

# 基于巡逻小车的人脸遮挡异常事件实时检测<sup>①</sup>

张伟峰, 朱 明

(中国科学技术大学 信息科学技术学院, 合肥 230026)

**摘 要:** 近年来, 随着城市化进程的加快, 银行、政府、学校等场所的安全越来越成为人们关心的问题, 智能监控已成为目前的一个研究热点问题. 本文主要研究室内巡逻小车监控下, 面部异常遮挡问题. 本文首先对监控视频进行前景提取; 接着, 基于提取的前景, 进行肩部定位、头部区域椭圆拟合; 然后通过肤色检测判断人脸区域; 最后, 通过 Haar 检测器检测人脸区域的眼睛和嘴巴, 以此来判断是否存在异常遮挡行为. 实验结果表明, 本文提出的算法能够实时、有效的检测出面部异常遮挡问题.

**关键词:** 人脸遮挡; codebook 模型; 椭圆拟合; 肤色检测; Haar 检测器

引用格式: 张伟峰, 朱明. 基于巡逻小车的人脸遮挡异常事件实时检测. 计算机系统应用, 2017, 26(12): 175-180. <http://www.c-s-a.org.cn/1003-3254/6122.html>

## Real-Time Detection of Face Abnormal Occlusion Based on the Indoor Patrol Car

ZHANG Wei-Feng, ZHU Ming

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

**Abstract:** In recent years, with the acceleration of urbanization process, people become more concerned about security in public places like banks, official buildings and schools etc. Therefore, intelligent monitoring has become a hot topic in current research. This paper mainly studies the face abnormal occlusion events based on the indoor patrol car. Firstly, we extract the foreground of surveillance video. Then, we locate shoulders and use an ellipse to fit the area of head based on the foreground. And then, we determine the face area through the skin color rate. Lastly, we detect the eyes and mouth of the area through Haar detector to determine whether there are abnormal occlusion events. The experimental results show that the algorithm proposed can detect abnormal occlusion events effectively.

**Key words:** face occlusion; codebook model; ellipse fitting; skin color detect; Haar detector

近年来, 智能视频监控系统已经成为一个非常活跃的研究领域, 并且产生了很多实用的应用, 如人脸识别, 行为识别, 异常事件检测等. 这些都极大的减少了人工检查和验证的工作量, 并能及时有效的给出提示. 人脸面部遮挡检测是智能视频监控系统的基本组成部分, 随着银行、政府等部门等敏感区对安全的迫切需求, 人脸面部遮挡检测已经成为一个非常热门的研究领域.

## 1 概述

监控视频中面部遮挡检测分为两个步骤: 面部检

测和异常检测.

当前已经有很多比较成熟的面部检测方法, 主要分为五个类别: 基于知识的方法<sup>[1,2]</sup>、基于特征的方法<sup>[3,4]</sup>、模板匹配、基于模板匹配的方法<sup>[5]</sup>、基于外观的方法<sup>[6]</sup>. 每种方法都有各自的优缺点: 基于知识的方法计算成本较低但检测的准确率低; 基于特征的方法(如皮肤颜色、面部特征等)很容易实现良好的性能, 但不能适用于复杂的场景; 基于模板匹配的方法, 当模板数量很大时非常耗时, 达不到实时检测的需求; 基于

<sup>①</sup> 基金项目: 中科院先导项目 (XDA06011203)

收稿时间: 2017-03-17; 修改时间: 2017-04-10; 采用时间: 2017-04-17

外观的方法比较准确,但外观条件比较复杂.以上方法在面部检测方面都有各自的优点,但这些方法对有遮挡面部的检测效果都不是很好.

面部异常主要是面部遮挡,面部遮挡可分为上遮挡和下遮挡.上遮挡即遮挡鼻子以上区域,主要戴墨镜遮挡眼睛;下遮挡即遮挡鼻子及以下区域,主要是使用口罩遮挡嘴巴及附近区域.面部异常检测方法也有很多,主要分为三种:基于肤色的方法<sup>[7,8]</sup>、基于学习的方法<sup>[9,10]</sup>、基于组件的方法<sup>[11,12]</sup>.基于肤色的方法,如果对整帧图片进行肤色检测,如果有类皮肤颜色的地板或物品,就容易出现较大的误检,本文采用固定区域的肤色检测,可以最大程度的减少误检的概率;基于学习的方法比较复杂,并且需要较多的训练样本,监控摄像头下遮挡人脸的样本采集比较困难;基于组件的方法,通过检测眼睛、鼻子、嘴巴等面部组件,来实验面部遮挡的检测.本文就是通过检测眼睛来判断是否存在上遮挡,通过检测嘴巴来判断是否存在下遮挡.

## 2 算法设计与实现

本文的基本思路是:首先,提取每一帧视频图像的运动前景,然后根据提取的运动前景进行人头区域定位,人头区域定位分为肩部定位和椭圆拟合两个步骤,接着对定位的人头区域进行肤色检测以确定是否是人的正脸区域,最后对正脸区域进行眼睛和嘴巴检测来判断是否有异常遮挡行为,具体过程如下.

### 2.1 前景提取

前景提取是检测人脸遮挡异常的第一步,通过前景提取步骤可以把监控视频序列帧中的前景提取出来.前景提取方法有很多,主要分为以下三类:帧差法、光流法、背景减除法.

帧差法是最为常用的运动目标检测和分割方法之一,基本原理就是在图像序列相邻两帧或三帧间采用基于像素的时间差分通过闭值化来提取出图像中的运动区域.由于相邻两帧间的时间间隔非常短,用前一帧图像作为当前帧的背景模型具有较好的实时性,其背景不积累,且更新速度快、算法简单、计算量小.算法的不足在于对环境噪声较为敏感,阈值的选择相当关键,选择过低不足以抑制图像中的噪声,过高则忽略了图像中有用的变化.对于比较大的、颜色一致的运动目标,有可能在目标内部产生空洞,无法完整地提取运动目标.

光流法的主要任务就是计算光流场,即在适当的平滑性约束条件下,根据图像序列的时空梯度估算运动场,通过分析运动场的变化对运动目标和场景进行检测与分割.通常有基于全局光流场和特征点光流场两种方法.最经典的全局光流场计算方法是L-K(Lucas&Kanada)法和H-S(Horn&Schunck)法,特征点光流法通过特征匹配求特征点处的流速,具有计算量小、快速灵活的特点,但稀疏的光流场很难精确地提取运动目标的形状.总的来说,光流法不需要预先知道场景的任何信息,就能够检测到运动对象,可处理背景运动的情况,但噪声、多光源、阴影和遮挡等因素会对光流场分布的计算结果造成严重影响;而且光流法计算复杂,很难实现实时处理.本文要实现实时检测异常人脸遮挡,因此光流法也不能满足要求.

背景减除法是一种有效的运动对象检测算法,基本思想是利用背景的参数模型来近似背景图像的像素值,将当前帧与背景图像进行差分比较实现对运动区域的检测,其中区别较大的像素区域被认为是运动区域,而区别较小的像素区域被认为是背景区域.研究人员已提出了许多背景建模算法,如codebook模型<sup>[13]</sup>、VIBE算法<sup>[14]</sup>、高斯混合模型<sup>[15]</sup>等.

codebook模型为每个像素建立一个包含一个或多个码字.检测过程包含在颜色和亮度两方面检查当前图像与背景模型的差异.如果一个新像素在某个码字的色彩失真量小于检测阈值,或者新像素的亮度在该码字的亮度区间内,该像素被归为背景.反之,此新像素被归为前景.

codebook模型可以更准确地检测前景,并能有效地处理包含移动背景的场景或照明变化,本实验是以室内巡逻机器人为平台,codebook模型特别适合本实验.

### 2.2 人头区域定位

在准确进行前景区域提取以后,我们基于提取的前景区域来进行头部区域的定位.

头部区域定位主要分为两个步骤:肩部定位和肩部以上区域的椭圆拟合.

图1是头部和肩部的示意图,其中头部区域是近似对称的椭圆形.如图中横线所示,肩部存在一个长度急剧变化的区域,我们就可以根据这个长度急剧变化的特征来定位肩部,并对肩部以上区域进行椭圆拟合,进而精确定位出头部区域.



图1 人头肩部示意图

### 2.2.1 肩部定位

根据上面所说的基本原理,我们做出如下定义: $f_{left}$ 和 $f_{right}$ 分别代表视频帧前景区域的最左边和最右边, $s_{bottom}$ 和 $s_{top}$ 分别表示头肩部区域的底部和上部, $height$ 表示头肩部进行统计的单位区域行高, $gap$ 表示当前行与上一行的差值, $m\_gap$ 表示最大的差值。

$$value = \sum_{i=s_{bottom}}^{s_{bottom}+height} \sum_{j=f_{left}}^{f_{right}} f(i, j) \quad (1)$$

式(1)表示头肩部单位行高前景像素个数的和。

#### 算法1. 肩部定位

输入: 运动前景图

输出: 肩部位置

1. 根据提取的运动前景进行初始化  $f_{left}$ 、 $f_{right}$ 、 $s_{bottom}$ 、 $s_{top}$ 、 $height = 5$ 、 $m\_gap = 0$ 、 $position = 0$ 、 $position\_value = 0$ 。
2. 计算底部小区域的值  $pre\_value = value$ 。
3.  $s_{bottom} = s_{bottom} - height$ , 如果  $s_{bottom} < s_{top}$ , 转到6。
4. 计算从底部小区域往上的小区域  $value$  值。
5. 相邻两小区域差:  $gap = pre\_value - value$ ,  $pre\_value$  赋值为:  $pre\_value = value$ 。如果  $gap < m\_gap$ , 则  $m\_gap = gap$ ; 最大差值位置  $position = s_{bottom} + height / 2$ , 最大差值处单位行高值:  $position\_value = value$ 。转到3;
6. 如果  $m\_gap / position\_value < threshold$ , 定位出肩部的位置即为  $position$  的最后取值, 否则就说明不是肩部位置。

### 2.2.2 椭圆拟合

根据上节我们定位出了肩部的的位置,可以得到头部的大致位置,即为肩部以上的区域。由于头部是近似对称的椭圆形,因此我们可以近似用矩形框框出头部区域。下面我们对框出的区域进行椭圆拟合。

平面上任意位置的椭圆方程可以表示为:

$$x^2 + Axy + By^2 + Cx + Dy + E = 0 \quad (2)$$

随机选取面部轮廓点  $P_i(x_i, y_i), i = 1, 2, \dots, N$  为要拟合椭圆上的  $N(N \geq 5)$  个点, 根据最小二乘原理拟合目标函数为:

$$f(A, B, C, D, E) = \sum_{i=1}^N (x_i^2 + Ax_i y_i + By_i^2 + Cx_i + Dy_i + E) \quad (3)$$

使式(3)达到最小值,可求出五个参数的值,进而求出了拟合的椭圆。

### 2.3 人脸区域判断

通过上述步骤,我们定位出了人头的区域,但定位出的人头区域有可能是人头的后脑部分,如果不加区分,很显然我们无法进行下面的人脸遮挡判断。因此我们需要通过肤色检测技术辨别出定位的区域是否是人脸区域。

肤色检测的原理是根据皮肤的固有颜色,在图像中选取相对应的颜色范围作为皮肤颜色。肤色作为人的体表显著特征之一,尽管人的肤色因为人种的不同有差异,呈现出不同的颜色,但是在排除了亮度和视觉环境等对肤色的影响后,皮肤的色调基本一致,这就为利用颜色信息来做肤色分割提供了理论上的依据。

在肤色识别中,常用的颜色空间为 YCbCr 颜色空间。在 YcbCr 颜色空间中, Y 代表亮度, Cb 和 Cr 分别代表蓝色分量和红色分量,两者合称为色彩分量。YcbCr 颜色空间具有将色度与亮度分离的特点,在 YcbCr 色彩空间中,肤色的聚类特性比较好,而且是两维独立分布,能够比较好地限制肤色的分布区域,并且受人种的影响不大。对比 RGB 颜色空间和 YcbCr 颜色空间,当光强发生变化时,RGB 颜色空间中 (R, G, B) 会同时发生变化,而 YcbCr 颜色空间中受光强相对独立,色彩分量受光强度影响不大,综上所述, YcbCr 颜色空间更适合用于肤色识别。因此,肤色检测的第一步就是把图像从 RGB 颜色空间转到 YcbCr 颜色空间。

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.564 & 0.098 \\ 0.148 & 0.291 & 0.439 \\ 0.439 & 0.368 & 0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (4)$$

由于肤色在 YcbCr 空间受亮度信息的影响较小,因此直接考虑 YcbCr 空间的 CbCr 分量,映射为两维独立分布的 CbCr 空间。在 CbCr 空间下,满足式(5)即为肤色区域,否则即为非肤色区域,由此形成二值图像。

$$\begin{cases} 77 < C_b < 127 \\ 133 < C_r < 173 \end{cases} \quad (5)$$



通过对 2.2 节定位出的人头区域做肤色检测,得到的肤色区域大于整个人头区域的 1/4 即可判断 2.2 节的区域为人脸区域。

#### 2.4 异常人脸遮挡判断

经过研究,室内人脸异常遮挡主要行为有戴墨镜、戴口罩、压低帽檐等,这几种遮挡人脸的行为也就是遮挡了眼睛或嘴巴。因此,我们可以根据这个规律,对 2.3 节判断的人脸区域进行眼睛和嘴巴的检测,由于本文实验平台是室内巡逻小车,因此可以较近距离的拍到行人,这样我们定位的人头区域就可以达到识别眼睛和嘴巴的大小。

根据上面分析,人脸遮挡主要分为上遮挡和下遮挡,上遮挡即通过墨镜等遮住眼睛,下遮挡主要通过口罩等遮住嘴巴,因此我们可以通过 Haar 检测器检测 2.3 节定位的人脸区域的嘴巴和眼睛,只有嘴巴和眼睛都检测到说明人脸没有遮挡,否则人脸存在异常遮挡。

若要用 Haar 检测器检测嘴巴和眼睛,首要就是要提取 Haar 特征, Haar 特征分为:边缘特征、线性特征、中心特征和对角线特征。这些特征组合成特征模板,特征模板内有白色和黑色两种矩形,并定义该模板的特征值为白色矩形像素和减去黑色矩形像素和。 Haar 特征值反映了图像的灰度变化情况。脸部的一些特征可以通过矩形特征简单的描述,如:眼睛要比脸颊颜色要深,鼻梁两侧比鼻梁颜色要深,嘴巴比周围颜色要深等。但矩形特征只对一些简单的图形结构,如边缘、线段较敏感,所以只能描述特定走向(水平、垂直、对角)的结构。在提取出 Haar 特征以后,我们利用积分图来进行计算,并把计算结果通过 AdaBoost 算法来训练,进而实现一个 Haar 检测器。

本文由于在固定人脸区域内进行眼睛和嘴巴的识别,我们利用 OpenCV 自带的眼睛和嘴巴检测器就可以达到需要的检测效果。

OpenCV 的训练文件在源码目录的 data 文件夹下,里面包含 haarcascades、hogcascades、lbcascades,在 haarcascades 文件下包含大量的针对不同目标的训练文件,如图 2 所示。

通过载入 haarcascade\_mcs\_mouth.xml 来识别嘴巴,通过载入 haarcascade\_eye\_tree\_eyeglasses.xml 来识别眼睛。正常的无遮挡人脸应该都可以在检测到眼睛和嘴巴,但是在实际小车巡逻场景中,由于小车的动态移动,必然会导致拍到的视频可能存在模糊,进而出现

误检现象。因此,我们规定连续出现 3 帧检测不到嘴巴,或者连续 3 帧检测不到眼睛,就说明出现了异常人脸遮挡行为。

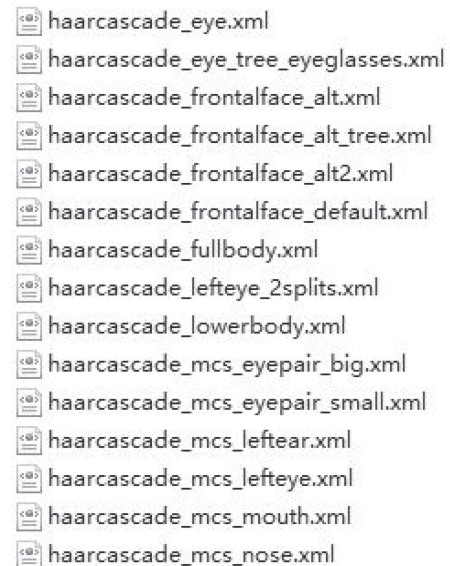


图 2 OpenCV 器官识别文件示意图

### 3 实验与分析

本实验的硬件平台为搭载摄像头的室内巡逻小车和处理实时视频流的 PC(Intel i5 3.3GHz + 6G 内存),软件平台为 VS2013+OpenCV。

#### 3.1 人脸遮挡异常检测实验

根据我们上述提出的算法流程,我们在实验平台上对每一步流程分别进行实验,并验证算法的有效性。



图 3 自制室内巡逻小车

### 3.1.1 前景提取

根据我们的巡逻小车的视频提取出前景,效果如图4.



图4 前景提取效果图

### 3.1.2 人头区域定位

根据我们的算法,我们先定位肩部,即途中蓝线;然后根据肩部框出头部区域,即图5中矩形框;然后进行椭圆拟合得到脸部区域,即图5中椭圆区域。

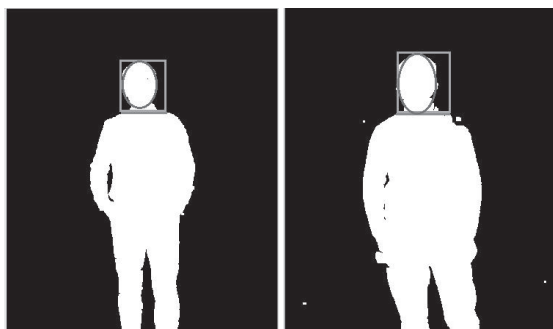


图5 人头区域确定

### 3.1.3 人脸区域判断

在图6中进行肤色检测,并把检测到为肤色的区域值除整个椭圆区域值,如果大于阈值说明是正面人脸。如图6检测结果都为正面人脸。

### 3.1.4 人脸异常遮挡

对上述区域进行 Haar 检测。在本实验中,如果被墨镜遮,则检测不到眼睛,就会发出面部遮挡的报警;

如果口罩遮挡住嘴巴,也会同样发出面部遮挡的报警。



图6 肤色检测

## 3.2 实验结果与分析

本实验数据为两段自录视频。一段是口罩遮挡住人脸,并把口罩取下的对比视频,总共750帧,其中有遮挡的帧数100帧,正常检测出93帧;另一段是墨镜遮挡住眼镜,并把眼镜取下的对比视频,总共1036帧,其中有遮挡的帧数224帧,正常检测出210帧。实验达到了很高的准确率。

## 4 结语

本文提出的面部异常遮挡检测方法,能够较为准确地定位脸部区域,并能实时检测出是否有面部异常遮挡行为。实验结果表明,本文提出的算法不仅满足了实时检测(每帧毫秒级),在准确率上也达到了传统经典算法的水平。但是本文对多人场合无法很好的完成检测。

下一步的工作,将侧重算法的改进,提出能够检测出人群异常面部遮挡行为,并搜集数据训练异常遮挡模型,达到精确检测效果。

## 参考文献

- 1 Yang GZ, Huang TS. Human face detection in a complex background. *Pattern Recognition*, 1994, 27(1): 53-63. [doi: 10.1016/0031-3203(94)90017-5]
- 2 Kotropoulos C, Pitas I. Rule-based face detection in frontal views. *Proc. of 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Munich, Germany. 1997. 2537-2540.
- 3 Yang L, Waibel A. A real-time face tracker. *Proc. of the 3rd IEEE Workshop on Applications of Computer Vision*. Sarasota, FL, USA. 1996. 142-147.
- 4 Yow KC, Cipolla R. Feature-based human face detection. *Image and Vision Computing*, 1997, 15(9): 713-735. [doi: 10.1016/S0262-8856(97)00003-6]
- 5 Craw I, Tock D, Bennett A. Finding face features. *Proc. of the 2nd European Conference on Computer Vision*. London, UK. 1992. 92-96.

- 6 Sung KK, Poggio T. Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, 20(1): 39–51. [doi: [10.1109/34.655648](https://doi.org/10.1109/34.655648)]
- 7 Lin DT, Liu MJ. Face occlusion detection for automated teller machine surveillance. *Proc. of the First Pacific Rim Conference on Advances in Image and Video Technology*. Hsinchu, Taiwan, China. 2006. 641–651.
- 8 Kim G, Suhr JK, Jung HG, *et al.* Face occlusion detection by using b-spline active contour and skin color information. *Proc. of 2010 the 11th International Conference on Control Automation Robotics & Vision*. Singapore. 2010. 627–632.
- 9 Priya GN, Banu RSDW. Detection of occluded face image using mean based weight matrix and support vector machine. *Journal of Computer Science*, 2012, 8(7): 1184–1190. [doi: [10.3844/jcssp.2012.1184.1190](https://doi.org/10.3844/jcssp.2012.1184.1190)]
- 10 Kim J, Sung Y, Yoon SM, *et al.* A new video surveillance system employing occluded face detection. *Proc. of the 18th International Conference on Innovations in Applied Artificial Intelligence*. Bari, Italy. 2005. 65–68.
- 11 Suhr JK, Eum S, Jung HG, *et al.* Recognizability assessment of facial images for automated teller machine applications. *Pattern Recognition*, 2012, 45(5): 1899–1914. [doi: [10.1016/j.patcog.2011.11.014](https://doi.org/10.1016/j.patcog.2011.11.014)]
- 12 Eum S, Suhr JK, Kim J. Face recognizability evaluation for atm applications with exceptional occlusion handling. *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Colorado Springs, CO, USA. 2011. 82–89.
- 13 Kim K, Chalidabhongse TH, Harwood D, *et al.* Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 2005, 11(3): 172–185. [doi: [10.1016/j.rti.2004.12.004](https://doi.org/10.1016/j.rti.2004.12.004)]
- 14 Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. on Image Processing*, 2011, 20(6): 1709–1724. [doi: [10.1109/TIP.2010.2101613](https://doi.org/10.1109/TIP.2010.2101613)]
- 15 Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. *Proc. of the 17th International Conference on Pattern Recognition*. Cambridge, UK. 2004. 28–31.