摘要

随着我国经济社会的快速发展以及生产生活方式的巨大转变，使越来越多的人成为久坐一族，因此长时间的伏案工作已经成为大部分人在工作和学习中的一种常态化办公方式。然而因坐姿不正确或坐姿不标准而产生的健康问题也越来越严重，诸如近视、颈椎病和腰椎病等，且年龄不断趋于年轻化。

由此，本文研究如何通过图像处理方法对用户不规范坐姿进行识别和提醒，从而有效地避免不良坐姿给身体带来的危害。本文的主要工作如下：

第一、通过对坐姿形态的分析，阐述了在日常生活中保持良好坐姿的必要性，并对坐姿行为进行了分类。

第二、对坐姿检测过程中涉及到相关检测识别算法进行了详细地阐述，介绍了目标提取、肤色特征提取、SURF 特征提取等算法的适应情况和优缺点。

第三、设计与实现了基于 OpenPose 的坐姿评估系统，对不良坐姿进行识别，并进行提示与规劝，该系统由数据采集，数据预处理，坐姿评估和可视化数据分析四个模块组成。

第四、对坐姿评估系统进行精度和稳定性测试，设计了两种试验方案，采集了大量数据进行分析，验证系统能否在保持高准确度的同时，仍然具有较强的稳定性。

经实验分析表明，该系统得出在高强度作业下，该系统能在保持 95%以上的准确度的同时，具有较强的稳定性，能够对用户坐姿进行有效的识别检测，具有较好的应用前景。

目录

第一章 绪论

机器视觉技术作为计算机领域的重要分支，其主导的人体运动行为分析始终是模式识别及计算机视觉方向的研究热点。目前，运动分析在智能检测领域有着广泛的应用。通过对人体行为进行建模，可以为人体行为检测提供数据，这有助于通过设计和使用检测系统进行人体行为方面的分析，对人体不良行为进行识别，为指导和纠正错误姿势提供数据支持。

1.1 选题的背景和意义

近年来久坐伏案型人群占人口比例逐年增加，久坐时人们很难一直维持正确坐姿，如果长期处于不良坐姿，人体眼部、颈椎、脊椎、腰椎等部位出现损伤引发病变的几率将大大提高，对于青少年，不正确的书写坐姿会影响其身体正常发育，而且使他们容易产生疲劳，更有可能引发近视等疾病。

1. 不良坐姿引发近视

2020 年 9 月至 12 月国家卫健委全面开展了近视专项调查，覆盖了全国 8604 所学校，共筛查 246.7 万名学生。调查显示，2020 年我国儿童青少年总体近视率 52.7%，高中生近视率更是高达 80.5%！世界卫生组织的另一组调查数据显示，中国青少年近视率为 70%，而美国的中小学生近视率仅为 10%。（数据引自《光明日报》）这样算下来，美国孩子的近视率，仅为中国孩子的 1/7，目前，在我国近视群体中，约四分之三的新发病例都发生在 11~16 周岁之间，据专家研究表明：不良坐姿会导致视距过近，长此以往，使眼睛睫状肌失去自我调节能力，造成近视。

2. 不良坐姿与颈椎疾病的关系

近年来，颈椎疾病的发病人群越来越趋向年轻化，越来越多的青少年也患有颈椎疾病。在我国某城市，有十分之一左右的中小学生患有颈椎疾病，而在重点中学患病率甚至更为严重，根据卫生健康委员会 2020 年的研究数据，我国中小学生患脊柱侧弯人数已经超过 500 万，并且还在以每年 30 万左右的速度递增。长期的不良坐姿是引发颈椎疾病的主要病因之一。

2. 错误坐姿与脊椎畸形关系

脊椎畸形主要原因是脊椎发生异常形变，导致患者驼背、双肩和姿势不对称等不良症状，在我国青少年人群中有较高的发病率,严重影响了他们身体正常生长发育,给他们造成心理上及生理上的危害。少年儿童常常忽视了在书写过程中养成正确的坐姿习惯。不规范的书写坐姿会使脊椎长时间处于不正常的生理曲度，长此以往就可能发生脊椎畸形。

从以上分析可知，良好的坐姿与我们的学习和工作息息相关，对于处于发育期的青少年，及时地纠正和保持正确的坐姿对孩子的健康成长非常重要。目前，预防方法主要靠老师、家长提醒，身体状况出现问题后通过物理类矫正器矫正坐姿。而长时间地佩戴物理类矫正器会使他们产生厌烦情绪。因此，本文研究如何通过图像处理方法对用户不规范坐姿进行识别和提醒，从而有效地避免不良坐姿给身体带来的危害。

1.2 国内外研究现状

近几十年来，人体姿态检测识别作为模式识别、机器视觉、图像处理以及传感器技术等诸多学科的下属分支一直是该领域内的热门研究方向之一，人体动作姿态的捕捉不仅具有高度的复杂性而且还具有灵活多变的特点。在目前关于人体姿态检测识别的研究中，无论是国内的研究人员还是国外的研究机构对姿态的检测识别工作主要是通过以下两种途径来实现：一种是基于机器视觉图像处理的非接触式识别方法，其侧重于在某一种特定场合下对动态视频流和静态视频可视化实时识别方面，主要功能是对人体姿态行为进行分析、归类和标记。该方法具备识别速度快，小范围内识别效率高的优点，另一种是基于传感器技术的接触式识别方法，使用传感器采集到人体动作姿势行为的数据信息，结合其本身自有的特性以及人体动作姿态行为的特征参数予以分析并建立相关模型，最后实现人体动作姿态行为特征的分类识别。

1.2.1 国外研究现状

国外的优秀学者对于坐姿识别的研究起步比较早，主要是从两个方面来研究：一方面是从头部中心轴、肩膀边缘所在线与水平方向所成角度的关系

来判断坐姿；另一方面是利用人体姿势轮廓特征和傅里叶变换系数作为分类特征，模糊神经网络分类识别人体姿势。

Alejandro Jaimes 等人在 2005 年提出一种人体有害坐姿报警方案。首先使用背景差分法对图像进行用户轮廓提取，然后对差分图进行水平方向上地投影，获得投影的最低点。通过最低点作一条水平分界线将帧差图中人体头部和躯体划分，并利用头部中心轴和双肩边缘轮廓所在线与分界线之间的角度关系来判断坐姿。此方法鲁棒性较差，容易受光照变化影响，且只是从人体轮廓的几何关系的角度判断坐姿，不能全面分析坐姿，识别率偏低。在 2007 年利用模糊神经网络检测用户由突发事件所发生的跌倒。该方法首先利用帧差法进行人体轮廓提取，将轮廓的长宽信息作为一类分类信息。其次，将轮廓进行水平和垂直方向上地投影，并对投影进行 傅里叶变换得到傅里叶变换系数，将其作为另一类特征。最后，使用模糊神经网络对姿势进行识别。该方法对跌倒检测有较好的效果，但主要是从站立、弯曲、坐及卧四个角度检测跌倒情况，没有对人体坐姿进行具体分析和判断。

由 Haritaolu 等人研制出一种实时监控系统—W4 系统，该系统不仅可以对人体定位及分割，而且可对站、坐、蹲及趟等人体姿势进行识别。此外，Guo 等人[12]在利用运动目标的投影信息的基础上,结合了 PCA 算法,对站立、躺、坐、行走等九种不同姿势进行识别；Cucchiara 等人[13]则是利用贝叶斯分类器,对蜷缩、坐、站立和躺四种差别明显的姿势进行分类。但这些方法都不针对坐姿进行具体分析和识别。

1.2.2 国内研究现状

相比国外而言，国内的研究者主要根据视频中人脸所占区域大小和所在空间 3 位置关系来进行研究。此外，还有从肤色特征或基于对称性定位肩膀的角度识别坐姿。LanMu 等人在 2010 年提出了一种方法。该方法根据视频中人脸所占区域大小和所在空间位置关系判别坐姿。在获得一个良好坐姿的标准图像后，通过使用 hausdorff 距离模板匹配检测和定位出后续视频序列中的人脸位置，并求解出人脸所占区域大小。用其与标准图像比较识别坐姿。若与标准图相比，待识图像中的人脸区域面积越大，则反映出人与显示

器相隔越近；若待识图像中人脸的位置比标准图像偏右，则表明人体左倾，反之亦反。此方法只从人脸所占区域大小和位置关系角度进行识别坐姿，鲁棒性不强。韩晓明等人在 2009 年以用户标准坐姿时的头顶位置为参考点来识别坐姿。该方法先用背景差分法获取用户标准坐姿轮廓，得到头部所在区域，并以此作为标准。若头顶位置高于参考区域的上界，则说明用户距离显示器较近，反之亦反；若头顶位置较标准区域偏左，则说明用户右倾，反之也亦反。这种方法具有局限性，当环境和用户改变时，设置也要随之改变。武松林、崔荣一等人在 2010 年利用肤色在 YCbCr 空间具有聚簇性提取坐姿肤色特征作为分类特征，然后使用 PCA 作为分类器对坐姿识别。该方法有针对性对坐姿分类进行了研究，但没有判断坐姿正确与否。王春阳 2013 年，通过分别对头部和肩膀的定位，并利用两者空间位置关系识别坐姿。首先通过肤色检测到人脸，通过对眼睛和嘴的识别进一步定位头部位置。然后利用背景相减法提取用户轮廓特征，并针对肩膀定位设计出矩形检测法检测肩膀位置，从而确定坐姿。

第二章 坐姿分析和坐姿目标提取

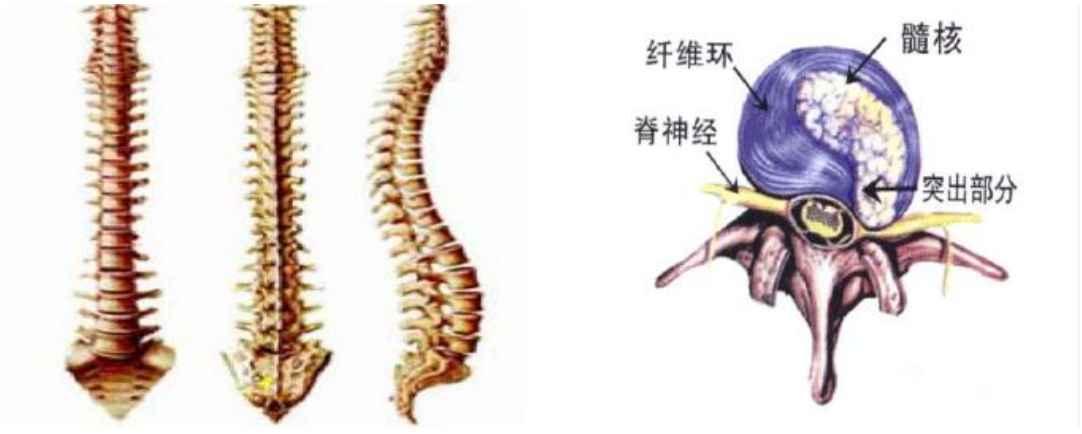
本文通过对坐姿形态的分析，阐述了在日常生活中保持良好坐姿的必要性，并对坐姿行为进行了分类。此外，介绍了常用的运动目标提取方法，用于视频图像中人体坐姿区域的提取，对之后的识别效果起着至关重要的作用。

2.1 坐姿分析

坐姿和脊椎生理曲度及眼距息息相关，不规范的坐姿导致脊椎生理曲度异常和眼距过近，引发脊椎及近视等疾病。

2.1.1 坐姿和脊椎生理曲度

脊椎是由 24 块椎骨构成，上承托颅骨，下联 髋骨，中附 肋骨，是人的“顶梁柱”，支持着躯干和保护人体内脏。图 2.1(a)是从前、后和侧面三个方向观察脊椎的结构形状图。从前后两个方向观察，脊椎 都是一条笔直的直线。但两者不同的是：前观时，脊椎从上到下逐渐加宽；而后观时，脊椎椎骨棘突连贯成纵嵴。当从侧面观察时，脊椎呈现出四个不同弧度的生理弯曲。这些生理曲度不但能够增加脊椎的弹性，而且在一定程度上减震，保护脊椎免收伤害。但长时间不正确的坐姿，会使脊椎异常弯曲，引发驼背和颈椎炎等疾病。



(a)脊椎结构图　　(b)椎间盘结构图

图 2-1 脊椎和椎间盘结构图

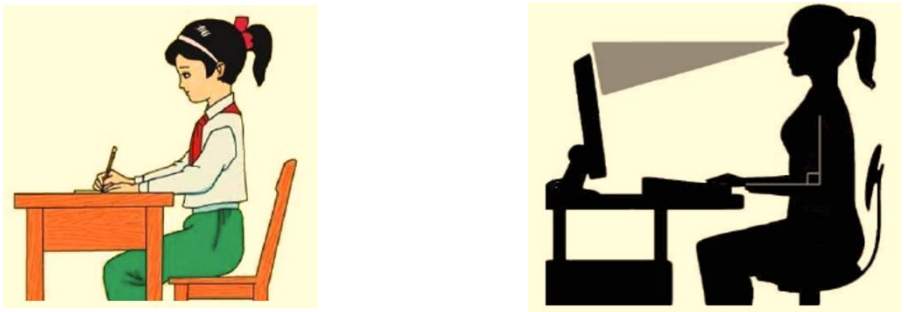
如图 2.1(b)所示，椎间盘是位于脊椎椎体之间的软骨组织，由上下软骨板、

纤维环及髓核三部分组成，构成一个有机的整体。其中软骨板起连接作用，上下衔接着椎体，纤维环充当着容器，包裹着髓核，而髓核则缓冲脊椎所受的压力。从图 2.1(b)可知，纤维环只有后侧较薄，当椎间盘受到上下两侧的压力后，容易迫使髓核往后侧方向突出，从而压迫脊神经，引发椎间盘突出疾病。

当坐姿正确时，人体脊椎会维持正常的生理曲度，让各椎间盘和周围肌肉组织承受最适合的压力并处于最佳状态，而长时间的不良坐姿，则会导致脊椎异常形变，造成驼背和颈椎不适合等症状。

2.1.2 正确坐姿的定义

保持良好的姿势不仅有利于身体健康，还能让人们保持注意力集中，提高工作和学习的效率。无论是读写姿势，还是使用电脑的坐姿，都必须做到“头正、身直、臂开、腿平”四部曲。头正，要求头部端正，不能左右倾斜和前后摇晃；身直，做到胸部张开，背部和腰部自然伸直，保持脊柱与臀部成一条直线；臂开，就是两个肩膀齐平，手臂自然张开放在桌上，有放松肩部肌肉的作用；腿平，即大腿保持水平，双腿自然弯曲下垂，脚掌以与肩膀相同的宽度平放在地上。



(a)正确的读写姿势　　(b)使用电脑时的正确坐姿

图 2-2 读写和使用电脑时的正确坐姿

正确的读写姿势如图 2-2(a)所示，根据教育部发布的相关文件[4]，其中提到需要家庭、学校等各方面的共同努力，引导青少年儿童养成良好的用眼读书习惯，要做到“三个一”：眼睛距离书本约为一尺，使睫状肌不需要过度收缩而让眼睛得不到休息，从而保护视力；身体距离课桌约为一拳，身体可以略微前倾，保证胸前不受压迫；手指距离笔尖约为一寸，这个距离刚好可以减轻手部疲劳，还可

以在书写时保持一个舒服的状态并提高书写质量。如图 2-3(b)所示是使用电脑时的正确坐姿，需要做到“三个调节”：调节椅背的位置，也可以根据需要使用靠垫，保证尽量将椅子坐满，让身体与椅子的靠背完全契合，以保持背部的生理弧度，能够增强舒适感，减少疲劳；调节办公椅的高低，使大腿和臀部受力均匀，并保证手臂与键盘平行，这样可以最大限度地减少手臂受力，有利于减少手臂疲劳；调节电脑屏幕的高低，使眼睛的平行视线略高于电脑屏幕。

2.1.3 不良坐姿的分类

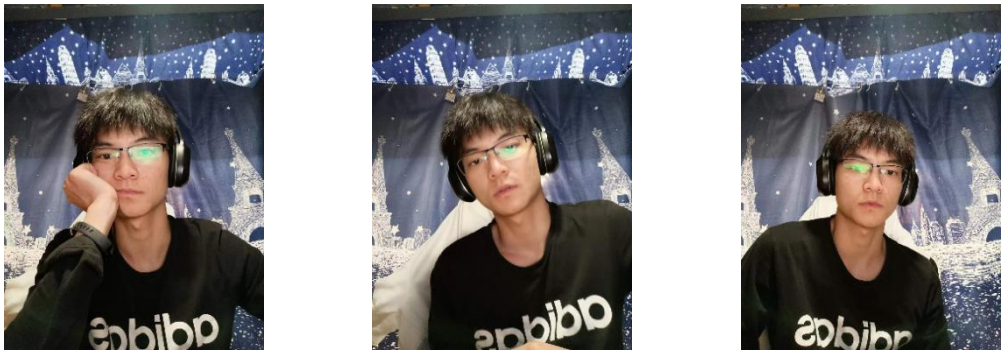
本文主要分析和研究了六种常见的不良坐姿，如图 2-3 所示是健康坐姿和不良坐姿的示意图，其中 2-3(a)为健康坐姿，其余为不良坐姿。

（1）头部倾斜（头部左偏、头部右偏）头部倾斜就是常见的歪头，在医学上也称为斜颈。通常情况下，人们以错误的姿势握笔时，会出现头部倾斜的情况。另外，在注意力过度集中时，人们的头部有时也会不自觉地歪向一侧，根据方向可分为头部左偏和头部右偏。斜颈容易造成人脸不对称、颈椎侧弯，还容易出现斜视、偏头疼等症状。

（2）肩膀倾斜（肩膀左倾、肩膀右倾）肩膀倾斜是人的肩膀处于一上一下的状态，根据方向可分为肩膀左倾和肩膀右倾。这种坐姿通常发生在因久坐疲劳需要调节坐姿，而将身体躯干向两侧偏移。此类身姿不正的状态，长久维持会压迫神经，导致脊柱变形，会出现脊柱侧弯、高低肩的情况，影响人们的身心健康。

（3）托腮（左手托腮、右手托腮）托腮指的是用手托着下巴，撑起整个头部的动作，通常人们在疲劳、打盹和思考问题的时候容易托腮。按个人习惯，可分为左手托腮和右手托腮。长时间托腮会给下巴造成非常大的挤压，容易在脸部形成皱纹，另外整个头部的重量都需要手掌和手腕来支撑，长时间托腮会导致骨骼变形。

(a)标准坐姿　(b)头部左偏　(c)头部右偏　(d)左手托腮



（e）右手托腮　（f）左耸肩　（g）右耸肩

图 2-3 坐姿分类

2.2 坐姿目标提取

2.2.1 常用的运动目标检测算法

运动目标检测技术要求从静态背景中提取运动前景[5]，这是视频分析中一个非常重要的步骤，已应用于多个领域。在医学上，用于分析生物组织运动等方面，提供了病理判断的参考依据；在现场监控等安全防范领域，相比纯粹依赖人眼进行监测的系统相比，使用基于运动目标检测的视频监控系统可以避免值班人员因主观判断造成的漏判、误判等情况，还能很大幅度的减少他们的工作强度，节约人工成本。因此，研究运动目标检测技术具有一定的理论意义和现实价值。

基于视频图像的运动目标检测一直都是研究人员的关注热点之一。运动目标检测主要是在视频中检测出变化区域，并将该区域中的运动目标从背景图像中单

独分离出来，这反映了人们对运动目标的定位和跟踪需求。根据实际的使用需求，运动检测包含静态背景和动态背景两种。通常需要把目标检测和分割操作放在目标分类和跟踪等过程中，然而，运动目标检测和分割会因周围环境的变化而受到不同程度的影响，实现起来比较困难。目前，运动目标检测主要有三类算法，分别为：光流法、帧间差分法和背景减除法。

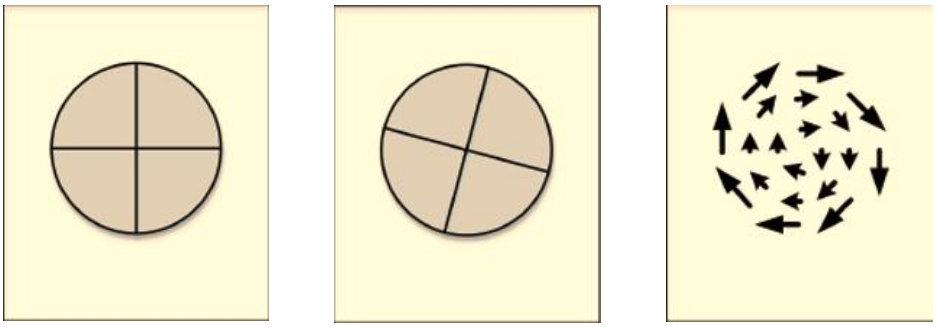
(1) 光流法

光流是图像中模式的运动速度，表示需要进一步处理的图像中与运动相关的亮度变化。光流法的原理是运用二维光流场映射三维速度场，以实现对运动图像的分析。光流场是一种二维瞬时速度场，光流计算为后续的高级处理提供了必要先决条件。如图 2-4 所示，展示了两幅连续图像对应的光流模拟示例图。

光流计算基于如下两个假设：

a) 亮度恒定：观察到的任何物体点的亮度随着时间的推移是恒定的；

b) 时间规律：相邻帧之间的时间比较短，位置的变化不明显。



(a)**时刻**t_1　　(b)**时刻**t_2　　(c)**光流**

图 2-4 **光流**

通常，这两个假设是不成立的，但对于两帧之间的小幅度运动和小间隔跳跃而言，它仍是一个不错的模型。假定一个目标像素在 t 时刻亮度为 I(x,y,t)，u、v 分别为该点光流矢量沿 x 和 y 方向的两个分量，且有 $u=\frac{d_x}{d_t}, v=\frac{d_y}{d_t}$ ，在 $t+\delta t$ 时刻运动了$[\delta x,\delta y]$后与 t 时刻具有相同的亮度，满足$\frac{dI(x,y,t)}{d_t}=0$，则：

$$\bar{L}(x,y,t)=I(x+\delta x,y+\delta y,t+\delta t) \qquad (2\text{-}1)$$

用泰勒公式对（2-1）式一阶展开并求 t 的偏导便可得到光流方程：

$$\nabla I^T v = - I_t \qquad (2\text{-}2)$$

在上述光流方程中，$v = [n,v]$是运动矢量，I_t是时间偏导。式（2-2）是线性欠约束方程组，若想求得光流，需加入约束方程，才能获得该方程的解。

稀疏光流主要计算图像上部分点的运动，主要的代表是由 Lucas 和 Kanade 两个人研究出来的 Lucas-Kanade 光流[6]，简称 LK 光流。LK 跟踪算法主要采用 Harris 角点跟随感兴趣点，可以应用于任何特征。稠密光流主要计算图像上所有点的偏移量，主要思想是基于光流场在同一运动物体上具有连续、平滑的特点，与稀疏光流有较大的不同。

光流是对象本身、对象所处的环境或相机在连续帧之间的移动所导致的目标运动。光流法的缺点是在大多数情况下存在计算复杂度高和耗费时间长的问题，并且容易受到噪声因素的影响，不利于实时处理。

（2） 帧间差分法

假定摄像机位置固定并且光线恒定，则在不同时间捕获的图像之间进行简单的减法就可以进行运动检测。帧间差分法[7]一般通过分析多张连续图像的像素灰度值来判断目标是否运动。如图 2-5 所示是帧间差分法的基本流程。

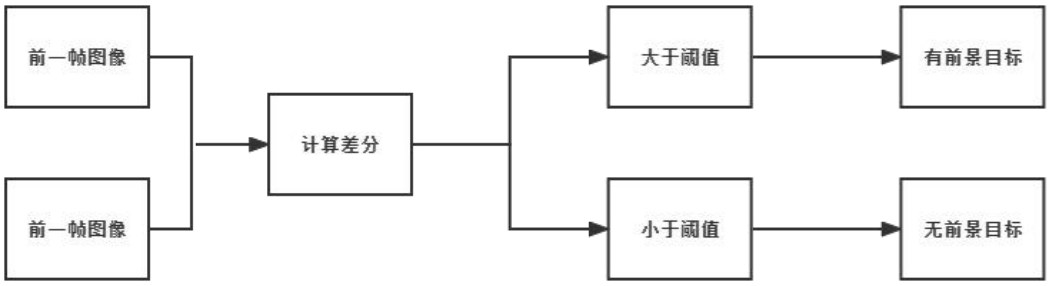


图 2-5 帧间差分法基本流程

假定取某连续的两帧灰度图像I_k,I_{k+1}，在图像上的某个像素点(i,j)在 k 时刻的灰度值记为 $t(i,j,k)$，在 $k+1$ 时刻的灰度值记为 $f(i,j,k+1)$，差分图像记为 $B(i,j)$，则有：

$$B(i,j)=\begin{cases}1 & |f(i,j,t)-f(i,j,t+1)|>T_1\\ 0 & \text{其他}\end{cases} \tag{2-3}$$

差分结果$B(i,j)$表示的是一个二值化图像，其中 1 表示图像中具有运动目标；其 0 表示图像中不存在运动目标；其中T_1 表示图像二值化过程中用到的阈值，直接决定着检测目标区域的准确性和灵敏度[8]。

帧间差分法进行运动目标检测的主要优点是算法和编程简单，抗噪能力强，能根据帧序列的运动快速适应，对目标运动具有较高的检测灵敏度。帧间差分法进行运动目标检测的主要缺点是检测位置不够准确，存在提取不到完整目标图像的情况。针对快速运动的目标和连续帧之间运动位移较大的目标，将会影响定位运动目标区域和提取特征参数的准确度，而针对慢速运动的目标和连续帧之间运动位移较小的目标，将会存在前后两帧中的目标几乎重叠而导致出现检测不到目标的情况。

帧间差分法进行运动目标检测的主要优点是算法和编程简单，抗噪能力强，能根据帧序列的运动快速适应，对目标运动具有较高的检测灵敏度。帧间差分法进行运动目标检测的主要缺点是检测位置不够准确，存在提取不到完整目标图像的情况。针对快速运动的目标和连续帧之间运动位移较大的目标，将会影响定位运动目标区域和提取特征参数的准确度，而针对慢速运动的目标和连续帧之间运动位移较小的目标，将会存在前后两帧中的目标几乎重叠而导致出现检测不到目标的情况。

(3) 背景减除法

背景减除法[9]是帧间差分法的一种特殊情况，主要是用当前帧图像减去背景帧图像，并获取运动区域的一种方法。背景减除过程的具体公式如下：

$$D_k=\begin{cases}1 & |f_k(x,y)-f_b(x.y)|>T\\ 0 & \text{其他}\end{cases} \tag{2-4}$$

式中，$f_k(x,y)$为某一帧图像，$f_b(x,y)$为背景图像，$D_k(x,y)$ 为帧差图像，T 为阈值。通过重复以上步骤处理每一个像素，最终能将运动目标完整地分割出来。如图 2-6 所示是背景减除法的基本流程。

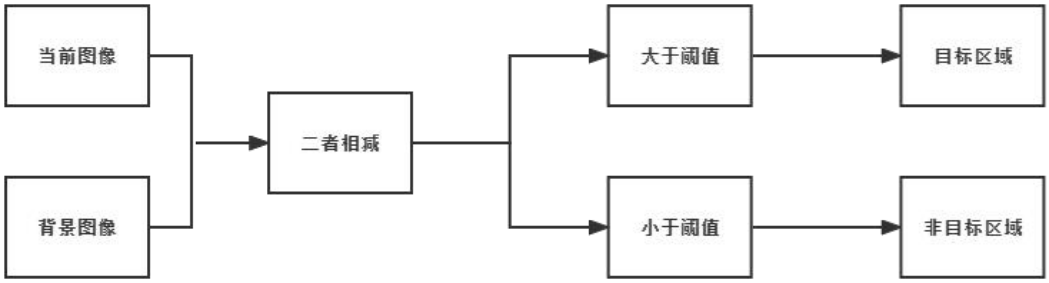


图 2-6 背景减除法基本流程

当背景图像在一段时间内基本没什么变化时，则不能继续使用预先存储的背景图像，而应该重新更新背景图像。该算法通常应用于固定不变或缓慢变化的背景，其中最重要的步骤就是获得静态背景图像。

2.2.2 基于 KNN 背景建模的背景减除法

在智能视频监控系统中，运动目标检测是首要目标，而背景目标的检测在后续的目标识别和跟踪中起着重要的作用。通常，使用背景建模技术就实现实时运动目标检测，也被称为背景估计。一般来说，拍摄背景是一个固定的场景，很少有变化；前景是指需要以静态背景为前提提取的感兴趣运动目标。通常，我们假定背景具有一些常规特性，可以用一个统计模型模拟描述，然后通过标记场景中与背景模型不相符的部分来检测入侵对象，完成对分类结果的后处理，以便达到检测运动目标的目的。背景建模的方法一般可以分成三个阶段：背景初始化阶段、前景检测阶段和背景维护与更新阶段。背景模型的建立并不是固定不变的，通常需要维护一个动态背景帧，首先对历史帧的像素点进行统计，并计算像素点的平均值；然后在更新一帧图像时，需要更新历史的 N 帧图像序列同时，用当前新帧数据替换最先前的帧数据，更新方式如下：

$$b_{\Pi+1}(ij) = \alpha t(i,j) + (1-a)b_n(i,j) \tag{2-5}$$

其中，参数 α 为学习率。

利用像素点的均值或中值来完成背景建模比较简单，但对于运动速度慢、亮度变化小的目标，存在统计缺失现象。为了这个解决问题，Zoran Zivkovic 和 Ferdinand van derHeijden 在 2006 年发表的论文中提出了 K 近邻思想的改进

方法[10]。

k 近邻算法（k-Nearest Neighbor，KNN）是一种非参数化模型的机器学习算法。非参数化模型不能通过一组固定的参数表示，随着训练数据的增加参数的数量也会递增。KNN 是一个非参数化模型的子类，表示基于实例的学习。KNN 没有明显的前期训练过程，是惰性学习算法的典型示例，这样的模型在训练阶段只是把样本保存起来，训练时间没有计算成本，等到获取测试样本之后再进行处理。k 近邻法的三要素：k 值选择、距离度量和分类决策规则。

选取不同的 k 值会使 k 近邻法的分类结果有显著的不同，若 k 值较小，则会增加整个模型的复杂度；若 k 值较大，则会使模型变得简单。因此，在工程实践中通常使用交叉验证来确定最优 k 值。也就是说，平衡过拟合和欠拟合的关键在于找到合适的 k 值。k 近邻模型常用的特征空间距离计算方式如下：

$$d\left(x^{(i)},x^{(j)}\right)=\sqrt[p]{\sum_k |x_k^{(i)}-x_k^{(j)}|^p} \tag{2-6}$$

式中，当 p 2 ，则为欧几里得距离；当 p 1 时，则为曼哈顿距离。KNN 算法的预测结果表示这 k 个实例中出现次数最多的类别标记，也就是常用的“投票法”。

如图 2-7 说明了新的数据点（图中问号的区域）如何根据周围距离最近的 5 个点进行投票抉择而被标记上三角形图标：

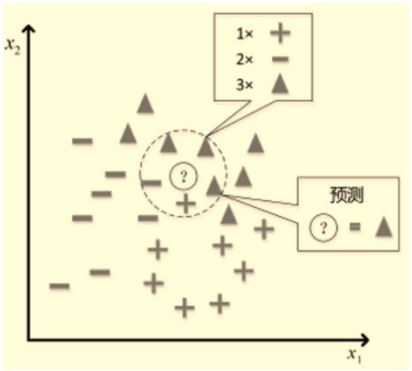


图 2-7 近邻分类示意图

2.2.3 坐姿目标图像提取

运动目标检测的前提是将前景目标和静态背景在场景中进行分离。其中，光流场法的计算复杂，不适合现场实时监测；帧间差分法虽然简单，但检测结果并不完整。考虑到

提取坐姿目标时的环境变化相对较小，且场景中摄像头也是固定的，本文选用背景减除法来完成运动目标提取，但是会存在一些空洞和噪声，因此，本文在提取到原始运动目标后，加入相应的形态学处理来填充空洞。经过形态学滤波，可以有效地去除二值化图像中的噪声点，连通了运动目标的整个区域。接下来，需要提取运动目标区域的轮廓，得到轮廓的外接矩形，以提取完整的运动目标区域。如图 2-9 所示是运动目标提取的流程。

(1) 形态学处理

形态学[11]，主要的研究对象是图像的形态特征，通过提取图像内的分量信息来表达和描绘图像的基本特征和结构，它是图像处理中应用最为广泛的技术之一。在数字图像处理的应用中，形态学可以简化图像数据，消除不必要的结构并保持图像的基本形状特征，以便进一步的图像处理操作。利用形态学方法对图像进行处理和分析即是对目标的形态学分析，基本的形态学运算主要包括：腐蚀操作[12]、膨胀操作[13]、开运算[14]和闭运算[15]。

A. 膨胀操作

膨胀是最基本的形态学运算之一，可以扩展目标像素点。假设 A 为图像像素集合，B 为结构元素，则膨胀操作可以表示为：

$$A \oplus B = \left\{ z | \left(\widehat{B} \right)_z \cap A \neq \emptyset \right\} \tag{2-7}$$

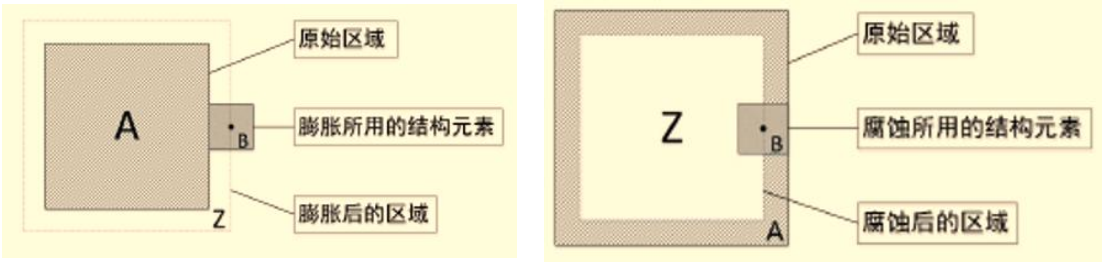
形态学膨胀操作的思路是运算前 A 与 B 分别为两个区域，如图 2-10（a）所示，B 区域中的黑点表示 B 的中心点，B^表示 B 相对于自己中心对称变换后图形，运算后相当于 B 对称，沿着区域 A 的边界遍历一圈，区域 B 的中心扫过的区域加上 A 本身的区域就是 A 膨胀区域 B 的结果，其中 zB 表示将 B 平移，使其中心点位于 z 位置。

B. 腐蚀操作

腐蚀是“缩小”或“细化”二值图像中的对象，主要规则是输出像素的值是所有输入像素的最小值。则腐蚀操作可以表示为：

$$A \oplus B = \{ z | (B)_z \subseteq A \} \tag{2-8}$$

形态学腐蚀操作的思路是运算前 A 与 B 分别为两个区域，如图 2-8（b）所示，使用 B 对 A 进行腐蚀就是沿着区域 A 的内部边界遍历一圈，平移区域 B 形成的集合区。B 的中心形成的轨迹即是腐蚀后 Z 的边界，$(B)_z$ 表示将 B 平移，使其中心点位于 z 位置。



(a)膨胀示意图　　（b）腐蚀示意图

图 2-8 形态学膨胀和腐蚀示意图

C. 开运算和闭运算

开运算和闭运算都继承自腐蚀和膨胀操作，并且由这两个基本的操作组合而成。开闭运算都能够使目标的边缘更加光滑，并保持目标的整体位置和形状不变。其中，开运算和腐蚀操作比较类似，可以消除物体的像素边缘，但消除的边缘像素数量没有腐蚀操作那么多，它主要的作用是打破狭窄的连接，消除小凸起；闭运算将保留背景区域中与结构元素形状相似的像素，填补物体之间宽度小于结构元素的间隙和孔洞。

形态学的开运算就是对图像进行先腐蚀后膨胀，两个操作使用的结构元素相同。开运算可以表示为：

$$A \circ B = (A \Theta B) \oplus B \tag{2-9}$$

其中，A 表示二值图像像素集合，B 为结构元素。

形态学的闭运算就是对图像进行先膨胀后腐蚀，前后使用相同的结构元素。闭运算可以表示为：

$$A \cdot B = (A \oplus B) \Theta B \tag{2-10}$$

（2）提取轮廓外接矩形

轮廓是对物体形状的有力描述，在图像分析和识别中起着重要的作用。当检测出目标的轮廓后，根据目标的形状完成对目标的边界提取，通常边界提取包括矩形框、圆形框和椭圆型框等操作。在前一小节中，对运动目标图像进行形态学操作后消除了二值图像中的大部分噪声和空洞，连通了整个目标区域。接下来，需要借助像素点的坐标获取目标轮廓的外接矩形。轮廓提取的步骤为：首先二值化处理源图像，然后利用边缘点连接的层次差别提取出一组位于结构特征高的区域点集构成的集合，最后得到的点集就可能就是对象的轮廓。

第三章 坐姿多特征提取和融合

特征是表征目标运动状态 的信息。提取出图像中的坐姿区域后，需要进行特征数据的采集，即对其中的重要特征进行提取。 特征是表征目标运动状态 的信息。这些信息一方面可以从图像中获取，例如颜色和纹理结构等，另一方面也可以从运动学角度获取到目标的速度和位置等信息。姿势识别的主要依据是特征，其选择和提取直接影响着识别正确率。

然而每种特征是从不同的角度描述运动特性，不能够全面地涵盖姿势的所有信息。因此，将多种特征融合使用，可以取长补短，包含更丰富的特征信息，也就能够更准确地表征姿势。

3.1 图像预处理方法

在对图像进行特征提取前，需要对图像进行预处理，经过预处理的图像相比较原始 RGB 图像，信息更加简单直观，有些特征是可以直接提取或者经过简单计算就可以找到。首先需要将三通道的 RGB 原始图像进行灰度处理，变成单通道的灰度图像，这样减少了计算量；其次需要对图像进行直方图均衡化，这样能够增强图像的对比度；最后需要对图像进行滤波，抑制噪声。

3.1.1 灰度空间转化

对于彩色图片，一般采用的方法是转换到灰度空间，这样做的好处是，减少了运算复杂度。对于彩色方式现在有很多种类的颜色空间，不管哪种方式，一般有三个或者四个基元素，通过组合这些基元素来产生所有的颜色。RGB 颜色空间是最常用的一种颜色空间，它跟人眼采用相似的工作机制，分别为红色、蓝色和绿色，有时为了表示透明颜色也会加上第四个元素 alpha（A）。

(a)三通道

(b)单通道

图 3-1 三通道和单通道

3.1.2 直方图均衡化

直方图是一张图像中最为直观的像素值分布体现，也是最容易得到的信息，从一张图像的直方图中，可以看到这张图像的像素值的大体分布，进而了解到这张图片的一些信息。直方图均衡化后的直方图分布

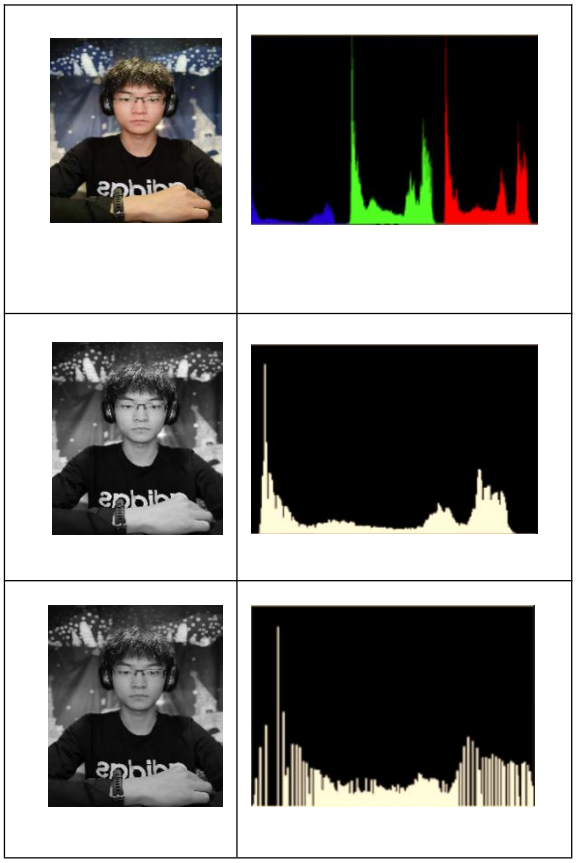


图 3-2 直方图均衡化后的直方图分布

但是从实验环境中的得到的图像可能会受到当前周围环境因素的影响，如光照、摄像头、传输过程失真、图片编解码错误等，为了得到最佳的检测图形，需要对原始图像进行直方图均衡化的处理。 如图 3-2 所示，直方图均衡化可以过滤掉一张图中较大或较小的像素值，使其增加或者减小，这样有助于减少一些斑点噪声的影响。并且直方图均衡化可以提高图像的对比度的，如本文中，需要对笔部特征的检测，若背景较为昏暗，此时可以使用直方图均衡化，增加对比度，方便检测。

3.1.3 滤波

滤波一般指消除一些噪声，对于图像的滤波其实就是减少图像中的噪点或者失真，尽量保留图像的细节特征，对图像的噪声进行抑制，平滑处理。图像滤波的目的简单来说就两个：消除图像中的噪声；抽出对象的特征作为图像识别的特征模式。在一幅灰度图像中，灰度剧烈变化的点属于高频部分，图像中较为平坦的、灰度变化不大的点属于低频。 在频域中，对于一幅图像，高频部分代表了图像的细节、纹理信息；低频部分代表了图像的轮廓信息。根据图像的高频与低频的特征，可以设计相应的高通与低通滤波器，高通滤波可以检测图像中尖锐、变化剧烈的区域；低通滤波可以让图像变得平缓光滑，滤除图像中的噪声。 均值滤波是最简单的一种滤波操作，给定图像中目标像素一个模板，该模板包括了其周围的临近像素，计算该模板内图像对应像素的平均值，替换图像的目标像素的像素值。均值滤波是典型的线性滤波，也是归一化之后的方框滤波，主要方法为邻域平均法。缺陷是不能很好的保护好图像细节，在图像去燥的同时也破坏了图像的细节部分，从而使图像变得模糊。 中值滤波的基本思想是用像素点邻域灰度值的中值来代替该像素点的灰度值，该方法在去除脉冲噪声、椒盐噪声的效果不错，可以去除图像中像素极大或者极小值，能够保留住图像的边缘细节信息。它首先需要将像素值排好序，然后取中值，把中心点的值用中值代替，让周围的像素值接近真实值，从而消除孤立的噪声点，是一种基于排序统计理论的能有效抑制噪声的非线性信号处理技术， 高斯滤波是一种线性平滑滤波，适用于消除高斯噪声，高斯滤波用临域内像素的加权平均灰度值去替换模板中心像素点的值，当然这个周围点的半径范围是可以设置的。但是周围点对当前像素点的影响存在一个权重问题，也就是离当前点越近的权重越大。所以用二维的高斯函数来计算

权重如公式 3-1 所示。

$$G_0(x,y) = A\,exp\left(\frac{-(x-u_x)^z}{2\sigma_x^2} + \frac{-(y-u_y)^2}{2\sigma_y^2}\right) \quad (3\text{-}1)$$

图像的高斯模糊过程就是图像与正态分布做卷积。

其公式化的描述一般如下所述：

$$H(\beta) = d^{-1}(x)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi)w(\xi,x)d\xi \quad (3\text{-}2)$$

$$d(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\xi,x)d\xi \quad (3\text{-}2)$$

其中的 w 为基于空间距离的高斯权重；而 d(x)的作用类似于空域滤波中加权后再除以加权系数之和；f(・)表示图像的一个像素。

以上两个公式都是用连续函数表达，采用的是数学上的双重积分符号来表达对像素点的加权求和。如果是采用离散的求和符号表达，则采用如下表达式：

$$g(x,y) = \sum_{s=-a}^{a}\sum_{t=-b}^{b} w(s,t)f(x+5,y+t) \quad (3\text{-}3)$$

w(s,t)即滤波模板中的加权系数，a 和 b 为模板的大小，一般选择奇数，f(x+s,y+t)为滤波模板覆盖处的图像像素值。

图 3-3 均值滤波、中值滤波、高斯滤波图像

高斯模糊的应用场景一般作为退化函数使用，可以去除图像噪声，Canny 边缘提取的第一步就是高斯模糊，以此来消除噪声干扰，用高斯模糊去噪对于随机噪声效果明显。如图 3-4 所示，是三种滤波的效果图。

图像滤波后的质量评价（Image quality assessment，IQA）通常有以下几种评价指标：

a.结构相似度(Structure similarity Index, SSIM)

结构相似度是从亮度、对比度与结构来对两幅图像的相似性进行评估。在实现上，用均值来表示亮度，用均值归一化的方差表示对比度，用相关系数即统计意义上的协方差与方差乘积比值来表征结构。此外，对于局部可抵抗失真程度突变，SSIM 效果比较好。实际是对各种局部窗口的 SSIM 做平均，为了防止出现块效应需要用高斯加权函数对每个局部的统计值进行加权。

结构相似度一般会以亮度因子、对比度因子、结构相似因子来表示，它们的表达式见公式 3-6 所示。

$$L(x,y)=\frac{2u_xu_y+}{u_x^2+u_x^2}\frac{c_1}{+c_1}$$

$$C(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_x^2+C_2} \tag{3-6}$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}$$

公式 3-6 中，μx、μy 分别表示图像 x 和 y 的均值，σx、σy 分别表示图像 x 和 y 的标准差，σxσx、σyσy 分别表示图像 x 和 y 的方差。σxy 代表图像 x 和 y 协方差。C1，C2 和 C3 为常数，是为了避免分母为 0 而维持稳定。通常取 C1=(K1×L)2, C2=(K2×L)2, C3=C2/2, 一般地 K1=0.01, K2=0.03, L=255。

最后 SSIM 指数如公式 3-7 所示。

$$\text{SSIM}\,(x,y)=L(x,y)\times C\left(x,y\right)\times s(x,y) \tag{3-7}$$

当 C3=C2/2 时，则公式 3-7 可以简化公式 3-8。

$$\text{SSIM}\,(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{\left(u_x^2+\mu_y^2+C_1\right)\left(\sigma_x^2+\sigma_y^2+C_2\right)} \tag{3-8}$$

可以看到，结构相似度的优点是可以较好地反映人眼主观感受，但是缺点也很明显，就是计算量有点大，计算稍复杂。

b.均方误差(Mean square error, MSE)

均方误差的原理是对原始帧和失真帧做差取平方求和，此外一些评价指标如绝对误差(Mean absolute error, MAE)、根均方误差(Root mean square error, RMSE)、标准偏差(Standard deviation, STD)效果也类似

它的公式见公式 3-4 所示。

$$MSE=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X(i,J)-Y(i,j)\right)^2 \tag{3-4}$$

均方误差有着计算简单的优点，但是一次同时也会有和主观评价差距较大的缺点。

b.峰值信噪比(Peak signal noise ratio, PSNR)

峰值信噪比是峰值信号的能量与噪声的平均能量之比，通常表示的时候取 log 变成分贝，由于 MSE 为真实图像与含噪图像之差的能量均值，而两者的差即为噪声，因此 PSNR 即峰值信号能量与 MSE 之比。最普遍、最广泛使用的评鉴画质的客观量测仍然是 PSNR，虽然不完全和人眼看到的视觉品质一致，但目前仍作为对照其他指标的基线。

它的公式见公式 3-5 所示。

$$PSNR=10\,log_{10}\frac{MaxValue^2}{MSE}=20\,log_{10}\frac{MaxValue}{\sqrt{MSE}} \tag{3-5}$$

峰值信噪比同样有着计算简单的优点，但是也有颜色变化会拉低分数，和主观评价有一定差距的缺点。 通常用来评价一幅图像压缩后和原图像相比质量的好坏，显然压缩后图像一定会比原图像质量差。

表 3-1 图像质量评估指标

指标	说明
PSNR（峰值信噪比）	取值为正数，值越高，压缩后失真越小，反之失真越大。
SSIM（结构相似性）	取值范围为 0-1，值越大则图像质量越好，反之质量越差。

一般 PSNR 的值大于 40 说明图像质量极好，低于 20 说明图像质量极差。经过计算，本文中的均值滤波、中值滤波和高斯滤波的 PSNR 值和 SSIM 值分别为如表 3-2 所示。其中高斯滤波的效果要比均值滤波以及中值滤波都要好得多，因此根据本文的实验环境，高斯滤波能在减少噪声的同时，还能保留大部分有用的信息。

表 3-2 均值滤波、中值滤波和高斯滤波的 PSNR 值和 SSIM 值

	均值滤波	中值滤波	高斯滤波
PSNR	41.243	39.032	41.971
SSIM	0.999611	0.999201	0.999621

3.2 特征提取方法

人体坐姿能够通过手和脸的空间位置关系来描述，而脸和手有着鲜明的肤色特征。因此，可对坐姿图像进行肤色检测得到其肤色特征图。肤色检测主要以颜色空间变换和肤色建模两个步骤为主[16]。常见的颜色空间有 RGB、HSV 及 YCbCr 颜色空间等。

3.2.1 坐姿肤色特征提取

以T_d 和T_u 分别表示肤色检测时亮度变化对色度的影响临界值； $B_{Cb}(Y)$、$B_{Cy}(Y)$ 表示小于T_d 和大于T_u 时模型修正肤色区域宽度； $M_{Cb}(Y)$、 $M_{Cr}(Y)$ 表示中轴线；$Cb'(Y)$和$Cr'(Y)$为修正后肤色特征区域。I 通过公式(3.1)和(3.2)转换为I'。

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} R \\ G \\ B \end{pmatrix} \cdot \frac{255}{R+G+B} \tag{3-1}$$

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.311 & 0.500 \\ 0.500 & -0.419 & -0.018 \end{pmatrix} \begin{pmatrix} r \\ g \\ b \end{pmatrix} \tag{3-2}$$

在 YCbC r 颜色空间对 Cb 和 Cr 进行如下非线性变换。

$$Ci'(Y) = \begin{cases} \left(Ci(Y) - M_{Ci}(Y)\frac{B_{Ci}}{B_{Ci}(Y)} + M_{Ci}(T_u)\right) & if\ Y < T_d\ or\ T_u < Y \\ Ci(Y) & if\ Y \in [T_d, T_u] \end{cases} \tag{3-3}$$

通过(3.3)式的非线性变换,求解到 YCbCr 空间中肤色聚类的分布情况,然后在 Cb - Cr 子空间中进行投影,位于式(3.4)的椭圆内部区域即为肤色区域。

$$(x - eC_x)^2/_{A^2} + (y - eC_y)^2/_{B^2} = 1 \tag{3-4}$$

式中

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} Cb' & -C_f \\ Cr' & -C_y \end{pmatrix} \tag{3-5}$$

在实际坐姿检测中，该模型会对亮度较低或较高的区域产生检测误差，将低亮度区域非肤色区域检测成肤色，而高亮度区域的肤色区域检测成非肤色。因此，在亮度小于 75 时，将 A 、 B 缩小为原来的 Y/ 75 ；在亮度大于 240 时，增大为原来的 Y/ 240 ;在大于 75 而小于 240 时， A 、 B 不变。

考虑到对肤色特征图二值化处理时，简单地将肤色区域的像素灰度值设为 255 ，非肤色区域的像素灰度值设为 0 ，丢失了肤色区域所携带的信息，所以将 I 中 Cb 、Cr 颜色分量在椭圆区域内的像素生成比例灰度特征值，而超出椭圆区域的部分的像素灰度值赋值为 0 ，如图 3.4 所示。灰度特征经过 PCA 算

法[17]降维处理构成灰度特征向量，向量元素记为f_{ijl}，即获取的第 j 类坐姿的第 l 个肤色特征向量元素。

图 3-4 肤色检测中比例灰度特征图

3.2.2 坐姿 SURF 特征提取

Herbert Bay 等人在 2006 年提出的 SURF[48](Speeded Up Robust Feature)特征，与 SIFT 算法相比，具有更快的运算速度。它是通过寻找图像中的一些“稳定点”20 确定 SURF 特征点，这些点对光照条件的改变不敏感且较稳定地对应对象的姿势，比如与领域内像素亮度差异较大的点和位于轮廓上的边缘点等[18]。该算法包括积分图像卷积、Hessian 矩阵构建、尺度空间生成、特征点的确定和精确定位及主方向的确定等步骤。

⑴ 积分图像卷积

积分图像加快计算出矩形区域内所有像素点灰度值的总和，提高算法的时间效率。设积分图像中任意一点 I(i, j) 的灰度值为 n(i, j) ，该值为原图像左上角到该任意点 I(i, j) 相应的对角线区域灰度值的总和，即：

$$n(i,j) = \sum_{i`<i,j`<j} p\left(i^{'},j^{'}\right) \tag{3-6}$$

可以通过简单的加减运算快速计算出矩形区域内的像素点的总和，来提高算法的时间效率。例如先通过加减运算对图像进行积分图像操作，并将结果保存在与原图对应的矩阵 M 中，当要对图像中的某矩形区域内的像素点求和时，只要对照积分图像矩阵查找出矩阵顶点 A 、 B 、 C 及 D 的积分值，通过式子

Σ=A-B-C+D 就能计算出结果。积分图像法实际上是利用计算出的 n 个互不相交的矩形区域内像素和的值，递推出其它未知值，从而有效地避免了重复计算。

⑵ Hessian 矩阵构建

与 SIFT 算法相比，SURF 通过 Hessian 矩阵来替代高斯滤波在不同尺度空间进行特征提取，来提高算法的时间效率。给定图像 f(x, y) 中一个点 (x, y) ，其 Hessian 矩阵 H(x, y, σ) 在(x, y)处，尺度空间为σ时的定义如下：

$$H_{(x,y,\sigma)} = \left[\frac{L_{xy}(x,y,\sigma)L_{xy}(x,y,\sigma)}{L_{xy}(x,y,\sigma)L_{yy}(y,y,\sigma)}\right] \tag{3-7}$$

⑶ 尺度空间生成

与 SIFT 类似，SURF 特征也是通过构建尺度空间的方式使特征具有尺度不变性。在图像处理领域，常常用图像金字塔来实现尺度空间。在 SIFT 算法中，图像金字塔由图像降采样形成一系列尺寸大小不一的单元组成，每个单元又由几张不同尺度图像构成。高斯模糊的过程中，不改变高斯模板的尺寸大小，只在同一单元中改变高斯模板的尺度。而在 SURF 算法中，图像的尺寸大小始终保持不变，只在同一单元组内改变高斯模板尺寸大小，不同单元组之间改变高斯模板的尺度值，对图像进行高斯滤波构建尺度空间。如图 3.11 所示，左边是 SIFT 算法按通常的方式的构建金字塔状的尺度空间，其中图像的尺寸大小是变化的。而右边则是 SURF 算法通过改变高斯模板尺度大小而不改变图像的尺寸的方式构建金字塔空间。由于 SURF 算法节省了图像降采样的过程，且能在图像金字塔各个单元组并行处理高斯模糊过程，所以，SURF 算法较 SIFT 算法运行速率有了很大的提高。

⑷ 特征点的确定和精确定位

经 Hessian 矩阵处理得到感兴趣点之后，在 3 3 3 领域空间内，使用非最大抑制法挑选出极大值特征点。如图 3 - 5 所示，用 ‘x’标记待检测像素点，将待检测点与同一尺度层的周围 8 个及对应的上下两层 9 个像素点相比较，若其响应值为最大值时，选定为特征点。然后，在图像及尺度空间中插值微调，得到更精确的特征点尺度值和位置。

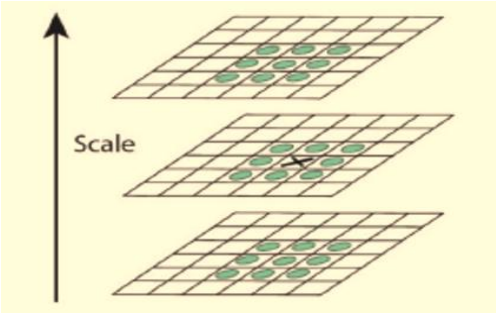


图 3-5 SURF 特征点精确定位示意图

⑸ 选取特征点主方向确定

在以特征点为中心，半径为6π的圆形领域内，使用 Harr 滤波对图像进行检测。按与圆心距离大小给这些响应值赋高斯权重系数，使得离圆心近的响应值赋值较大的权值，反之亦反。然后，求解出领域内以60π为步长划分的扇形区域中 Harr 滤波响应值总和，形成六个新的不同矢量。然后，以其中最长矢量方向作为特征点的主方向。

3.3 特征融合

将多种特征融合，可以取长补短，包含更丰富的特征信息，也就能够更准确地表征姿势。采用肤色特征与 SURF 特征集成可以让单一特征之间相互补充，提高图像分类识别的性能。特征融合主要是将已有的多个特征结合起来，形成新的特征，丰富了图像信息，并消除了因不同特征间相关性所带来的冗余信息。特征融合主要有两种方法：串行融合和并行融合，其中串行融合方式是把各类特征向量通过线性连接的方式进行组合；并行融合方式是将各类特征向量通过一个复合特征向量组合起来。

设 V 为 N 类特征加权融合向量，坐姿有 M 类，L 为第 i 类特征的第 j 种坐姿类别的特征维数，采用肤色特征与 SURF 特征融合，$v_{i1}....v_{i7}$依序对应正确坐姿、右手托腮、左手托腮、头往右斜、头往左斜、右肩偏上、右肩偏下等 7 类坐姿。

检测用户后，用肤色模型提取目标肤色特征,并用 PCA 算法降维，使得各坐姿类别均为L_1维肤色特征向量，然后对目标提取 SURF 特征，使得各坐姿类别均

为L_2维 SURF 特征向量。

3.4 数据处理方法

数据是非常重要的，数据的好坏将直接决定着最后识别正确率的高低，影响最后的实验结果。 但是特征提取完的数据，一般是夹带着有漏检、检错、异常的数据，这样的数据不能直接用于分类训练，会影响最后的识别准确率。还需要对提取的特征数据进行数据处理，这样最后得到的数据才会用于最终的分类训练。

3.4.1 缺失数据处理

在获取信息和数据的过程中，会存在各类的原因导致数据丢失和空缺，比如没有采集到数据，但是由于使用的采集数据的软件环境不同，比如采用 C/C++ 语言编写的程序会将未赋值的静态变量或全局变量自动置为 0。对于针对这些缺失值的处理方法，主要是基于变量的分布特性和变量的重要性采用不同的方法。主要分为以下几种：

哑变量填充：若变量是离散型，且相同值较多，可转换成哑变量，若某个变量存在十几个不同的值，可根据每个值的频数，将频数较小的值归为一类，降低维度。此做法可最大化保留变量的信息。

插值法填充：包括随机插值，多重差补法，拉格朗日插值，牛顿插值等。

定值填充：常规定一个不可能获得的值来表示，比如-1 或者当前位数的最小的负值。

统计量填充：若缺失率较低并且重要性较低，则根据数据分布的情况进行填充。对于数据符合均匀分布，可用该变量的均值进行填补，对于数据存在倾斜分布的情况，可以采用中位数进行填补。

删除变量：若变量的缺失率较高（若占绝大部分），且重要性较低，可以直接删除变量。

3.4.2 异常数据处理

异常值是数据分布的常态，离群点处理处于特定分布区域或范围之外的数据通常被定义为异常或噪声。异常分为两种：“伪异常”，由于特定的业务运营动作

产生，是正常反应业务的状态，而不是数据本身的异常；“真异常”，不是由于特定的业务运营动作产生，而是数据本身分布异常，即离群点。主要有以下检测离群点的方法：

基于聚类：利用聚类算法，丢弃远离其他簇的小簇。

基于距离：通过定义对象之间的临近性度量，根据距离判断异常对象是否远离其他对象，缺点是计算复杂度较高，不适用于大数据集和存在不同密度区域的数据集。

基于密度：离群点的局部密度显著低于大部分近邻点，适用于非均匀的数据集。

简单统计分析：根据箱线图、最大值最小值分布、各分位点判断是否存在异常，例如 pandas 的 describe()函数可以快速发现异常值。

基于绝对离差中位数（MAD）：计算各观测值与平均值的距离总和，放大了离群值的影响。这是一种稳健对抗离群数据的距离值方法。

常见的处理手段有以下两种：

1.根据异常点的数量和影响，考虑是否将该条记录删除。

2.平均值或中位数替代异常点，简单高效，信息的损失较少。

由于本文中对于坐姿区域特征的提取采取了缩小检测区域，并且剔除了上端的八分之一的区域和下端三分之一的区域，因此本文的数据异常点相对于总数数据量并不是太多，通过绝对离差中位数方法找到异常值，本文采用直接剔除该组异常点的方法进行数据处理。测试得到了 500 条数据，如表 3-3 所示，统计了各位置的数据个数，最大值，最小值等信息，640*480 图片规格为例。

表 3-3 特征数据的基本数据

项目	NoseX 坐标	NoseY 坐标	NeckX 坐标	NeckY 坐标	L_SX 坐标	L_SY 坐标	R_SX 坐标	R_SX 坐标

总数	500	500	500	500	500	500	500	500
平均值	296	263	321	407	180	385	483	378
最大值	320	272	326	418	185	410	493	397
最小值	267	254	316	396	174	360	473	359

第四章基于 OpenPose 坐姿评估系统设计与实现

本章主要通过搭建基于 OpenPose 的坐姿识别系统对不良坐姿进行识别，并对不良坐姿进行提示与规劝，避免用户长期因不良坐姿而导致产生疾病。

4.1 运行环境

4.1.1 软件要求

- window 11 64bit
- openpose 1.7.0
- cuda 10.2
- cuDNN 8.3.3
- visual studio 2019
- cmake 3.20.2

4.1.2 硬件参数

这里给出 GPU 版本 OpenPose 的硬件参数，相比 CPU 版本精度更高，当然硬件要求也更高。

表 4-1 GPU 版 OpenPose 的硬件要求

项目	最低配置	本文配置
显存	1.5G	6G
运行内训	2G	16G
CPU 核心数	8 核	16 核

4.2 系统设计

根据需求分析，系统设计大致分为四个模块，它们分别是：数据采集模块，数据预处理模块，坐姿评估模块，可视化数据分析模块。

数据采集模块：主要实现的功能是采集骨骼点数据和骨骼图，通过 OpenPose 进行指定骨骼关节点的跟踪以及采集。

数据预处理模块：主要实现的功能是对骨骼点进行连接，计算头部和肩部的偏转角。

坐姿评估模块：主要实现的功能是对坐姿进行识别和评估，利用数据处理模块处理好的数据作为参数，并对已识别出的坐姿进行分类，当识别出为不良坐姿时，对用户进行提醒纠正，评估过程中收集用户坐姿信息保存为 csv 格式。

可视化数据库分析模块：利用 Python 脚本对坐姿评估模块采集的 csv 数据文件进行可视化数据分析，生成本地分析图表和通过 web 服务显示报表。

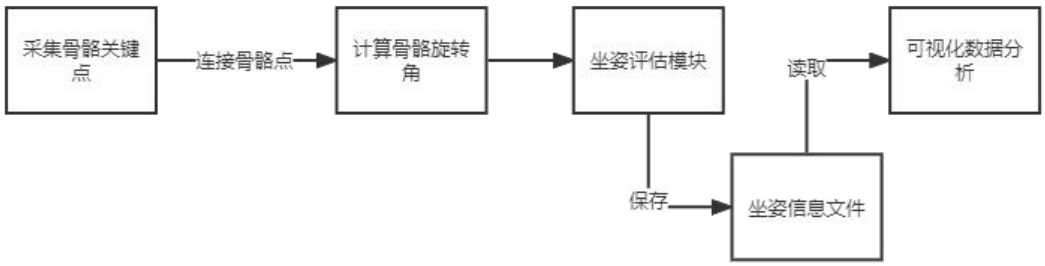


图 4-1 四个模块之间的工作流程图

4.3 各模块具体实现

本节主要介绍四个模块在整个系统中的具体实现和设计细节。

4.3.1 数据采集和预处理模块

首先 OpenPose 会对摄像头采集的图片进行自下而上的人体姿态估计，再通过 CPM(卷积姿势机)的方法，检测出人关键点的位置,得到检测结果是通过预测人体关键点的 heatmap，这样就可以看到每个人体关键点上都有一个高斯的峰值，

代表网络预测出这里是一个人体的关键点,在得到检测结果之后，采用 PAF(部分亲和场)对关键点加测结果进行连接,显示在屏幕上。

图 4-2 数据采集和预处理模块

4.3.2 坐姿评估模块

该模块被分为五个部分，分别为间隔采样，图像压缩，坐姿评估算法，语音提示和久坐提醒，可视化数据分析。

4.3.2.1 间隔采样

（1）间隔采样：主要指的是在时间维度上对摄像头采集到的连续图像进行间隔采样。

（2）采样定理：在进行连续信号离散化的过程中，当采样频率大于信号中最高频率的 2 倍时，采样之后的离散信号可以完整保留原始信号中的信息。

坐姿是一种持续性动作，因此，可以通过选用合适的帧率和采样间隔 t，使采样频率大于坐姿变换频率的 2 倍以上，就可以完整保留信息。所以，为了减少模型估计时间和系统流畅度，节省算力和降低系统硬件门槛，我们采用间隔取样的方式采集信息，流程如图 4-3 所示。

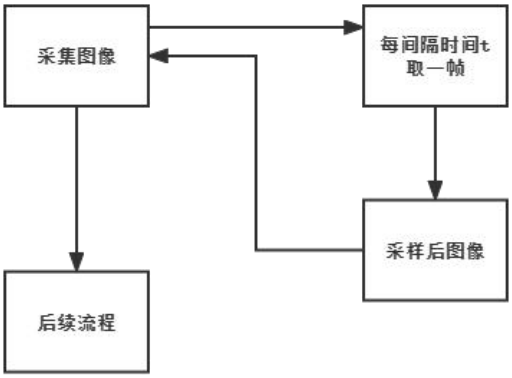


图 4-3 间隔取样流程

4.3.2.2 图像压缩

（1）原因：估计人体结构模型中的每个骨骼点时都针对于图像中的所有像素，因此图像的大小直接决定了模型的估计时间。采用几何变换将图像的像素进行压缩，可以减轻模型估计时的计算压力, 图像压缩流程如图 4-4。

（2）常见图像压缩方法：最近邻法、双线性插值、双三次方插值、像素区域重采样等。

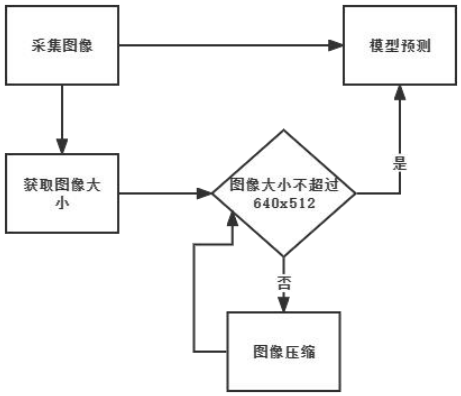


图 4-4 图像压缩流程

4.3.2.3 坐姿评估算法

该部分中，使用坐姿评估算法处理数据预处理模块提供的骨骼点信息，对坐姿进行分类，对不正确的坐姿进行判断。

给定关键点 Nose(0)，Neck(1)，Left_Shoulder(2)，Right_Shoulder(5)，L，N，m，h 分别代表(1,2)，(0,1)，(5,1),(0,2)间的差值，如图 4-5。

$$\frac{m}{n} > 0.13 \qquad (4\text{-}1)$$

$$\frac{N}{L} < 0.25 \;\&\&\; \frac{N'}{L'} < 0.25 \;\&\&\; \frac{N}{L} \times \frac{N'}{L'} > 0 \qquad (4\text{-}2)$$

$$\frac{N}{L} < 0.1 \;\&\&\; \frac{N'}{L'} < 0.1 \qquad (4\text{-}3)$$

当式（4-1）成立时，可判断颈部出现问题，当式（4-2）和式（4-3）都不成立时，可判断肩膀出现问题，并将结果显示在界面的结果框中，如图 4-6。

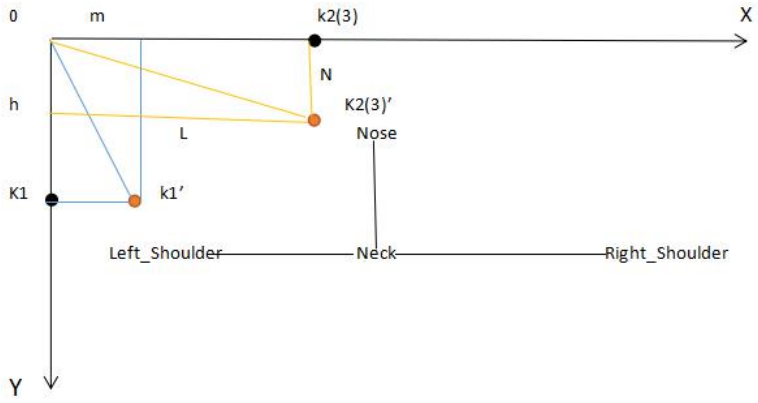


图 4-5 骨骼点间的关系

图 4-6 UI 界面和结果输出

4.3.2.4 语音提示和久坐提醒

为了提醒用户端正坐姿和定期放松，我们添加了语音播报和久坐提醒功能，语音播报模块部分程序实现如表 4-2 所示，语音提醒的设计逻辑是每三次取样视为一个周期，当一个周期内连续两次及以上错误坐姿时系统会进行语音提醒，然后进入下一个检测周期，值得注意的是若一个周期内两次错误非连续出现，则不会报警，久坐提醒的设计逻辑是当用户坐下后系统即刻计时，久坐时长达到一小时后系统会发出语音播报，提醒用户适当活动，久坐提醒部分功能实现如表 4-3 所示。

表 4-2 语音播报功能部分程序实现

```
void MSSSpeak(LPCWSTR speakContent) {
    ISpVoice* pVoice = NULL;
    if (FAILED(::CoInitialize(NULL))) {
        MessageBox(NULL,(LPCSTR)L"COM接口初始化失败！",(LPCSTR)L"提示",
        MB_ICONWARNING | MB_CANCELTRYCONTINUE | MB_DEFBUTTON2);
    }
    HRESULT hr = CoCreateInstance(CLSID_SpVoice, NULL,
        CLSCTX_ALL, IID_ISpVoice, (void**)&pVoice);
    if (SUCCEEDED(hr)) {
        pVoice->SetVolume((USHORT)100);
        pVoice->SetRate(0);
        hr = pVoice->Speak(speakContent, 0, NULL);
        pVoice->Release();
        pVoice = NULL;
    }
    ::CoUninitialize();
}
```

表 4-3 久坐提醒功能部分程序实现

```
if (getTimestamp() - lastRestTime >= 60 * 60) {
    MessageBox(GetForegroundWindow(), TEXT("Take a break, let's do someactivity~"),
    TEXT("SitePoseMonitor"), 1);
}
```

4.3.2.5 可视化数据分析

我们为了提升用户体验，提高交互性，对实验数据进行封装，以可视化的数据进行展现。具体实现如下，利用 numpy，padas，matplotlib 和 TailWindCss 第三方库编写 Python 脚本来对坐姿评估模块收集和保存的 csv 数据文件进行数据分析，生成本地分析图表和通过 web 服务生成数据报表，如图 4-7。

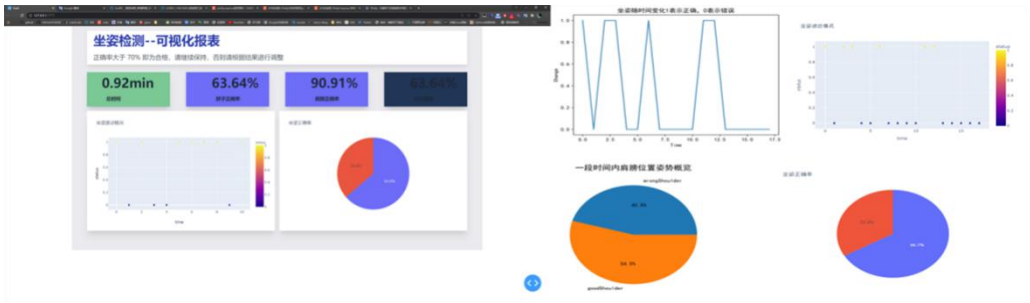


图 4-7 数据分析结果

4.4 精度和稳定性测试

对坐姿评估系统进行精度和稳定性测试，设计了两种试验方案，验证系统能否在保持高准确度的同时，仍然具有较强的稳定性。

（1）实验一：系统识别精度测试

我们邀请了三位同学进行实验，没人进行 300 帧识别，统计每个人识别个数并取平均值，计算识别率，得到结果如表 4-2 所示，平均识别率高达 96.4%，系统精度满足要求，达到预期。

表 4-2 坐姿各特征识别正确率

坐姿类别	总帧数	识别个数	识别率（%）
正确坐姿	300	289	96.3
头部左偏	300	280	93.3

头部左偏	300	293	97.6
左手托腮	300	285	95.0
右手托腮	300	291	97.0
左肩偏高	300	297	99.0
右肩偏高	300	289	96.3
平均识别率			96.4

（2）实验二：系统稳定性测试

坐姿是一种长时间的行为，为了验证该系统是否具有长时间稳定工作能力，我们邀请了四位实验者进行 40min,50min,60min,70min 的长时间高强度测试,

统计总帧数和识别正确率，如表 4-3，试验结果表明识别正确率在 96.5%左右小范围浮动，具有良好的稳定性。

表 4-3 强度测试的识别正确率

时间(min)	**总帧数**	**识别正确率(%)**
40	1200	95.4
50	1500	96.7
60	1800	96.6
70	2100	97.2

经实验分析表明，该系统仍然能在保持 95%以上的准确度的同时，具有较强的稳定性，具有较好的应用前景。

第五章 总结

本章主要总结全文和介绍我们创作过程中遇到的困难，开发感悟以及未来展望。

5.1 总结

本文在分析坐姿对人们身心健康的影响后，设计了一种基于机器视觉的人体坐姿检测系统，并有非常好的识别准确率。本文主要内容与结论有：

1.通过对坐姿形态的分析，阐述了在日常生活中保持良好坐姿的必要性，并对坐姿行为进行了分类。此外，介绍了常用的运动目标提取方法，用于视频图像中人体坐姿区域的提取。

2. 提取出图像中的坐姿区域后，对特征数据进行采集，对其中的重要特征进行提取，并对肤色和 SURF 两种特征进行融合。

3. 设计了基于 OpenPose 的坐姿评估系统，对不良坐姿进行识别，并进行提示与规劝，根据需求分析，该系统由数据采集，数据预处理，坐姿评估和可视化数据分析四个模块组成。

4. 对设计的系统进行精度和稳定性测试，分析得出在高强度作业下，该系统能在保持 95%以上的准确度的同时，具有较强的稳定性，具有较好的应用前景。

5.1 遇到的困难

在创作的过程中，从搭建环境到项目完成，我们遇见了许多大大小小的问题，比如,如何验证模型估计精度，如何实现坐姿评估算法，如何记录系统运行中的各种数据，如何设计界面，如何添加界面模块，如何添加语音模块，如何利用分析坐姿数据进行数据分析，框架 API 有哪些等等，虽然有这么多拦路虎，但是通过自己的苦心钻研和导师的指点迷津都一一解决了，其中坐姿估计算法设计，语音提醒模块，数据分析模块都花费了我们大量的心血，在做评估算法选择时起初采用的是角度方案，结果不尽人意，查阅大量论文书籍后，我们结合平面几何

知识，采用角度和比例评估算法，精度大大提高，让人欣慰。在语音提醒模块中，C++广泛使用的有两种方式可以实现语音播报 winAPI 的 Beep/PlaySound 函数和 Microsoft Speech SDK,经过不断试错，发现 SDK 更加稳定和强大，最终我们选用了后者。数据分析模块可以说是我们花费心血最多的部分了，一开始提出这个想法时，我思考了很久，最终设计了一套解决方案，将头部，肩部，颈部，坐姿时长，检测次数设置为全局变量，放置在合适的位置，在系统运行过程中，不断导出 csv 格式的数据，通过编写好的 python 脚本进行数据分析，该 python 脚本有 400 多行，采用了数据分析三剑客 numpy,padas,matplotlib 和 TailWindCss 框架，可生成可视化报表和用户坐姿的饼图，折线图，散点图，饼图等。为了实现一个较为复杂的功能，需要用到很多知识，经过不断地学习思考，想出解决方法，这锻炼了我们的思维能力以及综合运用知识的能力，过程往往比结果更重要，是这几周做项目的一个很重要的收获。

5.2 展望

本文针对计算机视觉在坐姿识别方向的应用进行了坐姿识别方法的研究，设计了基于 OpenPose 的坐姿检测系统，并取得了一定的成果。但是随着更加深入的研究，发现该系统还存在一些不足之处，还需要再今后的工作中进一步研究和完善。具体有以下几个方面：

（1）采用更轻量级的框架，将系统集成到嵌入式设备中，使该系统更加小型化、轻量化，便于广泛使用。

（2）采用多角度多机位对目标进行数据采集，提升检测精度。

（3）如今深度学习的研究和发展非常火热，并在目标识别等邻域中取得了不错的成绩，将深度学习应用到坐姿识别中也是今后的研究方向之一。

参考文献

[1]. 孙幸欣，周峠，姜斌，等. 办公座椅坐姿行为的聚类与分析[J]. 林业工程学报, 2018,3(5) : 158 - 164. DOI: 10.13360/j.issn.2096 - 1359.2018.05.02.

[2]. CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity filds[C]//2017 IEEE Confer-ence on Computer Vision and Pattern Recognition (CVPR), Iuly 21-26, 2017.Honolulu, HI. New York, USA: IEEE, 2017. DOI:10.1109/cvpr.2017.143.

[3]. 张鸿宇，刘威，许炜，等. 基于深度图像的多学习者姿态识别[J]. 计算机科学, 2015,42(9):299-302.

[4]. 教育部等八部门关于印发《综合防控儿童青少年近视实施方案》的通知[J].中华人民共和国国务院公报,2019(03):29-34.

[5]. Paul, N.1;Singh, A.1;Midya, A.1;Roy, P.P.2;Dogra, D.P.3.Moving object detection usingmodified temporal differencing and local fuzzy Journal of Supercomputing.2017,Vol.73(No.3)：1120-1139.

[6]. Lucas B. D., Kanade T. An iterative image registration technique with an application tostereo vision[C]. International Joint Conference on Artificial Intelligence. MorganKaufmann Publishers Inc. 1981.

[7]. 徐君妍,袁址赟,崔宗勇,曹宗杰.基于帧间差分的 ViB e 运动目标检测[J].科学技术与工程,2017,17(27):82-87.

[8]. 张广栋.拥挤人群异常情况探测研究[D].吉林:长春理工大学,2017.

[9]. 袁益琴,何国金,王桂周,江威,康金忠.背景差分与帧间差分相融合的遥感卫星视频运动车辆检测方法[J].中国科学院大学学报,2018,35(01):50-58.

[10]. Zivkovic, Zoran1;van der Heijden, Ferdinand2.Efficient adaptive density estimation perimage pixel for the task of background subtraction[J].

[11]. Lopez, P;Lira, J.;Hein, I..Discrimination of ceramic types using digital image processingby means of morphological filters[J].Archaeometry.2015,Vol.57(No.1)：146-162.

[12]. 孟建军,程思柳,李德仓.基于形态学处理的轨道扣件定位算法研究[J].计算机仿真,2019,36(11):105-109+170.

[13]. 何希,吴炎桃,邸臻炜, 陈佳 .基于图形 处理器 的形态学重 建系 统[J].计算机应用,2019,39(07):2008-2013.

[14]. 李帅,侯德华,高杰,童峥.基于数学形态学的路面裂缝图像处理技术[J].公路工程,2018,43(02):270-274.

[15]. 周乐前,郭斯羽,温和,张翌.大结构元素二值形态学基本操作改进算法[J].计算机工程与应用,2016,52(01):190-194.

[16]. 陈锻生，刘政凯. 肤色检测技术综述[J]. 计算机学报, 2006, 29(2):194-207

[17]. M Hubert，PJ Rousseeuw，KV Branden.ROBPCA: A New Approach to Robust Principal Component Analysis[J]. Technometrics, 2010, 47(1):64-79.

[18]. Josef Sivic,Mark Everingham,Andrew Zisserman. Person spotting:v-ideo shot retrival for face sets[C].ACM Conference on Image and Video Retrieval (CIVR).Lecture Notes inComputer Science, 2005:226-236.