

〈红外应用〉

基于近红外图像的嵌入式人员在岗检测系统

苏育挺，陈耀，吕卫

(天津大学 电气自动化与信息工程学院, 天津 300072)

摘要: 在岗检测是现代安防领域中视频分析的一个重要研究方向, 应用领域非常广泛。本文设计并实现了一种嵌入式人员在岗检测系统, 为了提高此嵌入式系统的运行速度, 提出了改进的人脸特征点检测方法; 并且为了提高系统的检测准确率, 建立了一个近红外人脸样本库。该系统通过近红外摄像头采集实时图像, 然后进行人脸特征点检测, 获取被检测人的面部信息。根据违规行为判断准则, 判断当前是否出现违规动作并且发出警报。实验结果表明: 在规定条件下, 系统的人脸特征点检测准确率达到了 95%, 针对两种异常情况的检测准确率也都超过了 94%, 具有良好的实时性能。

关键词: 近红外; 嵌入式; 在岗检测; 人脸特征点检测

中图分类号: TP391 文献标识码: A 文章编号: 1001-8891(2019)04-0377-06

Embedded Personnel On-the-job Detection System Based on Near-infrared Image

SU Yuting, CHEN Yao, LYU Wei

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: On-the-job detection is an important research direction for video analytics in the field of modern security, with a wide range of applications. This study designs and implements an embedded personnel on-the-job detection system. To improve the running speed of this embedded system, we proposed an improved face landmarks detection method, and to improve the accuracy of detection by the system, we established a near-infrared face database. This system initially collects real-time images through a near-infrared camera; subsequently, it performs face landmark detection to obtain the facial information of the detected person. According to predefined rules to identify violations, the system decides whether an illegal action has occurred and sends out alarm. The experimental results show that the accuracy of face landmark detection by the system reached 95% under the specified conditions, and the detection accuracy rate for two abnormal conditions exceeded 94%, both while maintaining a good real-time performance.

Key words: near-infrared, embedded, on-the-job detection, face landmarks detection

0 引言

在岗检测是指对视频画面中的特定人员进行监测。它是现代安防领域中视频分析的一个重要研究方向, 应用领域非常广泛。随着行为检测算法技术的发展, 现在已经可以使用计算机来执行在岗检测这一任务。当计算机识别到画面中特定人员出现异常情况时, 会直接发出警报。与传统的单纯依靠人力进行在

岗检测相比优势明显, 而且可以与近红外图像获取设备以及嵌入式设备结合使用, 实现全天候不间断值守同时又足够轻便, 应用潜力巨大。

在岗检测的核心步骤是行为检测, 对此国内外有诸多研究。例如, 文献[1]提出了一种基于关节点信息的人体行为识别新方法, 利用 Kinect 采集人体关节点坐标并构造 3 维空间向量, 然后计算结构向量之间的角度和向量模的比值, 进行人体姿态描述, 同时以一

收稿日期: 2018-07-20; 修订日期: 2018-10-25。

作者简介: 苏育挺 (1972-), 男, 博士, 教授, 主要从事数字视频编码, 数字视频处理, 信息隐藏与数字水印, 多媒体信息被动取证等方面的研究。E-mail: jxcy@tju.edu.cn。

基金项目: 国家自然科学基金 (61572356)。

段时间内连续的姿态序列作为行为表示特征量，最后计算测试行为模板与参考行为模板之间的相似度以实现行为识别。文献[2]提出了基于时间段网络的深度行为识别算法，该方法结合了稀疏时间采样策略和视频等级监督方法，可以使用整体的动作视频进行便捷和有效的学习。文献[3]提出了一种基于鲁棒且有效的视频表示的动作识别方法，通过显式相机运动估计来改善流行密集轨迹特征，使用 SURF (Speeded up robust features) 描述符^[4]和密集光流提取帧之间的特征点进行匹配。但是目前业内还没有一个成熟的针对嵌入式系统下的行为检测算法，大部分算法都需要性能较好的运行平台才能达到预期效果，而本文设计的嵌入式系统采用的硬件平台受性能所限，暂时无法流畅运行当下流行的机器学习算法，目前只能采用一些对硬件要求较低的传统检测方法。

由于在岗检测系统中的检测区域主要集中在人头部区域，因此行为检测可以细化到头部的行为检测。本嵌入式系统从人脸特征点 (Facial landmarks) 检测入手，通过特征点来分析当前人员的具体行为。目前可以使用一些公开算法来进行检测，比如 Dlib 库的人脸特征点检测^[5-6]、SeetaFace 人脸特征点检测^[7-8]以及剑桥大学的 CLM-framework 人脸跟踪器^[9]。依据嵌入式平台的性能水平，此系统采用了对硬件要求较低的 Dlib 人脸特征点检测方法。

为了满足全天候 24 h 不间断检测的要求，本系统通过近红外光 (Near infrared, NIR) 摄像头采集人脸图像信息。近红外光处于可见光和中红外之间，波长范围是 0.78~2.5 μm，属于不可见光。近红外成像在抑制环境光照影响的同时，也不易受到环境温度的影响^[10]。不会对被测人员的工作产生干扰，并能够使系统不受光照因素的影响而实现全天候工作。

本系统利用 Dlib 库对采集到的近红外图像进行人脸特征点检测，并为了提高此系统在嵌入式板卡中的运行速度，提出了人脸特征点检测的改进算法；此外还建立了一个近红外人脸样本库并训练了相应模型；依据设定好的报警阈值，对被测人员当前状态进行监测与报警；最后在模拟场景中对系统进行准确性和实时性检验。

1 系统结构

该人员在岗检测系统采用的嵌入式平台为恩智浦 (NXP) 公司生产的 iMX6Q 核心板^[11]。该核心板集成了 Cortex-A9 四核处理器，工作频率达到 1 GHz，可以较为流畅的运行此检测系统。

此系统可以分为 3 个模块：图像获取模块、行为检测模块与异常警报模块，总体框架如图 1 所示。

1.1 图像获取模块

图像获取模块采用了近红外摄像头，以便全天候获取被测人员的脸部实时图像，其分辨率为 1280×720，帧率设定为 25 帧/s，以保证能获取到实时图像。但为了保证画面中的人脸大小符合要求，对摄像头和人员间距离有硬性要求，目前距离 x 设定在 1 m 以内；并且摄像头与人眼的仰角 β 大小不得超过 60°，符合一般情况下的需求。摄像头及其安装位置示意图如图 2 所示。

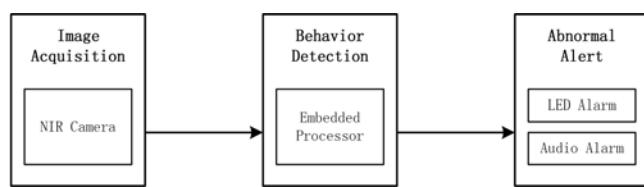
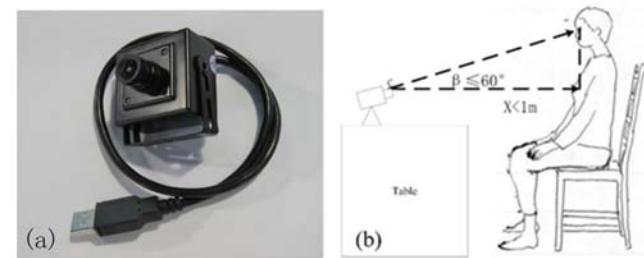


图 1 系统总体框架

Fig.1 General framework of the system



(a) Example of a camera (b) Example of installation location

图 2 摄像头及其安装位置示例

Fig.2 Example of a camera and its installation location

1.2 行为检测模块

行为检测模块基于 iMX6Q 核心板构成，负责进行图像预处理以及人脸特征点检测。当此系统检测到异常情况时，会把报警信号传输至后续模块。具体的异常判别标准将在行为检测流程部分详细介绍。

1.3 异常警报模块

异常警报模块采用 LED 灯闪烁警报和音频警报的组合报警方式。当发生异常情况时，LED 灯由绿色光变为红色光并且不停闪烁，同时音频报警频率为 3 s/次，此频率可根据需要更改。系统是 24 h 值守的，将会不断重复上述过程，从而实现系统的在岗监测与实时报警。

2 行为检测流程及算法介绍

2.1 行为检测流程

该系统核心部分为行为检测模块,它完成了对传输画面的人脸特征点检测和异常行为检测等功能。行为检测模块具体流程如图3所示。

在图3中,首先图像获取模块把采集到的视频帧依次传输给行为检测模块,然后利用训练得到的近红外人脸模型进行人脸特征点检测,若检测不到就发出警报,若检测正常就进入异常行为检测阶段。

目前系统设计要检测的异常行为包括抽烟和疲劳,且抽烟的违章等级更高,作为首先被检测的动作,如果出现抽烟现象持续时间达到1 s就发出警报;根据判定依据,若在一段时间内闭眼时间达到一定阈值就认定是疲劳状态并发出警报。如若一切正常就继续对下一帧进行检测,如此循环往复便实现了24 h值守。具体的检测算法将在下一部分介绍。

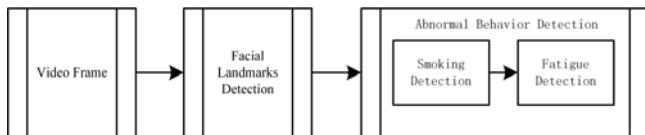


图3 行为检测流程

Fig.3 Process of behavior detection

2.2 算法介绍

当获得一帧图片后,Dlib的人脸特征点检测算法会估计一个大致的特征点位置,为此会先进行人脸检测,然后在人脸区域内将默认的特征点位置作为检测的初始位置(即图4中 $T=0$ 的情况),之后通过将均方误差作为损失函数进行迭代回归,得到最终的人脸特征点位置。通过计算出的68个特征点能够定位出人脸轮廓、眉毛、眼睛、鼻子和嘴巴等部分。某一帧图像 I 迭代的公式表示如(1)所示:

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \quad (1)$$

式中: $\hat{S}^{(t+1)}$ 表示第 $t+1$ 次迭代估计的位置; $\hat{S}^{(t)}$ 表示第 t 次迭代估计的位置; $r_t(\cdot, \cdot)$ 表示本次迭代的回归器。图4展示该估计过程的一个实例,其中 T 表示迭代次数。



(a) $T=0$ (b) $T=1$ (c) $T=2$ (d) Final position

图4 Dlib 特征点回归过程

Fig.4 The process of landmarks regression

Dlib 的人脸检测采用了方向梯度直方图(Histogram of oriented gradient, HOG)特征^[12],这是一种图像局部重叠区域的密集型描述符,通过计算局部区域的方向梯度直方图来构成特征。然而,由于Dlib自带特征点检测模型是基于自然光照下的图像所训练的,不太适用于该系统所需要的近红外光图像。因此本文需要训练近红外光下的人脸特征点检测模型,实验所用数据库以及模型将在下一章进行介绍。

2.2.1 抽烟检测

异常行为级别最高的是抽烟,烟头点燃时的中心温度高达700℃,在经过灰度化的近红外图像中看起来就是白色的亮点。经过特征点检测后,根据找到的嘴部位置可以框定一个比嘴略大的矩形框,系统将在这个范围内进行亮点检测。

首先对图像进行灰度化处理,可以得到每一个像素的灰度值,然后对图像进行二值化处理,这里将阈值设为253。如果出现像素面积达到一定值(为了排除环境干扰,这里设定为20)的亮点区域并且持续时间达到1 s,系统就会立即发出警报。二值化的判断过程如公式(2)所示:

$$dst(x, y) = \begin{cases} \text{maxval} & \text{if } \text{src}(x, y) > \text{thresh} \\ 0 & \text{if } \text{src}(x, y) \leq \text{thresh} \end{cases} \quad (2)$$

式中:(x, y)表示某点坐标, $dst(x, y)$ 表示某点最终的二值化结果; maxval 表示二值化最大值,在这里就是255; $\text{src}(x, y)$ 表示某点灰度值; thresh 表示二值化判断阈值。

亮点检测效果如图5所示。



图5 亮点检测效果

Fig.5 Bright spot detection effect

2.2.2 疲劳检测

如果没有检测到抽烟动作,会继续检测是否处于疲劳状态,在这里采用业界常用的PERCLOS(Percentage of eyeLid closure over the pupil over time)^[13]判断准则作为检测依据。PERCLOS准则最早由卡内基梅隆研究所提出,现已被广泛用于疲劳检

测当中,是一个非常成熟的检测准则。PERCLOS 准则通过计算单位时间内眼睛闭合状态的帧数 n 与单位时间内的总帧数 N 的比值得到。业内普遍认为当 PERCLOS 值超过 0.4 时,被检测人员处于疲劳状态。PERCLOS 准则的公式表达如(3)所示:

$$f = \frac{n}{N} \times 100\% \quad (3)$$

使用此准则有一个关键点在于闭眼状态的判定,以往业内都采用 P80 原则,即眼部闭合面积大于等于眼部面积的 80% 时,就判断为闭眼。但是由于本文采用了特征点检测方法,因此可以转化为对特征点的计算。

本文在此提出了一种新的判断方法,即将每只眼睛的横纵间距之比作为判断依据。如图 6(a)所示,由特征点模型即可得到眼睛特征点以及对应编号, x 和 y 分别为某个特征点的横纵坐标,在这里选定编号为 36、38、39、40 的特征点,眼睛的横纵间距之比如式(4)所示:

$$W/H = (x_{39}-x_{36})/(y_{40}-y_{38}) \quad (4)$$

利用这个特征就可以判断是否闭眼,经多次实验测量后,将此阈值设为 3.2。闭眼检测效果如图 6(b)所示。然后利用 PERCLOS 准则进行判断,当此值大于等于 0.4 时,即认定为疲劳状态。

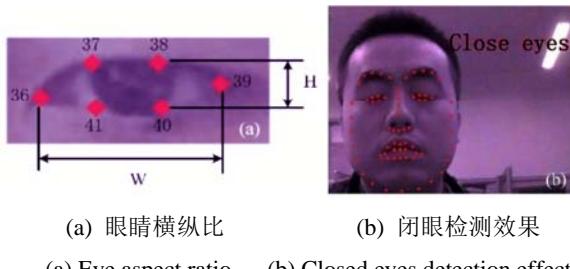


Fig.6 Detection of closed eyes

2.3 人脸特征点算法改进

在实验过程中发现人脸特征点检测速度较慢,帧率只有每秒 10 帧左右,达不到系统需要的实时性要求。从特征点定位过程来看,原来的 Dlib 算法会先进行人脸检测,而此人脸检测采用的是 HOG 特征,计算较为复杂,受限于板卡性能,对检测速度的影响会大大增加。并且初始位置的定位方式比较粗糙,需要进行的迭代次数会比较多。为了达到系统实时性要求,本文将对 Dlib 人脸特征点检测采取如下的两步方案进行改进:

第一步改进是将人脸特征点定位中的 HOG 人脸检测换成计算量小的 Haar 人脸检测,降低检测耗时。此人脸检测算法基于 Viola 等^[14]提出的 Adaboost 算法,采用了矩形特征来实现,也称为 Haar 特征。

第二步改进是除第一帧图像外,之后每一帧图像估计的大致特征点位置是上一帧图像的最终特征点位置。这样就能减少每一帧图像的迭代次数,提高了特征点定位的速度。这是因为本系统应用环境中被检测人一般进行小幅度的运动,而运动具有连续性,非剧烈运动情况下连续两帧图像之间的人脸位置必然很接近,在前一帧基础上进行回归运算会降低迭代次数,提高特征点定位速度。该过程在数学形式上如式(5)所示:

$$\hat{S}_{f+1} = S_f \quad (5)$$

式中: S_f 表示第 f 帧图像最终的特征点位置; \hat{S}_{f+1} 表示第 $f+1$ 帧图像的特征点初始估计位置。

3 实验结果分析

本文实验中采用的硬件平台是恩智浦公司的 iMX6Q 核心板,操作系统为 ARM 版本的 Ubuntu 12.04, 编程语言为 C++, 另外还利用 OpenCV 库^[15]进行一系列的图像处理。

3.1 实验数据库介绍及模型训练

为了建立一个近红外人脸样本库,本文采集了一百位人员的 10000 张姿态各异的近红外人脸图像作为正样本;另外从网上和现实生活中又采集了 20000 张左右的负样本图像。样本库示例图片如图 7 所示。构建好了样本库就可以进行模型训练,利用 Dlib 提供的训练函数,可以得到近红外光下的特征点检测模型。近红外光下的 68 个特征点检测结果如图 8 所示。



(a) 正样本示例;
(b) 负样本示例
(a) Sample pictures of positive samples; (b) Sample pictures of negative samples

图 7 人脸样本库示例图片

Fig.7 Sample pictures of face samples lib



图8 人脸特征点检测效果图

Fig.8 Effect picture of face landmarks detection

3.2 系统检测准确度

本文的检测准确率采用了国际流行的 FDDB (Face detection data set and benchmark) 人脸检测算法评价标准^[16], 若一次检测输出区域为 d_i 和 d_j , 对应的标准区域为 l_i 和 l_j , $\text{area}(\cdot)$ 表示区域面积, 则检测结果 S 可以用公式(6)评价:

$$S(d_i, d_j) = \frac{\text{area}(d_i) \cap \text{area}(l_j)}{\text{area}(d_i) \cup \text{area}(l_j)} \quad (6)$$

为了验证本文改进的人脸特征点检测方法确实提高了检测速度, 遂设计了一组对比实验。让改进算法(improved)与Dlib原有算法(original)对近红外人脸数据库上的所有正样本进行检测, 检测模型使用本文训练的近红外人脸特征点模型。对比结果如表1所示。

从表格中能看到对于同样的测试视频, 两种特征点检测方法的准确率变化不大, 新的特征点检测方法检测速度提升明显, 平均每帧检测耗时降低了47 ms。综合来看, 新方法明显优于Dlib原有的检测方法。

表1 两种特征点检测方法的对比结果

Table 1 Comparison between two methods of landmarks detection

Method	Accuracy/%	Average detection time per frame/ms
Improved	95.4	50
Original	95.3	97

为了验证训练得到的近红外人脸特征点检测模型(NIR)比Dlib自带的自然光检测模型(Natural)更加适用于此系统, 增加一组对比实验。使用改进算法, 统计两种模型对近红外人脸数据库上所有正样本的检测准确率, 近红外模型的检测结果就采用表1中改进算法的实验结果。最后得到的实验结果如表2所示。

从实验结果来看, 可以得到本文训练得到的近红外人脸特征点检测模型更加适用于此系统的结论。对于同一数据集, 新模型检测精度比改进前的自然光模型提高了2.8%, 而且两者平均检测速度几乎相同。

表2 两种人脸特征点检测模型对比结果

Table 2 Comparison results of two kinds of face feature point detection models

Model	Accuracy/%	Average detection time per frame/ms
NIR	95.4	50
Natural	92.6	51

为了获得系统对两种异常情况的检测准确率, 找了10位测试人员, 每人都模拟两种异常情况若干次, 并且由近红外摄像头实时采集图像。采用改进算法以及近红外模型, 最后得到的检测结果如表3所示。

3.3 系统的实时性

此系统基于嵌入式平台进行开发, 因此对检测实时性提出了要求, 一般来说为了保证检测和报警的及时, 要求检测帧率不得小于15帧/s。经过多次实验测量后, 该系统在算法改进后检测的帧率基本在20帧/s左右, 符合实时性要求。

表3 系统针对两种异常情况检测准确率

Table 3 The system detects accuracy for two abnormal conditions

Abnormal behavior	Accuracy/%
Smoking	94.3
Fatigue	96.8

4 结论

本文提出了嵌入式人员在岗检测原理流程及方法, 其中为了提高系统在嵌入式板卡上的运行速度, 改进了Dlib原有的人脸特征点检测方法, 兼顾了系统的检测精度和运行速度。当采用1280×720分辨率的近红外摄像头时, 人脸特征点检测准确率达到了95%, 针对两种异常情况的检测准确率也都超过了94%。而且系统的运行速度也符合实时性要求, 超过了15帧/s。该系统还采用了LED灯闪烁警报和音频警报的组合报警方式, 且可根据需要随时调整报警频率。下一阶段任务将尝试采用当下更为先进的机器学习算法进行模型训练与特征点检测, 并且对代码在CPU上进行Open CL加速, 以达到实时性要求。

参考文献:

[1] 田国会, 尹建芹, 韩旭, 等. 一种基于关节点信息的人体行为识别新方法[J]. 机器人, 2014, 36(3): 285-292.

TIAN G H, YI J Q, HAN X, et al. A novel human activity recognition method using joint points information[J]. Robot, 2014, 36(3): 285-292.

- [2] WANG L, XIONG Y, WANG Z, et al. Temporal segment networks: towards good practices for deep action recognition[C]//European Conference on Computer Vision, Springer, 2016: 20-36.
- [3] WANG H, Oneata D, Verbeek J, et al. A robust and efficient video representation for action recognition[J]. *International Journal of Computer Vision*, 2016, **119**(3): 219-238.
- [4] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features[C]//European conference on computer vision, Springer, 2006: 404-417.
- [5] Dlib. Dlib C++ Library[DB/OL]. Dlib, 2018. [2018-10-25]. <http://blog.dlib.net/>.
- [6] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1867-1874.
- [7] 山世光. Seeta Face Engine[DB/OL]. Git Hub, 2016[2018-10-25]. <https://github.com/seetaface/SeetaFace Engine>.
SHAN Shiguang. Seeta Face Engine[DB/OL]. Git Hub, 2016. [2018-10-25]. <https://github.com/seetaface/SeetaFace Engine>.
- [8] WU S, KAN M, HE Z, et al. Funnel-structured cascade for multi-view face detection with alignment-awareness[J]. *Neurocomputing*, 2017, **221**: 138-145.
- [9] Baltrušaitis T, Robinson P and Morency L P. 3D constrained local model for rigid and non-rigid facial tracking[C]//Computer Vision and Pattern Recognition (CVPR), 2012: 2610-2617.
- [10] 俞红兵, 乔亚. 基于核 DCV 算法的主动近红外人脸识别方法[J]. 红外技术, 2014, **36**(10): 807-811.
- [11] 恩智浦公司. i.MX6Q: i.MX 6Quad Processors- High-Performance, 3D Graphics, HD Video, Arm® Cortex®-A9 Core[DB/OL]. 2018[2018-10-25]. <https://www.nxp.com/cn/products/>.
NXP Corp. i.MX6Q: i.MX 6Quad Processors- High-Performance, 3D Graphics, HD Video, Arm® Cortex®-A9 Core[DB/OL]. 2018 [2018-10-25]. <https://www.nxp.com/cn/products/>.
- [12] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005, **1**: 886-893.
- [13] Dinges D F, Grace R. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance[J]/US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006, 1998.
- [14] Viola P, Jones M. Fast and robust classification using asymmetric adaboost and a detector cascade[C]//Advances in Neural Information Processing Systems, 2002: 1311-1318.
- [15] Bradski G, Kaehler A. *Learning OpenCV: Computer vision with the OpenCV library*[M]. Boston: O'Reilly Media, Inc., 2008.
- [16] Jain V, Learned-Miller E. Fddb: a benchmark for face detection in unconstrained settings[R]. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.