

文章编号: 2095-2163(2020)10-0006-05

中图分类号: TP391

文献标志码: A

基于人脸识别的客流分析技术研究与实践

姚 砺, 孙 汇

(东华大学 计算机科学与技术学院, 上海 200000)

摘要: 在复杂的场景中,多目标跟踪算法会由于遮挡出现目标丢失和误报的情况,导致目标跟踪失败。针对以上问题,本文提出基于人脸识别的多目标跟踪算法。首先,采用基于深度学习的YOLOv3进行目标行人检测;其次,对目标行人脸部特征点进行人脸姿态估计,用欧拉角表示人脸朝向信息,再对正向人脸通过人脸特征比对实现数据关联,非正向人脸采用基于Deep SORT的多目标跟踪算法进行目标跟踪。实验证明,在一定程度上提高遮挡情况下多目标跟踪的准确性,降低了行人ID跳变次数,在实时性和鲁棒性上也有良好的效果。

关键词: YOLOv3; 人脸姿态估计; 人脸特征比对; Deep SORT; 多目标跟踪

Research and Implementation of Passenger Flow Analysis Technology Based on Face Recognition

YAO Li, SUN Hui

(College of Computer Science and Technology, Donghua University, Shanghai 200000, China)

[Abstract] In complex scenes, the multi-target tracking algorithm will cause target loss and false alarms due to occlusion, resulting in serious target tracking failure. In response to the above problems, this paper proposes a multi-target tracking algorithm based on face recognition. First, YOLOv3 based on deep learning is used for target pedestrian detection, and then face pose estimation is performed on the target pedestrian, Euler angle is used to represent face orientation information, and then the positive face is used to realize data association through face feature comparison. The forward face adopts the multi-target tracking algorithm based on DeepSORT for target tracking. This method proves through experiments that the accuracy of multi-target tracking under occlusion is improved to a certain extent, the number of pedestrian ID jumps is reduced, and it has good effects in real-time and robustness.

[Key words] YOLOv3; Face pose estimation; Face feature comparison; Deep Sort; Multi-target tracking

0 引言

在科技高速发展的时代,智能化商场是未来的发展趋势,商场中准确的客流分析系统有利于优化资源分配以及预判商业发展趋势。高准确率以及高性能的目标检测与跟踪是优化客流分析系统的重要条件,基于深度学习技术的目标检测和跟踪具有更高的准确率,有效的降低了因遮挡造成的目标丢失和误报的情况,并且具有更好的鲁棒性。

目前,比较流行的多目标跟踪算法是基于检测的目标跟踪,检测部分在很大程度上影响了目标跟踪的最终效果^[1]。在线的多目标跟踪算法比离线模式的多目标跟踪算法具有更高的准确性^[2],更适用于实时的多目标跟踪。但是在线的多目标跟踪算法采用逐帧关联的方式,只能截止到当前帧的信息,容易受到遮挡和外观变化的影响^[3]。Bewle 在基于检测的在线多目标跟踪框架中提出深度关联度量的跟踪算法 Deep SORT 来解决多目标跟踪问题,采用

卡尔曼滤波进行预测,利用运动信息和外观特征相融合的度量方式进行数据关联^[4],但对复杂环境下的多目标跟踪,仍会由于遮挡出现误检和漏检的问题。

针对复杂的商场场景中的客流分析,基于行人检测的目标跟踪算法很容易由于遮挡问题以及行人体型、衣服颜色等外观特征信息影响跟踪结果,导致跟踪误检、漏检。本文在基于行人检测的目标跟踪基础上,提出增加人脸特征比对来提高跟踪算法的准确性,通过对检测算法得到的目标行人进行人脸姿态估计,得到脸部朝向信息,判断目标行人是否是正向人脸,对正向人脸的目标行人采用人脸特征点比对算法实现数据关联,非正向人脸采用运动信息和外观特征相融合的度量方式进行数据关联,实现多目标跟踪。相对于传统的多目标跟踪算法,本文提出的算法有效解决了目标因遮挡等复杂情况造成的跟踪失败问题,并有效的减小误检、漏检对跟踪造

作者简介: 姚 砺(1967-),男,博士,副教授,主要研究方向:图像处理; 孙 汇(1995-),女,硕士研究生,主要研究方向:图像处理。

通讯作者: 孙 汇 Email: 1228006939@qq.com

收稿日期: 2020-04-24

成的影响。

1 基于人脸识别的多目标跟踪方法

基于人脸识别的多目标跟踪方法主要包括行人检测、人脸特征比对、基于 Deep SORT 的多目标跟踪。首先采用 YOLOv3 检测算法对视频帧进行行人检测;其次,对目标行人脸部 6 个关键点采用 3D 脸部模型对应投影到 2 维图像坐标系中的点集之间的

变换关系矩阵来进行人脸姿态估计,并用欧拉角表示脸部朝向信息;最后,对正向人脸采用人脸特征比对算法进行数据关联,非正向人脸采用运动信息和外观特征相融合的度量方式进行数据关联,实现多目标跟踪。本文提出的多目标跟踪算法框架如图 1 所示。

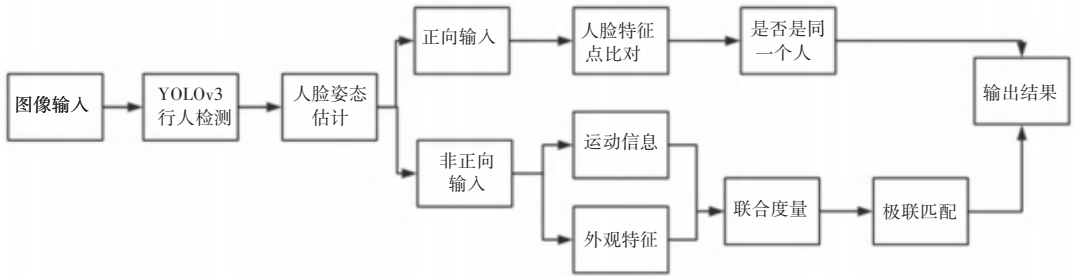


图 1 基于人脸识别的多目标跟踪方法框架

Fig. 1 Multi-target tracking method framework based on face recognition

2 基于人脸识别的多目标跟踪关键技术

本文采用基于深度学习的 YOLOv3s 算法进行行人检测,对目标行人中的正向人脸采用基于特征点的人脸比对算法,对目标行人中的非正向人脸采用基于 Deep SORT 的多目标跟踪算法,实现复杂的场景中的客流分析系统,提高客流分析的准确率和速度。

2.1 基于深度学习的行人检测

在基于检测的多目标跟踪算法中,检测器的检测效果直接影响多目标跟踪的最终结果。基于深度学习的检测方法主要分为以 SSD、YOLO 为首的 one-stage 和以 RCNN 系列为首的 two-stage 两种。one-stage 的目标检测方法直接通过主干网络给出类别和位置信息,经过单次检测生成最终的结果,two-stage 检测算法将检测问题划分为两个阶段,首先产生候选区,然后对候选区分类。相对来说 two-stage 算法在准确度上有优势,one-stage 算法在速度上有优势。YOLOv3 是使用一个 CNN 网络直接预测不同目标的类别与位置的 one-stage 算法,精确度高并且速度快,能够满足实时的检测和检测的准确性^[5]。

YOLOv3 是一种端到端的实时深度学习目标检测算法,基本思想是首先将输入图片压缩到 416×416,通过特征提取网络(Darknet53 without FC layer)对输入图像提取特征,得到 S×S 大小的特征图,然后将输入图像分成 S×S 个单元网格,每个单元负责检测中心点落在该单元内的目标,每个

单元格会预测 B 个边界框和置信度。现在 YOLO 已经衍生到第三个版本,相比于前两个版本,YOLOv3 最明显的改进是采用了类似 Feature Pyramid Networks for Object Detection(FPN)架构来实现多尺度预测^[6],每个尺度预测 3 个边界框,明显提升了对小目标检测的效果,另外特征提取采用了分类效果更好的骨架网络 darknet53,其特征提取器包含了 52 个卷积层和一个全连接层,加入了多个连续的 3×3 和 1×1 的卷积层。相比于 YOLOv2 采用的 Darknet-19 网络结构使用了残差单元,提高了检测的速度。本文采用 YOLOv3 得到的检测效果与其他检测算法的对比见表 1,可以看出 YOLOv3 在实现相同准确度下要显著地比其他检测方法速度更快。

表 1 本文基于 YOLOv3 的检测效果与其他算法的对比

Tab. 1 This paper compares the detection effect of YOLOv3 with other algorithms

Method	Map-50	Time
[B]SSD321	45.4	61
[C]DSSD321	46.1	85
[D]R-FCN	51.9	85
[E]SSD513	50.4	125
[F]DSSD513	53.3	156
[G]FPN FRCN	59.1	172
YOLOv3-320	51.5	22
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

原有的 YOLOv3 训练的是 COCO 数据集,训练

种类有 80 个,因为本文只针对行人进行检测,所以本文使用 VOC2007 数据集中 2000 张行人图片以及 2 000 张 MOT16 数据集中的行人图片,将数据集中 70% 作为训练集,30% 作为测试集。在训练中,设置每次迭代要训练的图片数量为 64, subdivision = 2, 最大迭代次数为 50 200, 学习率为 0.001。训练迭代到 760 次时,最终的 loss 收敛到 1.2, Avg IOU 的值为 0.89, 平均精度 (map) 为 0.55, 准确率 (AP) 为 0.8 左右。

2.2 人脸特征比对

多目标跟踪算法在复杂的场景中,会出现同一个目标行人由于遮挡被识别成不同的 ID,也会出现不同的目标行人由于衣服颜色等外观特征被识别成同一个 ID 的情况,为了提高多目标跟踪的准确性,本文增加人脸特征比对,对目标行人中的正向人脸通过基于脸部特征点的人脸比对算法进行数据关联。

2.2.1 人脸姿态估计

为了得到正向人脸,需要对检测到的目标行人通过人脸姿态估计,获得人脸朝向的角度信息。人脸朝向角度可以用三个欧拉角表示,即 pitch、yaw、roll,分别代表上下翻转、左右翻转、平面内旋转的角度。人脸姿态估计的核心思想是旋转三维标准模型一定角度,直到模型上“三维特征点”的“2 维投影”,与待测试图像上的特征点尽量重合。人脸姿势估计的过程为:

(1) 首先定义 n 个关键点的 3D 脸部模型,本文采用 $n = 6$,取鼻尖、下巴、左眼左角、右眼右角、嘴角左侧、嘴角右侧 6 个关键点。在实际应用中,不需要获取一个精确的人脸三维模型,直接采用世界中心坐标:鼻尖:(0.0, 0.0, 0.0),下巴:(0.0, -330.0, -65.0),左眼左角:(-225.0f, 170.0f, -135.0),右眼右角:(225.0, 170.0, -135.0),嘴角左侧:(-150.0, -150.0, -125.0),嘴角右侧:(150.0, -150.0, -125.0);

(2) 采用人脸特征点检测,得到 6 个关键点的 3D 人脸模型对应的 2D 关键点;

(3) 求解 3D 点和对应 2D 点之间的转换关系,得到旋转矩阵。姿态估计过程中有 3 个坐标系,分别为世界坐标系、相机坐标系和图像坐标系。世界坐标系中的三维点 (U, V, W) 通过旋转矩阵 R 和平移向量 t 映射到相机坐标系 (X, Y, Z) ,然后通过相机的内参数矩阵映射到图像坐标系 (x, y) ^[7];

(4) 由旋转矩阵求欧拉角。先将旋转向量转化为四元数,再将四元数转化为欧拉角;

四元数的定义,式(1)

$$q = [w \ x \ y \ z]^T, \\ |q|^2 = w^2 + x^2 + y^2 + z^2 = 1. \quad (1)$$

四元数到欧拉角的转换公式(2)为:

$$\begin{pmatrix} \hat{\theta} \\ \hat{\rho} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} \hat{e} \\ \hat{e} \\ \hat{e} \end{pmatrix} \begin{pmatrix} \arctan \frac{2(wx + yz)}{1 - 2(x^2 + y^2)} \\ \arcsin(2(wy - zx)) \\ \arctan \left(\frac{2(wz + xy)}{1 - 2(y^2 + z^2)} \right) \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{u} \\ \hat{u} \end{pmatrix}. \quad (2)$$

最终得到的脸部旋转角度误差在 ± 4 度。

2.2.2 基于特征点的人脸比对

本文人脸比为 M : N 模式,对每一帧图像中的正向人脸与人像数据库进行比对,采用基于特征点的人脸识别算法。首先,对人脸图像提取特征点,通过差分高斯函数来计算图像的极值点,这些极值点具有尺度不变性和方向不变性,通过极值点与其他 26 个邻居像素点进行比较,筛选出一些高对比度和高质量的点作为特征点;其次,通过为每一个特征点计算梯度和朝向,为特征点指定方向,用一个 128 维的向量来表征图像特征,最后通过计算待测图像与数据库图像中的特征点向量的欧氏距离,得到互相匹配的特征点对,特征点匹配对最多的图像即为正确匹配图像。

将基于特征点的人脸比对算法在数据集 yale face-database 中实验,测试得到人脸比对结果平均准确率为 98%,比对效果良好,具有优秀的识别率,并且在时间复杂度上也具有良好的表现。

2.3 多目标跟踪

本文采用基于 Deep SORT 的多目标跟踪算法对目标行人进行跟踪,该算法是一种基于检测跟踪的在线实时跟踪算法。在实时目标跟踪过程中,提取目标的表观特征进行最近邻匹配,改善有遮挡情况下的目标追踪效果,减少了目标 ID 跳变的次数。同时采用递归的卡尔曼滤波和逐帧的数据关联,并使用级联匹配策略解决因遮挡导致的卡尔曼滤波梯度弥散问题,很好的改善了遮挡后跟踪序号跳变的问题。

2.3.1 数据关联

多目标跟踪的核心是解决目标关联问题,本文使用 Deep SORT 算法在目标关联中采用运动信息和外观特征相融合的度量方式^[8]。运动信息的关联使用检测与轨迹在卡尔曼滤波器预测的位置之间的马氏距离来描述运动匹配程度,马氏距离是一种有效的计算两个未知样本集的相似度的方法^[9],公

式(3):

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i),$$

其中, $d^{(1)}(i, j)$ 表示第 j 个检测到的目标和第 i 条轨迹之间的运动信息匹配度, d_j 表示第 j 个检测框的位置, y_i 表示第 i 个追踪器对目标的预测位置, S_i 表示第 i 条轨迹由卡尔曼滤波器预测得到的在当前时刻观测空间的协方差矩阵。由于在遮挡情况下, 单独使用马氏距离为匹配度量不精确, 会导致严重的 ID 跳变, 所以引入一个新的判断数据关联的标准, 即外观信息匹配。使用外观空间中第 i 个轨迹和第 j 个检测之间的最小余弦距离为度量标准, 具体公式(4)如下:

$$d^{(2)}(i, j) = \min\{1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in R_i\},$$

其中, r_j 表示每一个检测块 d_j 对应的一个特征向量, 满足 $\|r_j\| = 1$; r_k 表示每一个追踪目标储存的最近 100 帧成功关联的特征向量。

最后使用运动匹配度和外观信息匹配度的线性加权作为最终的联合度量, 式(5):

$$c_{(i, j)} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j)$$

马氏距离对于短期的预测和匹配效果很好, 最小余弦值对于长时间丢失的轨迹而言比较有效。

2.3.2 级联匹配

在复杂场景中, 行人聚集比较多, 跟踪目标容易被遮挡, 如果目标被长时间遮挡, 卡尔曼滤波预测的不确定性就会大幅度的增加, 同时状态空间内的可观察性也会很大程度地降低, 进而导致目标轨迹碎片增加以及轨迹不稳定。引入级联匹配, 其核心思想是由小到大对消失相同的轨迹进行匹配, 即优先考虑更常见的目标, 提高了多目标跟踪算法的准确性。

通过 Deep SORT 跟踪算法对 YOLOv3 检测出的行人进行跟踪, 取得较好的效果, 而且能够实现实时

的目标跟踪, 有效的提高客流分析系统的多目标跟踪。

3 实验结果与分析

3.1 实验数据与环境

为了验证增加人脸特征比对能否提高多目标跟踪的准确性, 本文采用数据集 MOT16 中的视频集进行测试。MOT16 是 2016 年提出的衡量多目标检测跟踪算法标准的数据集, 视频分辨率为 1920×1080。本文的实验平台采用 Linux 系统下的 Ubuntu16.04, GTX1060, 3.2GHz, 内存为 16 GB。

3.2 结果分析

对 MOT16 中的一个数据集进行测试, 采用 YOLOv3 检测算法实现的客流分析系统中的行人检测结果见表 2。对视频流进行行人检测, 统计了正确检测到的行人个数、错误检测到的行人个数以及未检测到的行人个数, 通过和实际存在的行人个数的比值得到检测算法的准确率、误报率和漏检率。

表 2 实验数据统计表

Tab. 2 Experimental data statistics table

客流分析	检测人数	实际人数
正确检出	47	
误检	2	52
漏检	5	

由表 2 结果可以行人检测的正确率为 90.3%, 误检率为 3.8%, 漏检率 9.6%, 对客流分析系统的行人检测效果良好, 并且检测速度为 65 ms/帧, 可以实现客流分析系统中的实时检测。

采用传统的基于 Deep SORT 多目标跟踪算法, 采用运动信息和外观信息相融合的方式进行数据关联, 将该算法在 MOT16 数据集测试, 得到的实验实验结果如图 2 所示。



图 2 基于 Deep SORT 多目标跟踪算法实验结果

Fig. 2 Experimental results of multi-target tracking algorithm based on Deep SORT

由图 2 可以看出, 帧 753 中的 1 号行人在经过帧 766 中 25 号行人的遮挡之后, 在帧 778 中重新出

现时被错误识别成了一个新的目标行人。

针对客流分析系统中的多目标跟踪, 本文实验

采用 YOLOv3 进行行人检测,对有正向人脸的行人采用人脸特征比对进行数据关联,对非正向人脸的目标行人采用基于 Deep SORT 的多目标跟踪算法进行数据关联,实验效果如图 3 所示。从图 3 中可以看出,该方法对目标行人进行了准确的检测和跟踪,并且在 1 号目标行人被 25 号目标行人完全遮挡之后,在重新出现时也能正确的关联。有效解决了基于多目标跟踪算法由于遮挡或者行人目标特征相近等问题导致结果失败的问题。

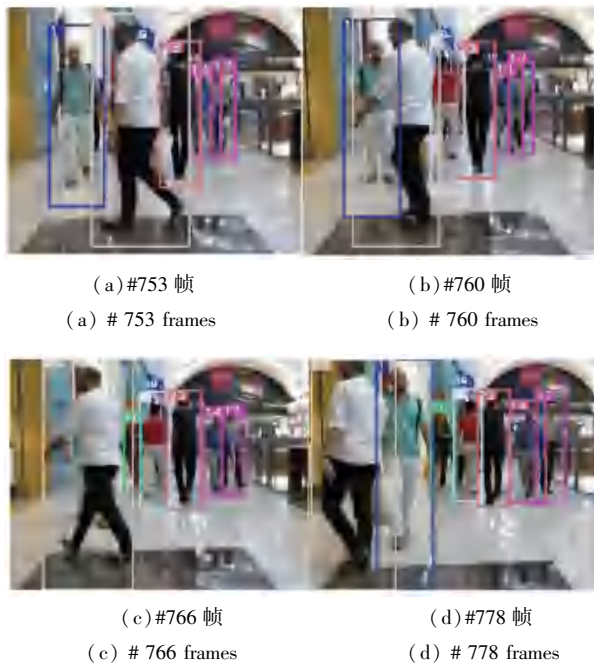


图 3 实验结果

Fig. 3 Experimental results

为了验证本文算法的有效性,本文采用 MOT 基准挑战中的评价标准:多目标跟踪准确率(MOTA),最多跟踪目标率(MT),最多丢失目标率(ML),ID 错误变换次数(IDs)。本文的跟踪算法与其他跟踪算法在 MOT 基准训练集的性能比较见表 3。

表 3 跟踪器在 MOT 基准训练集的跟踪性能比较

Tab. 3 Tracking Performance Comparison of Tracker in MOT Benchmark Training Set

跟踪算法	MOTA \uparrow	MT/% \uparrow	ML/% \downarrow	IDs \downarrow
EAMTT	52.5	19.0	34.9	910
SORT	59.8	25.4	22.7	1423
Deep SORT	61.4	32.8	18.2	781
Ours	66.7	34	18	650

实验结果表明,本文算法在客流分析系统中实现行人检测以及多目标跟踪都具有良好的效果,并

且在复杂的环境下也有良好的鲁棒性。针对多个目标时不仅能够实现准确的行人检测,而且能够在有遮挡的复杂环境中建立正确的数据关联,实现良好的多目标跟踪效果,减少了多目标跟踪中的目标丢失以及跟踪误报的问题,降低了 ID 跳变的次数。

4 结束语

为了在复杂的环境下进行准确的客流分析,本文提出基于人脸识别的客流分析方法,将人脸识别和多目标跟踪具体应用到智能化商场中,解决因遮挡造成的目标丢失以及误报的问题。

首先采用基于深度学习的 YOLOv3 算法进行行人检测,该算法在精确度相同的情况下,目标检测速度明显优于其他检测算法;其次,对检测到的目标行人通过脸部 6 个关键点的 3D 脸部模型对应投影到 2 维图像坐标系中的点集之间的变换关系矩阵,来获取人脸旋转和平移信息,采用 3 个欧拉角(pitch, yaw, roll)来表示脸部朝向信息,对于正向人脸采用人脸特征比对实现数据关联,非正向人脸采用运动信息和外观特征相融合的度量方式进行数据关联,实现在遮挡情况下,对目标行人的准确跟踪。

在数据集 MOT16 中的实验证明,本文的方法能满足实时的多目标跟踪,并且具有良好的跟踪性能,一定程度上解决了因长时间遮挡出现的行人跟踪误报或漏报的问题。

参考文献

- [1] 向伟. 基于检测的目标跟踪算法研究[D]. 上海交通大学, 2015.
- [2] BAE S H, YOON K J. Robust Online Multi-object Tracking Based on Tracklet Confidence and Online Discriminative Appearance Learning[C]// Computer Vision & Pattern Recognition. IEEE, 2014: 1208-1225.
- [3] 王慧燕, 杨宇涛, 张政. 深度学习辅助的多行人跟踪算法[J]. 中国图像图形学报, 2017, 22(3): 349-357.
- [4] BEWLEY A, ZONGYUAN G, RAMOS F, et al. Simple online and real-time tracking[C]//ICIP, 2016: 3464-3468.
- [5] 赵琼, 李宝清, 李唐薇. 基于改进 YOLOv3 的目标检测算法[J/OL]. 激光与光电子学进展: 1-16[2020-01-16].
- [6] 季航, 贾镔, 刘晓, 拓浩男. 一种基于 YOLOv3 的红外目标跟踪系统[J]. 电子设计工程, 2019, 22(27): 61-04.
- [7] 毋立芳, 张斯聪, 赵晓晴, 等. 一种人脸姿势估计新方法[J]. 信号处理, 2006, 22(1): 41-47.
- [8] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//IEEE International Conference on Image Processing(ICIP), 2017: 3645-3649.
- [9] 张巍. 基于深度学习的多目标跟踪技术研究[D]. 电子科技大学, 2019.