一种检测相关滤波器跟踪失败的方法[24],计算相关输出相应的峰值和 APCE,当相关滤波器跟踪失败时,使用卡尔曼预估器跟踪,框架如图 2 所示.

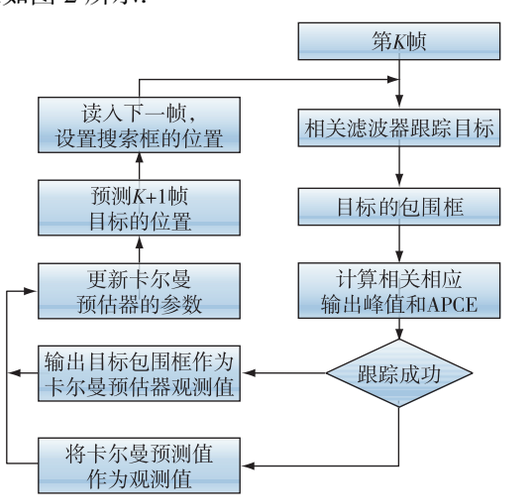


图 2　结合卡尔曼预估器和相关滤波器的单目标跟踪算法流程

Fig. 2　Flowchart of the single object tracking algorithm combining the Kalman filter and correlation filters

在多目标跟踪算法框架下,当前帧检测器检测到的目标包围框 $B=\{b_j\}_j^M$,其中 $b_j=(x_{b_j},y_{b_j},w_{b_j},h_{b_j})$,当前帧卡尔曼预估的目标状态 $P=\{p_i\}_i^N$,其中 $p_i=(x_{p_i},y_{p_i},w_{p_i},h_{p_i})$.则目标的运动相似性和形状相似性计算方法如式(2)和式(3)所示:

$$S_{\text{motion}}(p_i,b_j)=\mathrm{e}^{-\omega_1\left(\left(\frac{x_{p_i}-x_{b_j}}{w_{b_j}}\right)^2+\left(\frac{y_{p_i}-y_{b_j}}{y_{b_j}}\right)^2\right)},\tag{2}$$

$$S_{\text{shape}}(p_i,b_j)=\mathrm{e}^{-\omega_2\left(\frac{|w_{p_i}-w_{b_j}|}{w_{p_j}+w_{b_j}}+\frac{|h_{p_i}-h_{b_j}|}{h_{p_i}+h_{b_j}}\right)}.\tag{3}$$

2.3　融合多种信息的在线多目标跟踪算法

多目标检测算法的关键部分是当前帧的检测与之前帧的跟踪轨迹的数据匹配问题.本文使用匈牙利算法进行数据匹配,而且目标检测 i 和跟踪轨迹 j 的相似性计算如下:

$$S(i,j)=S_{\text{appearance}}(d_i,d_j)\times S_{\text{motion}}(p_i,b_j)\times S_{\text{motion}}(p_i,b_j).\tag{4}$$

在线多目标跟踪算法流程如表 1 所示.

表 1　在线多目标跟踪算法

Table 1　Flow of online multiple object tracking algorithm

算法 1　在线多目标跟踪算法

输入:第 t 帧图像,以及第 t 帧的目标检测集 $\boldsymbol{D}^t$,目标的跟踪轨迹集 $\boldsymbol{T}^{t-1}$.

输出:目标的轨迹集 $\boldsymbol{T}^t$

1) 利用式(4)计算目标检测 i 和跟踪轨迹 j 的相似性;

2) 利用匈牙利算法匹配目标检测集 $\boldsymbol{D}^t$ 和目标的跟踪轨迹集 $\boldsymbol{T}^{t-1}$,当相似性 $S(i,j)$ 超过阈值 θ,则目标检测 i 和跟踪轨迹 j 不相匹配;

3) 未匹配目标跟踪的检测集 $\boldsymbol{D}_{\text{fail}}^t$,这些被认为是新出现在场景的目标集合,其中的每一个检测分配一个卡尔曼预估器和相关滤波器结合的单目标跟踪器;

4) 未分配目标检测的跟踪集 $\boldsymbol{T}_{\text{fail}}^{t-1}$,记录其未分配的次数,当该次数超过某个阈值 τ,则认为目标离开场景,销毁目标对应的跟踪器;

5) 成功匹配目标检测和目标跟踪的集合,用检测集 $\boldsymbol{D}_{\text{success}}^t$ 设置卡尔曼预估器的观测值,更新卡尔曼预估器的参数.

3　实验结果

本文实验程序在 NVIDIA TITAN X 环境下,针对目标遮挡、快速运动以及目标交汇进行实验,实验证明,本文的多目标跟踪算法具有很强的鲁棒性.在 NVIDIA TITAN X 下,速度可以达到 25 fps 以上.如图 3 所示,本文的多目标跟踪算法可以在目标遮挡、快速运动等追踪困难场景下鲁棒地跟踪目标.如图 3a 所示,目标 ID3 和 6 在发生交汇的时候,得益于多种信息融合的目标特征表示以及卡尔曼预估器和相关滤波器结合的单目标跟踪算法,使多目标跟踪算法可以很好地跟踪各个目标.如图 4 所示,本文算法同样可以很好地跟踪密集的小目标.在目标发生形变及遮挡时,算法依然能够准确地跟踪目标.本文算法可以为后续目标行为分析、目标检索等问题提供有效的技术支持.

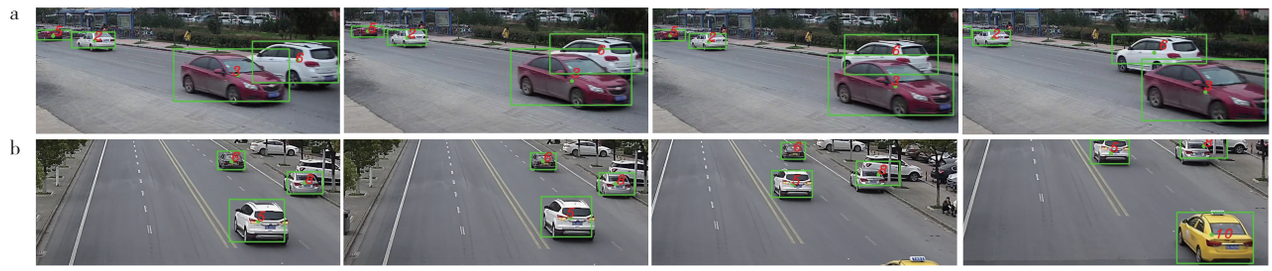


图 3　车辆追踪实验结果

Fig. 3　Experiment results of vehicles tracking

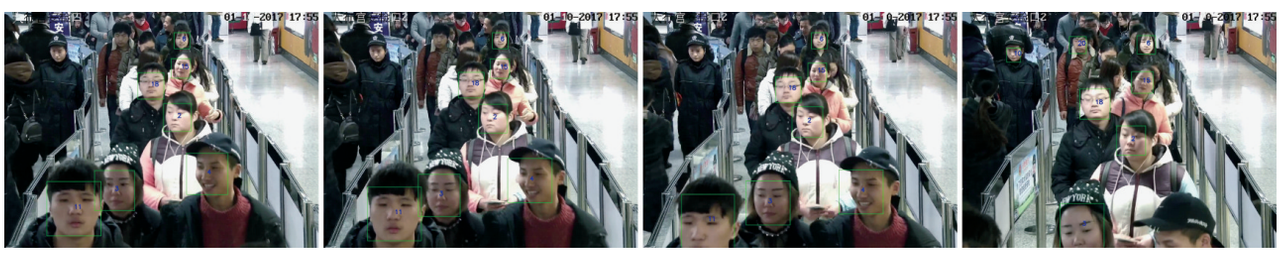


图 4　人脸追踪实验结果

Fig. 4　Experiment results of faces tracking

4　结论

本文针对目标帧间位移过大时导致目标在相关滤波器搜索区域消失的问题,利用卡尔曼预估器预测目标下一帧中的目标位置,以此位置为中心设置候选搜索区域可以很好地解决目标快速运动跟踪丢失的问题.针对目标被遮挡时,本文设计了多种信息融合的目标特征表示,准确地计算目标之间的相似性,而且目标完全遮挡时,可以利用卡尔曼预估器估计目标的后续位置,因此算法对目标遮挡问题鲁棒性较好.另外,实验证明,本文算法在各种场景下均能达到实时跟踪.

参考文献

References

[1] Milan A, Leal-Taixe L, Reid I, et al. MOT16: A benchmark for multi-object tracking[J]. arXiv e-print, 2016, arXiv:1603.00831

[2] Leal-Taixé L, Milan A, Reid I, et al. MOTChallenge 2015:Towards a benchmark for multi-target tracking[J]. arXiv e-print,2015,arXiv:1504.01942

[3] Viola P,Jones M.Rapid object detection using a boosted cascade of simple features[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003:511-518

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

[5] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE International Conference on Computer Vision and Pattern Recognition,2014:580-587

[6] Girshick R.Fast R-CNN[C]//IEEE International Conference on Computer Vision,2015:1440-1448

[7] Ren S P,He K M,Girshick R,et al.Faster R-CNN:Towards real-time object detection with region proposal networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,39(6):1137-1149

[8] He K M,Zhang X Y,Ren S Q,et al.Spatial pyramid pooling in deep convolutional networks for visual recognition [J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1904-1916

[9] Redmon J,Divvala S K,Girshick R,et al.You only look once:Unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788

[10] Redmon J,Farhadi A.YOLO9000:Better,faster,stronger [J].arXiv e-print,2016,arXiv:1612.08242

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

[12] Leal-Taixé L, Pons-Moll G, Rosenhahn B, et al. Everybody needs somebody:Modeling social and grouping behavior on a linear programming multiple people tracker [C]//IEEE International Conference on Computer Vision,2011:120-127

[13] Dicle C,Camps O I,Sznaier M,et al.The way they move: Tracking multiple targets with similar appearance[C]//IEEE International Conference on Computer Vision,

2014:2304-2311
[14] Fagot-Bouquet L, Audigier R, Dhome Y, et al. Improving multi-frame data association with sparse representations for robust near-online multi-object tracking [C] // European Conference on Computer Vision, 2016: 774-790
[15] Kim C, Li F X, Ciptadi A, et al. Multiple hypothesis tracking revisited [C] // IEEE International Conference on Computer Vision, 2015: 4696-4704
[16] Kieritz H, Becker S, Hubner W, et al. Online multi-person tracking using integral channel features [C] // IEEE International Conference on Advanced Video and Signal Based Surveillance, 2016: 122-130
[17] Sadeghian A, Alahi A, Savarese S. Tracking the untrackable: Learning to track multiple cues with long-term dependencies [J]. arXiv e-print, 2017, arXiv: 1701.01909
[18] Tang S Y, Andres B, Andriluka M, et al. Multi-person tracking by multicut and deep matching [C] // European Conference on Computer Vision, 2016: 100-111
[19] Jain V, Learned-Miller E. FDDB: A benchmark for face detection in unconstrained settings [R]. Technical Report UM-CS-2010-009, University of Massachusetts, 2010
[20] Urtasun R, Lenz P, Geiger A. Are we ready for autonomous driving? The KITTI vision benchmark suite [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3354-3361
[21] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596
[22] Danelljan M, Hager G, Khan F S, et al. Accurate scale estimation for robust visual tracking [C] // British Machine Vision Conference, 2014, DOI: 10.5244/C.28.65
[23] Liu R M, Li X L, Han L, et al. Track infrared point targets based on projection coefficient templates and non-linear correlation combined with Kalman prediction [J]. Infrared Physics & Technology, 2013, 57(2): 68-75
[24] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps [J]. arXiv e-print, 2017, arXiv: 1703.05020

An online real-time multiple object tracker with multiple information integration

LIU Zhonggeng[1] LIAN Zhichao[1] FENG Changju[1]
1 School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing 210094

Abstract The multiple object tracking (MOT) algorithm will fail when its target is occluded or in fast motion, furthermore, it cannot recover from drifting. To solve these problems, firstly, we employ integrated information to enhance the representation of objects, which includes the target's appearance, shape and motion information. By means of the integrated information, we can accurately calculate the similarity, which is as similar as possible between the same targets and as different as possible between the different targets. Secondly, we propose a novel real-time single object tracker based on the combination of the discriminative correlation filters (DCF) and the Kalman filters, which is robust to occlusion and fast motion. Extensive experiments have been done, and results show that the proposed MOT algorithm can accurately track the target in case of occlusion or fast motion in real time.

Key words multiple object tracking; multiple information integration; correlation filters; Kalman filter