

西南科技大学

高级图像处理程序设计报告 公园移动人群检测跟踪识别



姓 名：冯 波

学 号：7220180157

专 业：控 制 工 程

指导老师：路锦正老师

基于 Face_Recognition 的人脸检测跟踪识别

摘要：近年来，随着计算机技术的迅速发展，人脸自动识别技术得到广泛研究与开发，人脸识别成为近 30 年里模式识别和图像处理中最热门的研究主题之一。人脸识别的目的是从人脸图像中抽取人的个性化特征，并以此来识别人的身份。一个简单的自动人脸识别系统，包括以下 4 个方面的内容：(1) 人脸检测 (Detection)：即从各种不同的场景中检测出人脸的存在并确定其位置。(2) 人脸规范化 (Normalization)：校正人脸在尺度、光照和旋转等方面的变化。或者叫做 alignment，人脸对齐，人脸校准。(3) 人脸校验 (Face verification)：采取某种方式表示检测出人脸和数据库中的已知人脸，确认两张脸是否是同一个人。(4) 人脸识别 (Recognition)：将待识别的人脸与数据库中的已知人脸比较，得出给你的脸是库里的谁。

关键词：人脸跟踪、人脸检测识别、传统人脸识别

一、引言

随着计算机技术的高速发展和人工智能推向高潮，AI 技术在各个领域有着迫切的应用需求，从自动考勤、无人卡口，到智能家居、智能安防，工业自动化，再到临床诊疗，金融投资，用户群体从行业到企业再到个人。人脸识别作为计算机视觉和 AI 领域中十分重要的一个研究方向，在经过多年发展后，已逐步走向实用化，成功应用在身份验证，智能安防，图像检索，交通管制等公共及信息安全领域。

相比于目前常用的用于识别的生物特征，人脸识别凭借其便捷快速，高智能化，人机友好，安全稳定，不宜伪造等优势一跃成为了身份识别领域的热门技术。随着各个应用领域对于安全防护的要求越来越高，人脸识别技术也毫无疑问拥有巨大发展前景和市场空间。

一个人脸检测与识别系统，包括以下 4 个方面的内容：(1) 人脸检测 (Detection)：即从各种不同的场景中检测出人脸的存在并确定其位置。(2) 人脸的规范化 (Normalization)：校正人脸在尺度、光照和旋转等方面的变化，也称为人脸对齐。(3) 人脸表征 (Face Representation)：采取某种方式表示检测出人脸和数据库中的已知人脸。(4) 人脸识别 (Recognition)：将待识别的人脸与数据库中的已知人脸比较，

得出相关信息。

二、 传统方法

传统方法分析

1. 基于几何特征的人脸识别

前面有提到，Bledsoe 提出了最早的人脸识别相关的学术论文，其中用到的便是基于几何特征的方法。基于几何特征人脸识别的实现步骤：定位人脸的面部五官特征点，并测量这几个特征点间的欧氏距离，得到一组能够描述面部特征点的矢量，如特征点的位置坐标、宽度等，或者眉毛的弯曲程度、淡浓等，以及他们之间的联系。通过计算特征点间的距离，即可找出类似的人脸。该方法的优点主要有：因为只需要存储特征矢量即可，故所需的存储空间较小；对光照的变化具有一定的鲁棒性；这也是常用的人脸识别原理，易于理解。但是想要从一个未知的人脸图像中提取具有稳定性的特征矢量仍具有一定的复杂性，因此近年来对该算法的研究在逐渐地减少。

2. 基于特征脸的人脸识别

基于特征脸方法的核心技术是主成分分析 (Principal Component Analysis, PCA)。PCA 算法最早由 Sirovitch 和 Kirby 引入人脸识别领域，它是一种基于 K-L 变换 (Karhunen-Loeve transform) 的算法，通过对人脸图像进行统计特征提取，从而实现在子空间模式下进行人脸识别。这种方法主要是利用系数变换，因此具有简便、快捷、实用的优点。但该方法对训练样本与测试样本之间的相关性要求较高，且受光照和面部表情的变化影响较大，导致识别率较低。

3. 基于模板匹配的人脸识别

事先给定一些包含不同人脸特征尺度的样本作为模板的方法称为模板匹配方法。其主要原理是：通过比较待测人脸与给定模板的图像窗口，来判定图像窗口中是不是有待测人脸。然而该方法只是简单机械的对比未知图像和已知模板，因此无法处理姿态和尺度变化丰富的人脸，当然这种简单机械的方法比较容易实现。所以目前应用得也较少。

4. 基于神经网络的人脸识别

如今学术界讨论比较多的人脸识别算法便是基于神经网络的方法，BP (Back Propagation) 神经网络学习算法是该技术的核心内容。它是受生物神经网络单元间行为特征的启发而建立的一种仿生的运算模型。其主要步骤也依赖于所设计的神经网络：将待识别人脸图像中的像素点与事先设计好的神经网络中的神经元依次对应。

这种算法的特点是：较容易提取到人脸特征，通过学习就能得到人脸识别的规律，适应于不同的人脸，而且由于是并行处理信息，速度较快。但是若神经元数量太多，样本的训练时间就会增加。

5. 基于隐马尔可夫模型的人脸识别

Samaria 等人最先提出人脸马尔可夫模型理论，他将人脸划分为 5 个代表性特征区域：额头、眼睛、鼻子、嘴巴和下巴，并且这 5 个特征区域均可用一组特征值表示。这种方法的优点即是：由于特征区域间位置的相对性，无论人脸表情如何变化，该算法的识别精度都较高。但是这种方法实现起来不仅复杂度高，运算量大，而且还需要大量的内存用于存储特征。

6. 基于弹性匹配方法的人脸识别 基于动态链接结构 (Dynamic Link Architecture, DLA) 是弹性图匹配算法的核心思想。在 1992 年，Lades M 等人曾首次在人脸识别中应用该方法，并取得了较为满意的成果。它通过使用网状的稀疏图来表示人脸图像，并在网状稀疏图的节点处对人脸图像实现 Gabor 小波分解，从而获得特征向量标记，然后对模型图进行搜索，找到最相似于待测样本脸的图，并将图中的节点按照相似性进行依次匹配，形成一个变形图。弹性匹配方法能在光照不好或脸上有装饰的情况下使用，而且当数据库较大时，该方法可以利用神经系统中的突触将神经元按照图的结构划分为若干个组织，从而很大程度上减少了识别时间。但是正因为所有输入的人脸模型都需要计算，因此该算法不仅计算量大，而且需要的内存空间也很大。

7. 基于贝叶斯决策的人脸识别

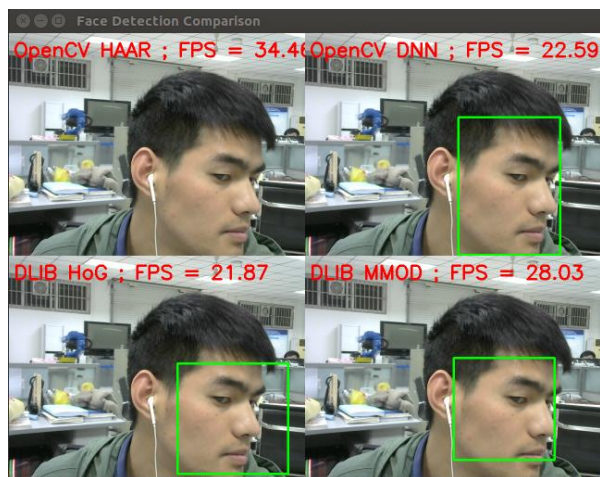
在样本有缺失的情况下，可以利用贝叶斯决策的方法实现人脸识别。该算法的主要步骤是：首先利用主观概率对某些未知的数据预估，然后使用贝叶斯公式校正发生概率，最后根据校正后的概率作出判断。贝叶斯决策算法主要是利用最大后验概率准则来解决样本有缺失情况下的分类问题，但因总需要计算概率，所以算法较为复杂。

8. 基于支持向量机的人脸识别

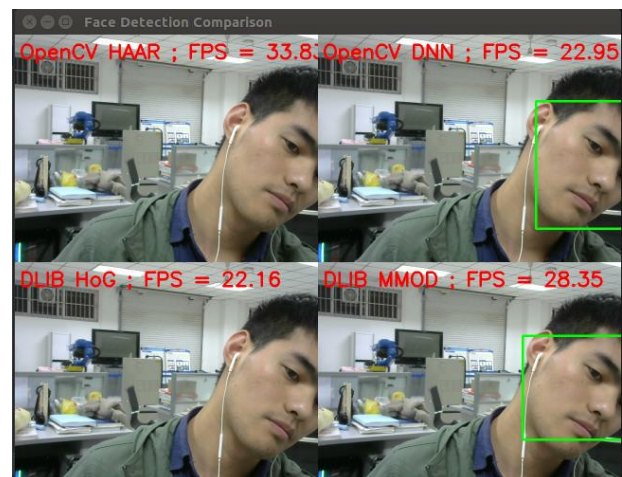
由于机器学习算法的流行，支持向量机 (support vector machine, SVM) 也成为现如今比较热门的识别方法，最早由 Vapnik 等人提出的。它的主要思想是致力于结构风险降到最低，常应用于分类回归的相关问题，其中最重要的原则是在进行样本训练时，学习机器的数量需要对应于训练样本数量。最早在人脸检测过程中使用 SVM 算法的是 Osuna 等人，用于真实人脸和虚假人脸的分类。SVM 算法与传统算

法不同的是,不需要对图像进行空间降维,而是根据实际情况,对图像进行空间升维,即将低维映射到高维,从而实现非线性问题到线性问题的转换。目前, SVM 算法是在训练样本定量的情况下,应用最多的机器学习方法。它依靠严谨的理论,能够很好地实现小样本量、非线性以及高维等实际应用中常见问题的解决。然而在进行样本训练时会耗费大量的内存是它唯一的缺点。

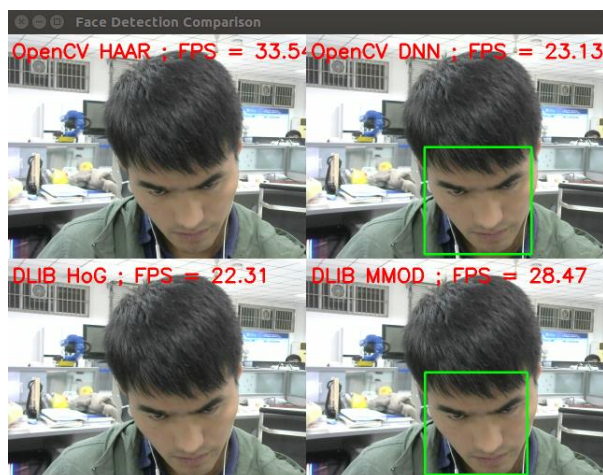
对以上部分的传统方案进行实现对比,实验结果如图 1 所示。



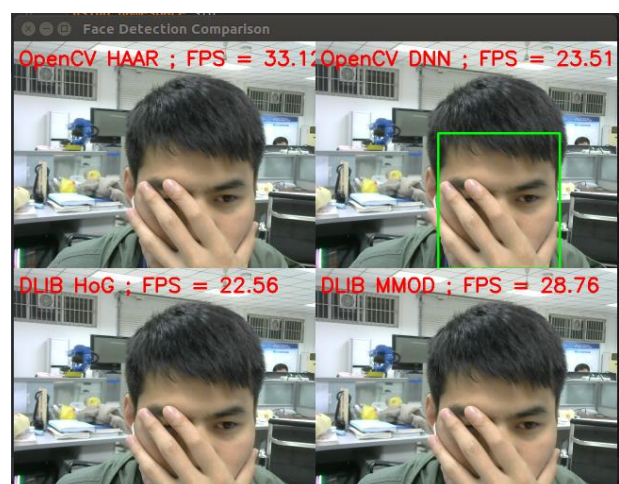
(a)



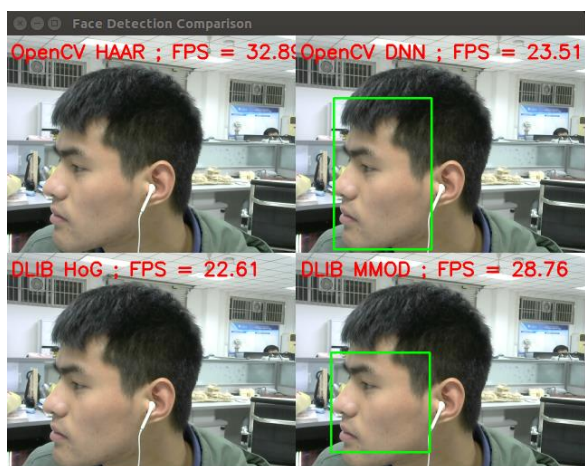
(b)



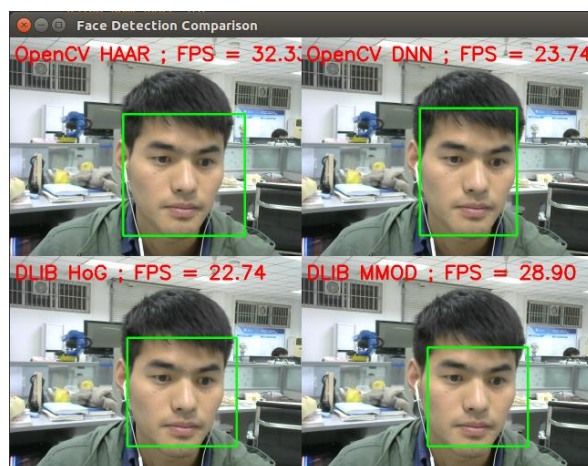
(c)



(d)



(e)



(f)

图 1. (a) (b) (c) (d) (e) (f) 几种传统方法的人脸跟踪检测算法对比

从实验过程做定性定量分析可知 DNN 的鲁棒性最高，对侧脸、遮挡等情况都具有很好的鲁棒性，运行速度 23 帧每秒左右也基本满足实时性，相对于其他的传统机器学习的方法更加的稳定。

三、 本设计原理及方案

本设计欲实现一个完整的人脸识别系统，最后达到从 USB 相机采集图像流进行实时的人脸检测并识别的功能。系统在图像采集后转换颜色为灰度，使用方向梯度直方图^[1] (Histogram of Oriented Gradients, HOG) 的方法进行人脸的检测；检测到人脸后，利用面部特征点估计^[2] (face landmark estimation) 的算法将人脸对齐；最后使用 FaceNet^[3] 提取脸部特征点，识别过程则采用支持向量机^[4] (Support vector machine, SVM) 分类器进行人脸识别。系统流程图如图 2 所示。



图 2.人脸检测流程图

3.1 HOG 人脸检测

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。(本质：梯度的统计信息，而梯度主要存在于边缘的地方)。首先将图像分成小的连通区域，称为细胞单元。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。

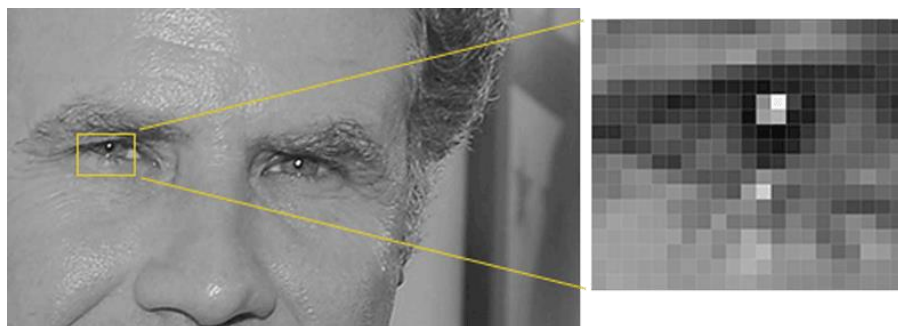


图 3.逐像素遍历卷积

计算人脸图像横坐标和纵坐标方向的梯度，并据此计算每个像素位置的梯度方向值；首先用 $[-1,0,1]$ 梯度算子对原图像做卷积运算，得到 x 方向（水平方向，以向右为正方向）的梯度分量 gradscalx ，然后用 $[1,0,-1]^T$ 梯度算子对原图像做卷积运算，得到 y 方向（竖直方向，以向上为正方向）的梯度分量 gradscaly 。然后再用以上公式计算该像素点的梯度大小和方向。将图像划分成小 cells（例如 $6*6$ 像素/cell）；统计每个 cell 的梯度直方图（不同梯度的个数），即可形成每个 cell 的 descriptor；将几个 cell 组成一个 block（例如 $3*3$ 个 cell/block），一个 block 内所有 cell 的特征 descriptor 串联起来便得到该 block 的 HOG 特征 descriptor。将图像内的所有 block 的 HOG 特征 descriptor 串联起来就可以得到图像的 HOG 特征描述器

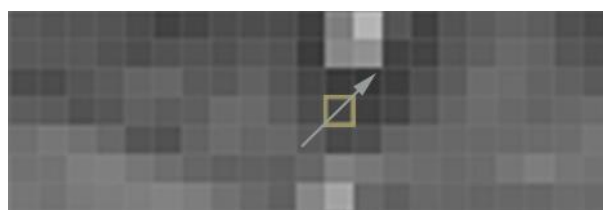


图 4.绘制明暗变化方向

3.2 面部关键点提取

使用 Vahid Kazemi 和 Josephine Sullivan 提出的面部对齐方法^[2]。对每个人脸提取 68 个关键点（landmark）：下巴，两只眼睛的外边缘，眉毛的内边缘等等。如图 5 所示。

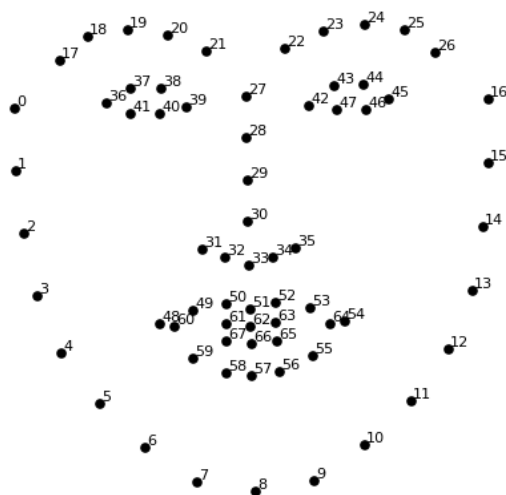


图 5.人脸的 68 个关键点示意图

提取出关键点后使用图像的仿射变换对脸部进行校正和裁剪处理，得到人脸的便于后续的特征提取，人脸对齐实验结果如图 6 所示，将 68 个关键点连成线后的效果，其中实验过程为实时对视频流进行处理。



图 6.人脸对齐实验

3.3 面部特征点估计及识别

参考文献^[4]，将校正裁剪居中的人脸图片传入训练好的 FaceNet 网络模型进行编码得到 128 个特征测量值并保存。最后，识别过程中将每个人脸表示为 128 维空间中的点，即是一个人脸为一个点。要检查两个人脸是否匹配，直接检查这两个点之间的距离是否小于 0.6（经验阈值）。

3.4 程序运行条件和步骤

- 系统平台：Ubuntu16.04 + cuda_10.0.130 + cudnn_v7.4.1.5(GTx 750Ti)
- 软件平台：Python3.5.2 + OpenCV3.3.1 + PyTorch1.0.1(自带 caffe)
- 将预识别的包含人脸的图片以人名命名放入 face_recognition/examples/ 文件夹中；
- 插入 USB 摄像机运行程序：python facerec_from_webcam_faster.py。

四、运行结果

SVM 分类过程中阈值取为 0.6，若低于 0.6 的阈值会使人脸更接近准确率更高！最终系统运行结果如图 7 所示。

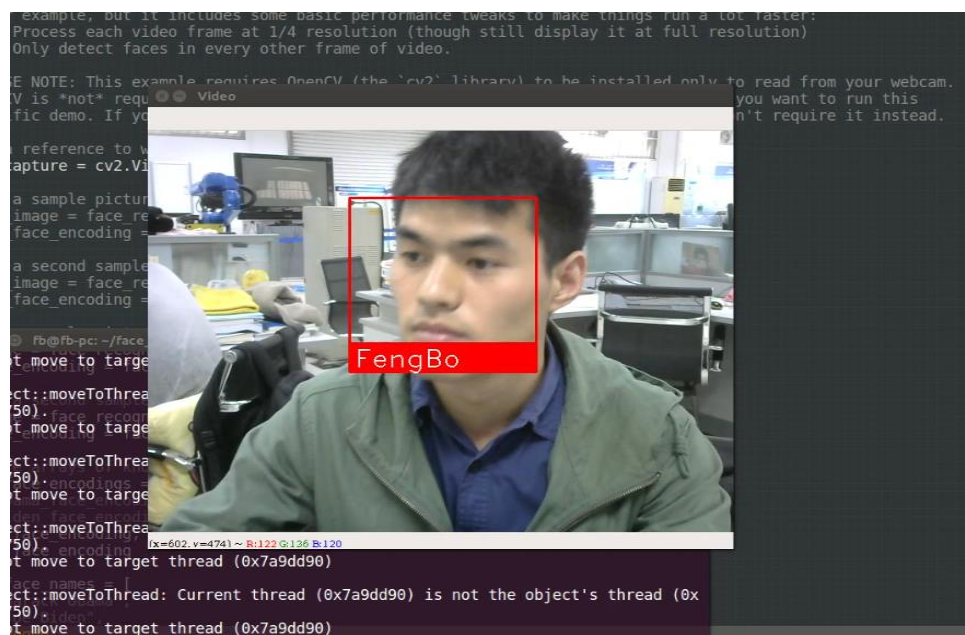


图 7.系运行截图

图 7 为在线测试运行截图，该系统能实时的进行人脸跟踪检测与识别。更重要的是本系统能够同时在线进行多张人脸的跟踪检测识别。在公园等场景则可以实现首先框选一张图片作为样本然后作为目标进行人脸的跟踪识别。

五、 讨论

在实验过程中可以看到，当人脸离摄像头较远时，人脸检测的效果不是很好，这一点可以通过程序来设定将

```
face_locations = face_recognition.face_locations(rgb_frame)
```

更改为：

```
face_locations=face_recognition.face_locations(rgb_frame,number_of_times_to_
upsample = 2)
```

这样能够检测更小的脸，但是代价就是运行帧率降低以及内存消耗更大。

总之，人脸识别作为一项非常普遍的应用目前已经有很多成熟的方案，同时也有很多更好更优秀的方案在涌现出来。十年前还停留在想象阶段的事情，现如今已经成熟投入使用，科技的进步与发展速度令人咂舌，就算作为一名平凡的普通人也唯有不断地学习，紧跟时代的步伐，才能享受科技带来的便利，才能享受当代生活的极致体验。

参考文献

- [1] N.Dalal, B.Triggs.Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 05)
- [2] V.Kazemi, J.Sullivan.One Millisecond Face Alignment with an Ensemble of Regression Trees. Computer Vision and Pattern Recognition, At Columbus, Ohio, USA
- [3] F.Schroff ,D.Kalenichenko,J.Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015
- [4] https://en.wikipedia.org/wiki/Support-vector_machine
- [5] <http://dlib.net/>
- [6] <https://cmusatyalab.github.io/openface/>

2.2 检测器

检测器主要是用来定位视频每一帧中目标的位置。TLD 中的目标检测器主要是一个级联分类器，其主要由方差滤波器、集成分类器以及最近邻分类器三部分组成。

目标在视频帧中任何可能出现的区域均可以成为目标候选区域。在通过检测器对目标候选区域进行判断时，需要连续的利用方差滤波器、集成分类器和最近邻分类器进行判断。一旦目标候选区域不符合三个部分中的任何一个部分，则候选区域将进入被拒区域，完全通过三个部分时，目标候选区域才有可能成为目标区域。

2.2.1 方差滤波器

方差滤波器是级联分类器的第一步，通过候选区域的方差与目标区域的方差比较，把方差小于目标区域方差一般的候选区域排除，因此方差滤波器可以排除大量的非目标区域。方差滤波器需要计算图像土语的灰度值方差，其计算表达式如下：

$$D(x) = E(x^2) - E^2(x) \quad (1)$$

其中， x 表示图像区域， $D(x)$ 表示图像区域的灰度值方差， $E(x)$ 表示图像区域的灰度值均值， $E(x^2)$ 表示图像区域灰度值平方之后的均值。

2.2.2 集成分类器

集成分类器是级联分类器的第二步，对通过方差滤波器的图像区域做进一步判断。集成分类器主要是由 n 个随机森林分类器构成的随机森林组成。图像区域输入到集成分类器之后，会得到相应的后验概率。通过后验概率的大小判断是否需要进入最近邻分类器。集成分类器的原理如图 2 所示，图中的实线表示一种可能的判断。

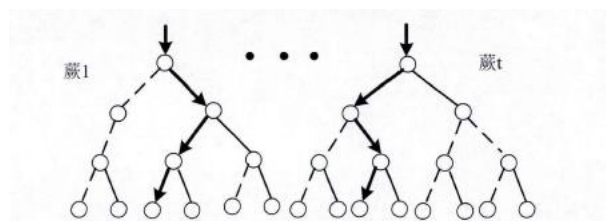


图 2 集成分类器原理

每个随机森林由一个树形结构组成，通过对图像区域进行判断，得到一个后验概率。如果 n 个随机森林分类器的后验概率平均值大于阈值，则将该图像区域判定为正样本，否则判定为负样本。

2.2.3 最近邻分类器

图像区域经过前面的方差滤波器和集成分类器判断之后,需要将其输入到最近邻分类器做进一步的判断。这一阶段主要判断图像区域和目标模型的相似度。判断方法如下:

$$S^+(p, M) = \max_{p_i^+ \in M} S(p, p_i^+) \quad (2)$$

$$S^-(p, M) = \max_{p_i^- \in M} S(p, p_i^-) \quad (3)$$

其中 P 表示图像区域, M 表示目标模型, 目标模型由多个正样本和负样本组成, p_i^+ 表示目标模型中的正样本, p_i^- 表示目标模型中的负样本。 $S^+(p, M)$ 表示图像区域 p 与目标模型 M 中的正样本相似度的最大值。 $S^-(p, M)$ 表示图像区域 p 与模型 M 中的负样本相似度的最大值, 通过这两个相似度, 最终可以求出图像区域与目标模型的最终相似度 S^r :

$$S^r = \frac{s^+}{s^+ + s^-}$$

(4)

2.3 跟踪器

TLD 中的跟踪器使用的是一种基于前向后向轨迹的 LK 光流法。在 TLD 算法中, 通过将视频帧均匀画网格, 选取每个网格的左上角顶点为特征点, 然后使用前向后向轨迹光流法来跟踪预测这些特征点在下一帧中的位置。假设当前帧为 I_t , 下一帧为 I_{t+1} , 普通光流法仅仅由当前帧 I_t 预测 I_{t+1} , 缺乏反向校验。而前向后向轨迹光流法则在由 I_t 预测出 I_{t+1} 之后, 利用 I_{t+1} 中预测得到的点, 在做反向预测, 即由 I_{t+1} 预测到 I_t , 通过这种前向后向轨迹预测可以得到一个位移偏差。如果反向预测得到的 I_t 中的特征点与原来已知的 I_t 中的特征点位移偏差超过阈值, 则将 I_{t+1} 中预测出的偏差较大的特征点排除。前向后向轨迹光流法的基本原理如下图 3 所示

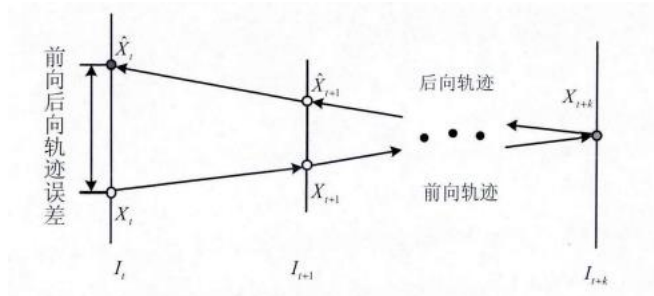


图 3 前向后向轨迹光流预测原理

2.4 学习器

TLD 的学习器又被称为 P-N 学习。P-N 学习主要是通过先验标记样本和未标记

的样本来学习分类器。P-N 学习主要由两类结构约束组成,即正约束(P 约束)和负约束(N 约束)。P-N 约束对未被标记的样本进行标记分类,之后训练分类器。

正约束是指将未知样本标记为正样本的约束条件,这里将靠近轨道附近的样本标记为正样本。负约束是指将位置样本标记为负样本的约束条件,这里将远离轨迹的样本标记为负样本。通过正负约束的相互作用来提高样本标记的准确性。

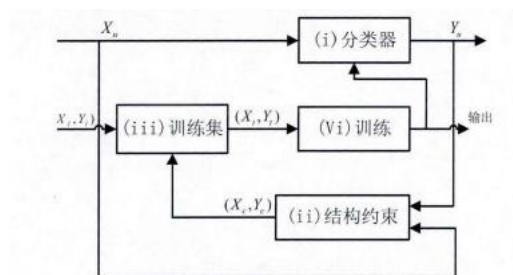


图 4 P-N 学习原理

如图 4 所示, 设 x 为特征空间 X 中的一个样本, y 表示对应的标记空间 $Y=\{-1, 1\}$ 中的一个标记, 那么可以用集合 $\{X, Y\}$ 来表示样本空间和对应的标记。P-N 学习根据已标记的样本集合 $\{X_i, Y_i\}$ 来建立分类器, 并且训练样本, 使用没有标记过的数据 X_u 来引导分类器工作, 提高分类器的性能。

三、 实验结果

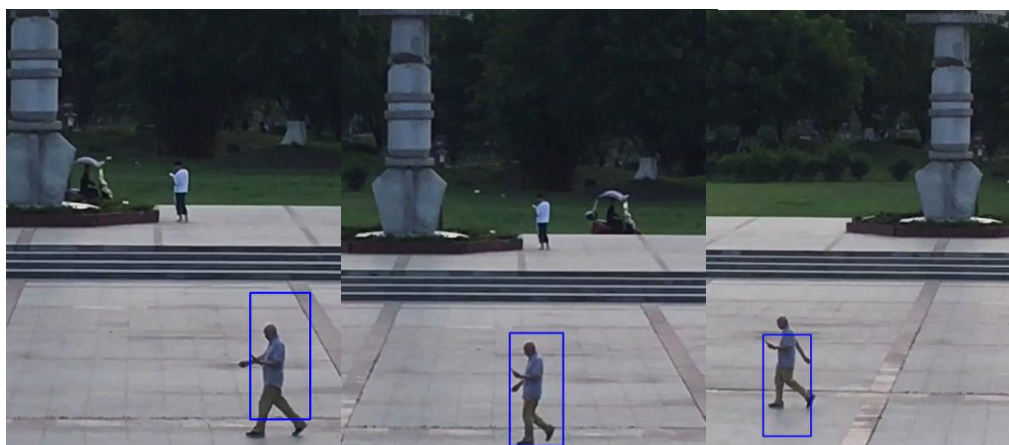


图 5 西南科技大学科技广场人体跟踪效果图

根据实验可知, TLD 跟踪算法处理视频的速度为 4 帧每秒, 由图 5 可知, TLD 跟踪算法随处理速度较慢, 但识别精度较高。

四、 总结

TLD 算法的优点在于能够通过在线学习不断的训练新的分类器并且能够不断的对分类器进行更新校正。由于算法继承了在线检测其, 因此在目标消失后具有良好的重检测能力, 一旦目标重新出现, 便能够及时的恢复跟踪。

TLD 算法的缺点在于不能够自动锁定跟踪目标, 需要人工确定跟踪目标。算法在目标遇到遮挡时很容易丢失目标或者造成误跟踪。

参考文献

- [1] 龚小彪. 基于 TLD 框架的目标跟踪算法研究[D].西南交通大学,2014.
- [2] 基于 TLD 和 fDSST 的长时间目标跟踪算法[J]. 李轶锴,吴庆宪,丁晟辉,胡鲲.电光与控制. 2019(04)
- [3] 采用核相关滤波的快速 TLD 视觉目标跟踪[J]. 王姣尧,侯志强,余旺盛,廖秀峰,陈传华.中国图象图形学报. 2018(11)
- [4] 基于多信息融合的 Mean-Shift 跟踪算法[J]. 郭瑞峰,张文辉,刘娜,彭战奎.传感器与微系统. 2018(11)
- [5] 融合位置预测的一致性匹配算法[J]. 李雪武,战荫伟,吕俊奇.计算机工程与设计.2018(09)
- [6] 基于改进的 TLD 目标跟踪算法[J]. 胡欣,高佳丽.计算机应用研究.2019(05)
- [7] 基于运动检测与 KCF 算法的目标跟踪方法[J]. 易诗,唐文.电视术.2018(09)
- [8] 基于 TLD 框架的车距检测[J]. 陈薇,卢冠男,陈梅.控制工程.2018(06)
- [9] 基于 FAST 特征点改进的 TLD 目标跟踪算法[J]. 毛晓波,周晓东,刘艳红.郑州大学学报(工学版). 2018(02)
- [10]目标跟踪算法研究及实现[D]. 郑鹏程.宁夏大学 2018