

电 子 科 技 大 学

硕 士 文 献 阅 读 综 述

学 号： 201822060706

姓 名： 朱恒伟

指导教师： 康波

学科专业： 控制工程

所在学院： 自动化工程学院

指导教师阅读后签字： _____

2020 年 03 月 01 日

1.研究背景及研究意义

随着科学技术的不断进步，人类在建筑领域取得了非常大的成就，并且随着基建能力的不断提升，城市建筑的面积不断的扩大，而且楼层的高度也在不断增高。以便在城市有限的空间中容纳更多的人类活动。不断扩大的建筑面积和不断增高的楼层高度让更多的人能够有机会在城市生活。

大型楼宇的出现，容纳了更多的人类活动。与此同时，这也为楼宇的安保、管理等工作带来了新的挑战。传统的纯人工管理方式已经不能够很好的保证楼宇的安全，而且花费如此巨大的人力和物力来管理大型楼宇是不合理的，而且非常不明智。与此同时，人类的大部分活动是在室内展开的，因此人们对室内的安全问题一直以来非常的关注，社会各界也对此高度重视。随着计算机技术的快速发展，目前在大多数楼宇中的关键通道上都安装上了摄像头，另外也或多或少的安装了烟雾报警器、火灾报警器、温湿度传感器等等设备。这些传感设备的出现为楼宇的安保和管理增添了更多的便利，为楼宇的安全监测提供的十分有用的数据。

其中视频监控的诞生初衷也是应对越来越复杂的安保工作，一直以来备受人们的青睐。但是，目前的大多数监控还是以传统的使用模式为主。由专人来负责查看已拍摄完成的场景信息，发现不法人员或者行为活动后再采取相关措施，或者发生相关纠纷或者意外情况时来调取监控视频进行取证。这虽然为事后取证提供的方便，但是不能及时的反馈信息，避免此恶劣事件的发生。随着科学技术的发展，目前人脸识别技术已经有相当的商业应用规模了，如果视频监控结合人脸识别技术，提前为人们预警非法人员的出现，这将很大程度上降低发生意外的概率。在不改变现有监控系统的安装和布局情况下，将监控摄像头作为人脸识别系统的输入，这将大大丰富与提升摄像头的作用，如果发现有不法人员出现，系统及时预警告知相关安保人员，提早进行预防，这将为楼宇的安全管理提供了非常有用的保证。

综合以上说明，现有监控网络结合人脸识别系统可以极大的提升监控系统的作用，并为楼宇安保工作提供有力的数据支持，方便大型楼宇的安全管理工作。因此，研究面向楼宇监控的人脸识别系统是非常具有实际意义的。

2.研究目标

面对各种各样的楼宇而言，不同的楼宇具有不同的特点，同样针对不同的特点需要不同的功能。常见的楼宇可以分为机关楼、商业楼（也称为写字楼）、居

民楼等。首先针对不同的楼宇特点说明楼宇监控的特点以及具备的功能。

机关楼一般只允许特定人员进入,其他不属于机关单位的人员是不允许进入的。因此系统需具备设置白名单功能,建立机关人脸库,并设置为白名单。对于机关内部人员可以任意活动,非机关人员出现在机关楼内时,系统将及时进行预警,并将非法进入的人员信息进行建库记录,根据多个相机拍摄到的信息系统能够绘制非法人员行踪轨迹,帮助机关单位进行查证。

商业楼一般包括商场和办公楼两大部分,是以底层为商场,高层为办公企业为主。对于商场而言是对所有人都开放的,而办公楼是具有非请勿入的特点。因此,在商场环境下,任何人员都可以随便出入,系统只进行区域人数统计,另外系统支持黑名单库建立,在此种场景下如果发现黑名单库中的人员出现时,将及时向相关部门预警,并将其人脸图像进行记录。通过挂载的多个相机的记录情况,系统能够自动绘制黑名单中人员的行踪轨迹图,以便调出来查看。而对于办公楼而言,一般不允许闲杂人员进入,但是又可以让外部人员出入谈论事情,一般是需要相关单位人员带领进入。考虑到一般办公楼是有高层专用电梯,闲杂人员一般不会到此,因此在这种场景下系统自动记录出现在视频区域的陌生人,并建立数据库,以便出现情况时的调查取证。

对于居民楼而言,出入口相对封闭,非小区内部人员不能随便进入,但是小区内的人员变动较大。比如:租房子的人员更换、亲朋好友的探访和家政服务人员等等。因此系统需具备快速增删改数据库的性能,方便发生人员流动时数据库的快速更新。在监控方面的应用类似于办公楼的功能,这里不再赘述。

为了兼容以上提到的楼宇功能,本文开发的面向楼宇监控的人脸识别系统需要具备以下功能:

- 1、系统支持挂载网络相机和智能相机作为输入;
- 2、系统能够批量与快速建立人脸信息数据库,并支持黑名单、白名单的设置;
- 3、对于视频中的在库人员能够自动检测和抓拍识别,并且实时显示;
- 4、支持已识别目标历史记录查看;
- 5、系统支持区域人数统计与行人追踪,并能够显示行踪轨迹;
- 6、系统支持黑名单预警。

论文将结合实际情况,设计并实现上面所提到的功能。

3.研究现状

视频监控的设计初衷是为了应对越来越复杂的安保工作,尤其是社会上发生

较大的纠纷或者遭受了恐怖袭击之后,将进一步刺激视频监控的安装与监控技术的发展。随着技术的发展和监控保有量的增加, 监控技术的发展也是相当迅速。而人脸识别技术的研究起源于上个世纪, 在本世纪初期迎来了高速发展。人脸识别包含两个大的问题, 一个是人脸的检测, 另外一个是人脸的识别。在视频监控中, 人脸识别相对更加复杂, 当目标通过视频监控的过程中, 会捕捉到多帧人脸图像。因此, 在此过程中又包含人脸跟踪的问题。人脸跟踪技术的发展也有 20 年时间^[1]。

随着技术的进步, 智能监控系统也具有了极大的发展。全世界各国也都在此方面投入了较多的物力、财力, 同样取得了不错的效果。美国的智能监控技术发展的最好, 目前已经具有比较成熟的系统, 并且一些知名企业和实验室也在不断的努力, 比如 Facebook、麻省理工学院计算机视觉实验室、加州大学伯克利分校的计算机视觉组^[3]。瑞典的 Axis 公司推出的智能视频组件, 具备十分优越的性能, 同时支持向用户开放接口, 支持二次开发。我国的监控技术虽然起步较晚, 但是在投入了大量人力物力之后发展的十分迅速。我国公安的“天眼”系统就是一个十分优越的系统, 目前已投入了使用^[1]。另外, 我国的车站检票、上班打卡和刷脸支付等功能已经投入了相当大的使用规模。国内高校也在这一块一直进行研究, 如清华大学、国防科技大学和华中科技大学等都开展了人脸检测与跟踪的研究^[2]。国内在计算机视觉领域比较出色的企业, 北京旷视科技推出的 Face++ 产品、虹软公司提供的人脸识别相关的解决方案具有不错的性能, 并向开发者开放了免费的 SDK 以供学习开发。百度、腾讯、阿里巴巴也在人脸识别领域取得了不俗的成绩。

3.1 人脸检测方法研究现状

人脸检测方法可以大致分为两种, 一种是基于先验知识的人脸检测方法, 另一种是基于统计模型的人脸检测方法^[4]。对于人脸而言, 先验特征有五官特征、几何特征、纹理和脸型边缘特征和肤色特征等。基于统计模型的人脸检测方法是依靠大量的人脸样本, 对样本进行分类后, 再通过机器学习的手段训练出能够判断是否是人脸的分类器, 用于区分人脸。

3.1.1 基于先验知识的方法

对于人脸的先验特征, 五官特征、几何特征、纹理和脸型边缘特征和肤色特征等, 这些都是相对固定的, 并且是有规可循的。因此基于先验知识就是通过这些特征来检测人脸。

Geng 等人基于图像局部特征提出了人脸检测算法^[5]。随后，杨光正等人提出了一种基于多分辨率分层结构的思想，分层结构逐渐精细，并且在第三层的精细检测中利用了对人眼、嘴唇和鼻子等的对应规则^[6]。Kanada 等人利用积分投影函数分析二值图像，根据特定变化点来确定特征点位置^[3]。基于先验知识的方法基本都是依靠人脸的一种或者多种先验特征实现的，由于比较简单并且相对固定，因此在比较复杂的环境背景中，人脸检测的效果并不理想。

3.1.2 基于统计模型的方法

基于统计模型的人脸检测方法也可以叫做基于外观的方法^[3]。由于人脸特征具有一致性，因此可以通过对大量的人脸样本图片进行分类，将人脸样本可以分为人脸区域和非人脸区域。通过对样本的分类结果进行统计分析，由此来训练人脸分类器。常见的人脸分类器有，线性子空间法^{[7] [8]}、人工神经网络法^[9]、支持向量机法^[10]和 adaboost 算法^{[11] [12]}等。这些方法都通过统计学知识分析训练了大量的样本，在复杂场景中的检测效果远优于基于先验知识，取得了非常不错的检测效果。但是，大量的样本分类与训练耗费了较长的时间。

当然，由于人脸检测在实际的应用中的环境更加复杂和不确定，因此，目前的人脸检测算法还没有一种能够适应全部的情况。人脸检测依然存在较多的挑战，也是我们后面努力的方向。

3.2 人脸跟踪方法研究现状

目标跟踪作为机器视觉中的热点问题，一直以来而备受关注。人脸跟踪问题也是属于目标跟踪的一个分支。首次提出人脸跟踪问题的论文是由 Azarbayejani 发表的^[17]。对于人脸的跟踪问题，是通过检测到人脸特征信息，利用算法在视频流中确定各个时刻人脸位置的过程。人脸跟踪主要由人脸特征表示、搜索方法和背景模型更新三个部分组成^[3]。人脸跟踪算法大致可以分为四类：基于模型跟踪^[13]、基于运动信息跟踪^[14]、基于人脸局部特征跟踪^[15]、基于神经网络跟踪^[16]。

人脸跟踪中的特征选择对跟踪人脸的准确度有很大影响。T. O.jala^[18]等人首次提出了 LBP (Local Binary Pattern) 特征。LBP 特征可以用于描述图像的局部纹理，具有旋转不变性与灰度不变性的特点^[3]，后来被引入到了人脸跟踪中来加以利用。Zdenek Kalal 等人^[19]提出的 TLD (Tracking-Leaming-Detection) 跟踪算法，将传统的跟踪算法与传统的检测算法合二为一，用以解决当被跟踪目标在被跟踪过程中发生的形变、部分遮挡、消失、重新出现等问题。与此同时，

通过一种改进的在线学习机制不断的更新目标模型以及相关配置参数,从而取得了更加稳定、更加可靠并且鲁棒性更高的跟踪效果。

人脸检测的主要关注点是检测的准确性,而人脸跟踪在保证人脸检测的准确性的同时又十分关注实时性。由于在视频监控中,人脸的运动是十分复杂与多变的,人脸的姿态变化、遮挡、表情变化还有消失等都为人脸跟踪带来了很大的难题。人脸跟踪本身又是一个非常困难过程,一般的实际使用场景中会将人脸检测的准确性与人脸跟踪的实时性进行折中处理,以满足实际的使用需求。

3.3 人脸识别方法研究现状

人脸识别可以分为人脸对比和人脸搜索两个具体任务。其中人脸对比是对给定的两个人脸进行对比,从而确定是否是同一个人,也就是所谓的“我是我”问题,属于一对一匹配任务。而对于人脸搜索问题而言,是将一个人脸拿到,再与库中的全部成员进行比较,从而确定此人是哪一个人,也就是所谓的“我是谁”问题,属于一对多匹配任务。对这两个问题进行抽象可知,归根结底的问题都是两张人脸的相似度衡量问题,所以人脸识别研究的是如何通过有效的手段准确并且高效的衡量人脸特征相似度的过程。

随着深度学习的广泛应用,其在人脸识别领域的优势也日益突显出来,同时人脸识别算法也得到了极大的发展。DeepFace^[20] 的实现使用了 3D 对齐技术,训练过程前两个卷积层共享卷积核,后三个卷积不共享卷积核,倒数第二层采用全连接层提取对应的人脸特征。DeepID 网络分为一代和二代两种,区别主要是损失函数的定义不同。DeepID^[21] 采用常见的损失函数 softmax 训练人脸识别网络,DeepID2^[22] 利用 identification 和 verification 信息联合监督,实现了更好的效果。随着网络结构的优化,人脸识别取得了更好的识别效果。再者,很多工作的重心转移到优化损失函数,以取得更好的识别效果。代表性的损失函数有 L-Softmax^[23]、A-Softmax^[24],其在几个公开数据集上都实现了更好的效果^[1]。谷歌提出的 FaceNet^[25] 网络结构,其网络结构更加灵活。并且利用了海量的人脸样本图片训练深层 CNN 网络,取得了很好的识别效果。

虽然目前的人脸识别算法在 LFW^[26]、YTF^[27] 等公开的测试集上性能十分优越,但是这些测试集的普遍质量较高,因此取得了十分不错的结果。而一旦应用在实际的应用场景中,识别性能会下降很多,主要原因是在实际应用场合,场景和环境相对复杂,并且采集到的图像质量相对较低,再者训练集的图片 and 实际场景差异很大。目前,亚洲人种的测试集相对较少,这在很大程度上限制了深度学习工程的应用。

4.技术路线与实施方案

4.1 技术路线

针对上面提到的楼宇中监控的特点，并提出了不同的楼宇监控需具备不同的功能，用以适应实际使用需求。另外再结合研究目标提到的功能，论文的主要研究内容分为以下几点：

一、系统的整体框架设计

- 1、系统人机交互界面与各功能模块的设计；
- 2、系统架构中的多线程设计，包括消息分发与接收、线程锁等；
- 3、网络通信的相关设计。

二、视频流中的行人检测与人脸识别算法研究

- 1、行人通过监控区域时，自动抓拍与最优帧提取算法的研究；
- 2、人脸区域检测与人脸识别的算法实现。

三、面向楼宇监控的人脸信息数据库的结构设计与管理

- 1、实现人脸信息数据库的创建，黑名单与白名单的设置；
- 2、研究人脸信息数据库在多线程系统中的增删查改策略，防止数据库信息被破坏。

四、面向楼宇监控的人脸识别系统的实现与联调测试

封装算法以及各个子模块，完成整个系统的搭建。挂载监控相机对整个系统进行逻辑与功能完整性测试，并测试系统性能。

针对上面提到的主要研究内容，首先查阅资料分析涉及到的领域以及知识。在获取监控视频时，系统通过相机接口或者网络 IP 加载相机视频流。获取到视频流后，将存在于视频监控区域的目标从监控场景中分离出来，拟采用 surendra 算法进行运动目标检测来检测行人^[28]，并抓取包含行人的图像。在此过程中，先对选取最优帧的思路进行研究，再实现具有行人轨迹预测的最优帧抓取方法，减少人脸检测与识别的次数，以此来减少系统的运算开销，提高系统实时性。在进行人脸检测和人脸识别之前，首先对图像进行预处理，有利于提高人脸识别的正确率。在背景繁杂监控视频中，人脸检测拟采用 Adaboost 算法，人脸识别拟采用 PCA 算法。在系统开发中，网络通信与算法均采用多线程开发，防止线程阻塞而导致的系统崩溃，另外对数据库操作时进行加锁，以保证数据库的原子性。

4.2 实施方案

通过对面向楼宇监控的人脸识别系统的深入分析和内容拆解，系统可以分为两个大的方向来实现，其一为监控视频中的人脸识别的实现以及识别结果和使用到的算法的优化，其二为面向楼宇监控的人脸识别系统的开发，主要包括 UI 界面设计、监控视频流接口设计、数据库结构设计与库管理、系统的整体框架设计等内容。制定了一下的实施方案：

- 1、整体方案设定，监控相机的部署与系统框架设计；
- 2、监控视频的获取，包括监控视频的接口、通信格式、编解码方式等；
- 3、在不同的视角、光线、距离下获取的行人视频，实现 surendra 算法进行运动目标检测，并验证检测结果；
- 4、研究抓取最优帧行人图片，实现并验证具有行人轨迹预测的最优帧抓取方法；
- 5、人脸检测与人脸识别的实现，测试耗时情况；
- 6、设计人脸信息数据库结构，并确定建立数据库的方法，确保数据库的快速录入的同时，支持黑白名单的设置；
- 7、交互界面设计，预警接口的设计，以及目标行人的轨迹绘制功能的实现；
- 8、系统的整体封装，包括多线程任务分配、人脸数据库管理、黑名单人脸预警、人机交互接口等。

参考文献

- [1] 张恒胜. 视频监控中人脸检测与识别的关键算法研究[D]. 电子科技大学, 2019.
- [2] 高俊艳. 基于视频流的实时人脸检测与跟踪[D]. 广东工业大学, 2019.
- [3] 赵曦敏. 视频中的人脸检测与跟踪算法研究[D]. 山东大学, 2018.
- [4] 刘鹏. 基于视频序列的目标人脸跟踪技术的研究[D]. 新疆大学, 2019.
- [5] Geng C, Jiang XD. Face Recognition Based on the Multi-scale Local Image Structures[J]. Pattern Recognition, 2011, 44(10): 2565-2575.
- [6] Huang TS, Yang GZ. Human Face Detection in A complex Background. Pattern Recognition, 1994, 27(1): 53-63.
- [7] Moghaddam B, Pentland A. Probabilistic visual recognition for object recognition[J]. Trans. IEEE Pattern Analysis and Machine Intelligence, 1997, 19(7): 696-710.
- [8] Sung K K , Poggio T. Example-Based Learning for View-Based Human Face Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1995, 20(1): 39-51.
- [9] Baluja S, Rowley H A, Kanade T. Neural Network Based Face Detection[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 1998, 20(1): 23-38.
- [10] Osuna E, Freund R, Girosi F. Training Support Vector Machines: an Application to Face Detection[C]. roceedings of the 1997 Conference on Computer Vision and Pattern Recognition. IEEE, 1997. 130-136.
- [11] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]. In: Anne Jacobs , Thomas Baldwin . Computer Vision and Pattern Recognition(CVPR). Proceedings of the 2001 IEEE Computer Society Conference on. Kauai, Hawaii: IEEE, 2001. 1: 511-518.
- [12] Hong L . Multiresolutional multiple-model target tracking[J]. IEEE Transactions on Aerospace and Electronic Systems, 1994, 30(2): 518-524.
- [13] Viola P, Jones M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154
- [14] Olson B T . Brill, \Moving Object Detection and Event Recognition Algorithms for Smart Cameras[J]. Proc Image Understanding Workshop, 1997.
- [15] Shi S , GAO, CHEN. Facial Feature Extraction Based on Facial Texture Distribution and Deformable Template[J]. Journal of Software, 2001.
- [16] Cui Y, Bo Z, Yang W, et al. End-to-End Visual Target Tracking in Multi-robot Systems Based on Deep Convolutional Neural Network[C]// IEEE International Conference on Computer Vision Workshop. 2018.
- [17] Azarbayejani A, Horowitz B, Pentland A. Recursive estimation of structure and motion using relative orientation constraints[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 1993: 294-299.
- [18] Ojalat, Pietikainen M, and Harwood D, Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions[C], Iapr International Conference on Pattern

- Recognitional, 2002:582-585.
- [19] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Computer Society, 2012, 34(7):1409-22.
 - [20] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.
 - [21] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1891-1898.
 - [22] Sun Y, Chen Y, Wang X, et al. Deep learning face representation by joint identification-verification[C]. Advances in neural information processing systems. 2014: 1988-1996.
 - [23] Liu W, Wen Y, Yu Z, et al. Large-Margin Softmax Loss for Convolutional Neural Networks[C]. ICML. 2016, 2(3): 7.
 - [24] Liu W, Wen Y, Yu Z, et al. Sphereface: Deep hypersphere embedding for face recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 212-220.
 - [25] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 815-823.
 - [26] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[C]. Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition. 2008.
 - [27] Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity[M]. IEEE, 2011.
 - [28] 朱雨佳. 基于 ARM 的人脸识别智能视频监控系统的设计与实现[D].哈尔滨理工大学, 2019.