

便携式复杂背景下标准化高清面舌图像采集装置研究

胡真晓^{1,3}, 杨朝阳^{1*}, 吴长汶³, 伍明诚², 王仁平², 阴亚东³

(1. 福建中医药大学中医证研究基地, 福建 福州 350122;

2. 福州大学物理与信息工程学院, 福建 福州 350108; 3. 福建省第三人民医院, 福建 福州 350122)

【摘要】目的 加快中医望诊装置在社区家庭等场景中的应用普及。**方法** 开展了复杂背景下便携式标准化高清面舌图像采集装置研究。完成了装置的硬件设计, 其尺寸为 $20 \times 22 \times 23 \text{ cm}^3$, 达到了便携性要求。基于现有 Mediapipe 人脸特征检测算法和飞行时间 (ToF) 测距算法, 提出了一种智能感知算法以获取使用考脸部区域的纵深、水平、垂直等 3 维度的位置信息; 提出一种实时摄像头位移动速算法以实现其快速升降; 将查找表与爬坡算法相结合, 实现了一种摄像头快速自动变焦对焦算法; 基于 Haar 级联分类器完成了一款轻量级舌体检测模型的设计; 实验结果显示, 在同等硬件资源下本文舌体检测模型准确率和漏检率基本与 YOLOv5 算法相当, 但其速度提升了约 23 倍。**结果** 通过装置的语音引导和摄像头位置自适应调整, 30 位未经培训人员能够快速完成位姿纠正及面舌像的标准化图像采集, 平均采集时间为 33 秒。**结论** 为中医望诊面舌图像采集装置的推广提供了一种新思路。

【关键词】 中医面舌诊; 图像识别; Haar 级联分类器; 位姿纠正

DOI 标识: doi:10.3969/j.issn.1008-0805.2024.15.34

【中图分类号】R-331 **【文献标识码】**B **【文章编号】**1008-0805(2024)15-3493-05

中医诊疗装置现代化一直是中医现代化的重要组成部分。在业界的不懈努力下中医诊疗设备产品层出不穷, 其中不乏设计优秀的产品已进入临床使用并取得了良好效果^[1,2]。随着电子信息技术发展, 中医装置现代化发展迎来契机。基于大数据 + 物联网 + 人工智能的中医诊断装置现代化发展方向也已经渐渐得到业界认可和共识。获取海量中医四诊数据, 开展基于人工智能的中医诊断模型的训练完善, 使之最终达到临床使用要求, 已成为中医诊断装置现代化发展的关键核心环节。

海量中医四诊数据的获取必然要求中医诊察装置的大规模社会普及。然而, 让中医诊察装置“走出”医院“下沉”社区家庭, 实现复杂背景下的稳定运行和良好效果却并非易事。以图 1 所示的当前主流面舌像望诊装置为例分析可知, 当前中医望诊诊察装备具有以下不足之处: (1) 使用过程中需要医生等专业人员的干预; (2) 需借助采集箱以实现人脸及舌部的标准化图像采集^{3,4}; (3) 使用过程中使用者位姿受到限制, 造成使用体验不佳; (4) 装备体积庞大, 很难满足社区家庭场景的便携使用要求。有鉴于此, 针对开放环境下便携式中医面舌望诊诊察装置的设计需求, 提出了采用智能位姿引导与自适应摄像头调整相结合的方式, 解决当前装置需人工干预、采集箱限制使用者位姿等问题; 针对便携式装置算力相对不足的缺点, 设计了基于 Haar 级联分类器的舌体检测算法, 可极大地降低

嵌入式算力要求, 最终实现了开放环境下较为满意的面舌高清图像标准化采集效果和较好的用户体验。



图 1 当前主流面舌望诊装备使用场景照片

1 便携式面舌望诊仪硬件结构

如前文所述, 现有面舌像望诊诊察装置大多需要使用体积庞大的采集箱, 难以实现便携化。虽然借助手机等移动智能终端可以灵活地进行图像拍摄, 但其所采集图像需要依赖后期图像处理以达到标准化要求, 这使得后续海量图像数据的处理变得异常困难。

收稿日期: 2024-03-04; 修订日期: 2024-09-20

基金项目: 福建高校产学研合作项目 (2020Y4017); 福建省自然科学基金面上项目 (2023J01398); 福建中医药大学财政资助专项 (X2020001)

作者简介: 胡真晓 (1998-), 男 (汉族), 浙江绍兴人, 福州大学硕士研究生, 主要从事嵌入式系统研究工作。

* 通讯作者简介: 杨朝阳 (1972-), 男 (汉族), 福建莆田人, 福建中医药大学医学院教授, 主要从事中医四诊装备、中医健康管理研究工作。

有鉴于此,设计一种便携式面舌图像采集装置,其结构如图 2 所示。该装置主要由控制模块、图像采集模块、语音提示模块和显示交互模块等构成;其整体尺寸为 $20 \times 22 \times 23 \text{ cm}^3$,能够满足便携式要求;其正面采用倾斜斜面设计,能保证图像采集模块中摄像头具有足够视野的前提下减少装置高度。

图像采集模块由红外飞行时间(Time of flight, ToF)测距单元、可变焦摄像头和丝杆电机构成。可变焦摄像头可实现宽于 $\times 4$ 倍的变焦范围;与常规基于定焦摄像头的便携式装置相比,本装置采用可变焦摄像头不仅能够确保面像和舌像均能实现高清采集,还能实现图像的光学缩放以实现标准化采像。装置内图像采集模块丝杆电机平行于外表面倾斜放置而摄像头则垂直固定于丝杆的滑台上;通过滑台移动改变摄像头高度以适配不同坐高的使用者。ToF 测距单元安装于丝杆上部,可实现使用者与装置间的距离感知^[5],以辅助使用者进行位姿纠正操作。装置控制模块基于安卓操作系统,负责了图像处理与识别、摄像头焦距调整、丝杆电机运动控制、用户互动等控制算法的运行。此外,装置还具备了喇叭、触摸显示屏等以方便与使用者进行互动操作。

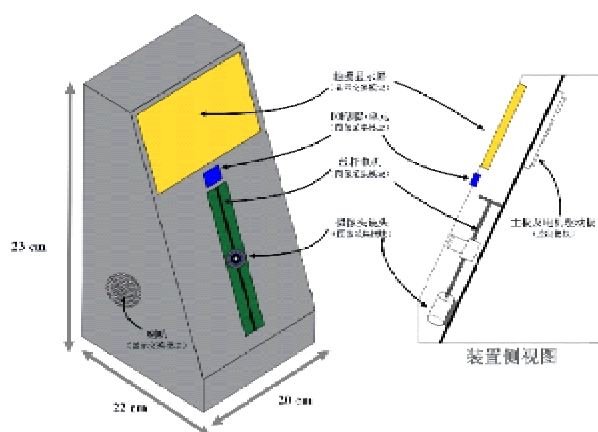


图 2 面舌望诊装置结构示意图

2 系统软件流程

本文所述便携式装置面舌图像采样总体流程与常规装置类似,相似部分流程的原理不再赘述。与常规装置不同,本装置无采集箱体的设计一方面可以减少望诊图像采集过程中对使用者位姿限制;使用者处于自然姿态采像可有效消除其心理焦虑、紧张、恐惧等因素以展现其自然神态,对精准掌握使用者健康状态具有重要意义;同时,无采样箱体的设计还能大大缩小装置尺寸,实现便携化。然而该方式无疑将引出图像色彩失真、背景干扰、内容标准化等一系列技术挑战。主要对如何解决图像内容标准化这一技术难题开展探索。针对这一难题,将采用使用者位姿纠正和图像采集模块调整两者相

结合的方式,从水平、垂直和纵深等 3 维空间进行纠正,达到面像和舌像采集内容标准化的要求。

2.1 面舌像标准化采集控制流程

使用者面像的标准化采集控制流程主要由人脸轮廓识别、纵深位置校准、水平位姿校准和垂直位姿校准等流程构成。考虑降低控制模块的算力要求,人脸检测、人脸关键点、目标检测追踪等功能则采用业界较为成熟的 Google 开源计算机视觉处理框架 Mediapipe 进行处理^[6];其能够计算得到使用者左右嘴角点、鼻尖点、左右眼眼角点等关键点的像素坐标。当 Mediapipe 检测到人脸轮廓后,将开启 ToF 模块得到使用者与装置之间的纵深距离;然后通过语音提示方式引导使用者进行纵深位置调整。当纵深距离控制在 $50 \sim 70 \text{ cm}$ 内时,纵深位置校准流程完成,此时摄像头可达到最佳摄像效果。

由于控制模块算力有限,频繁地调用 Mediapipe 将导致算力消耗过多而造成系统卡顿,影响用户体验甚至造成流程运行失败。为减少图像识别的次数,利用 Mediapipe 图像处理结果获取鼻尖点坐标后推算出面部中心的水平像素坐标,最终精确地获得其与目标像素坐标之间的水平像素距离 $D_{H,PK}$ 。根据我国颁布的《成年人头面部尺寸》数据可获知成年人两眼外宽平均值^[7],然后采用公式(1)可将水平像素距离转换为真实的水平物理距离。

$$D_{H,R} = \frac{W_{H,R}}{W_{H,PK}} D_{H,PK} \quad (1)$$

其中 $W_{H,PK}$ 为使用者两眼外宽像素距离; $W_{H,R}$ 为成年人两眼外宽物理距离,约为 97 mm 。通过公式(1)计算,可实现较为精确的水平物理距离感知,从而准确地引导使用者进行调整位置;图 3 显示了使用者水平位姿调整前后的效果,其中图 3(a)为调整前状态,图中红点表示则位姿校准时将使用到的关键点;中心红色矩形框表示鼻尖点预置区域,蓝色线表示鼻尖点与鼻尖点预置区域的水平/竖直像素差。图 3(b)为调整后的效果,此时鼻尖点处于预设矩形框内因而可认为位姿校准流程结束。通过以上步骤,可迅速指导使用者完成水平位姿校准。



图 3 水平位姿调整前后图片采样效果

使用者垂直位姿引导与水平位姿引导工作原理类

似,但实践中使用者进行垂直方向的位姿调整往往较水平方向位姿调整更为不便,因此本设计采用了使用者垂直位姿引导进行粗调后再驱动摄像头位移进行细调的方案。语音提示的使用者垂直位姿调整过程与水平方向类似,只需采用公式(1)完成图像中人脸垂直像素距离 $D_{V,EX}$ 与人脸物理垂直距离 $D_{V,R}$ 的转化即可。从成本、体积等多方面考量,摄像头位移操作中采用了较为经济的直流伺服电机驱动丝杆旋转。控制模块只能通过伺服电机运转时间来控制丝杆位移量。这不仅要求控制模块需精准地获知丝杆物理移动速度,同时还需结合不同摄像头焦距下的图像放大倍率,将丝杆物理移动速度转换为图像像素位移速度,过程较为繁琐。为此,本设计提出了一种基于丝杆位移实时测速的方案,大大地简化操作。为了与后续舌像采集操作顺利衔接,控制模块以图片中人脸左嘴角点的垂直坐标 P_{Y1} 为检测对象;通过驱动丝杆滑块移动一小段时间产生小偏量位移后,重新得到人脸左嘴角点的垂直坐标 P_{Y2} ,据此便可以得到人脸图像在像素层面上的位移速度。

$$v_{v, P\bar{X}} = \frac{\Delta D_{v, P\bar{X}}}{\Delta t} = \frac{P_{r2} - P_{r1}}{\Delta t} \quad (2)$$

进而可以计算出 $D_{t,pi}$ 所对应的直流伺服电机工作时间。

$$\ell_{\text{SH}} = \frac{D_{\text{F,PK}}}{v_{\text{F,PK}}} \quad (3)$$

最终使用者垂直位姿引导流程如图 4 所示。首先得出图像中人脸左嘴角点与采集图像中心的垂直像素距离 $D_{V, PX}$; 利用小偏量位移法计算直流伺服电机工作时间 t_{SM} 后驱动丝杆电机相应地进行滑块上摄像头的升降操作; 执行过程中, 如果摄像位移至丝杆游程终端, 则重新得出垂直像素距离 $D_{V, PX}$, 转化为物理距离 $D_{V, R}$ 并通过语音提示告知使用者。由于使用者往往很难做到精准的垂直方向位姿调整, 因此在使用者完成位姿调整后, 需重新检测 $D_{V, PX}$ 是否在预定阈值内。如果超出阈值, 则开始新一轮调整, 直至满足要求。经过使用者面像的标准化采集控制流程后, 使用者脸部将处于最佳摄像区域, 之后便可以进行摄像头的变焦对焦环节。

2.2 摄像镜头的变焦对焦算法

常规装置大多采用定焦摄像头进行面舌图像采集,其舌像往往通过裁剪面像采集结果中舌部器官部分内容而得,无法获得高清的舌像图片。与此不同,通过摄像头的变焦对焦精细操作可确保舌像图片的高清晰度。操作过程中,镜头的变焦操作依然以使用者嘴部区域为目标,确保使用者嘴部图像像素宽度达到预设值。当镜头变焦完成后进行对焦操作以获得高清晰度。由光学成像原理可知,镜头的焦距、物距和像距应

满足以下公式时才能保证摄像清晰：

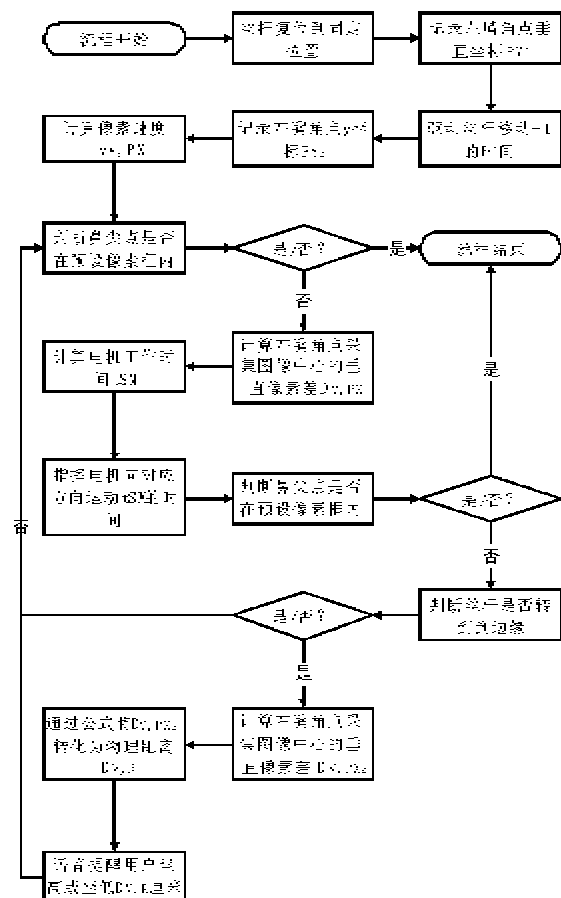


图4 使用者垂直位姿引导流程图

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (4)$$

其中 f 为镜头焦距, μ 为物距而 v 为像距。本设计中, μ 为使用者与装置之间的纵深距离; 如前文所述, 经位姿引导后, 使用者物理纵深距离 μ 大约处于 50 ~ 70 cm 区域。控制模块还需对像距 v 进行精确调整以满足公式(4)而实现清晰成像。

常规被动式自动对焦算法大多通过图像清晰度评价函数算出图像清晰度然后利用爬坡算法实现像距 μ 微调^[8];该方法调节精度高,然而需要频繁进行图像处理 and 镜头马达驱动,占用大量算力资源,往往很容易造成的系统延迟卡顿;同时由于非线性,利用爬坡算法进行大范围像距调整时很容易陷入局部最优解而操作失败。针对这种情况,本设计采用了基于查找表的粗调和基于爬坡算法的细调相结合的方式以减少算力消耗,实现对焦的精确性和快速性。其原理为,将镜头变焦区域与对焦区域各等分为 60 刻度,事先测出各刻度所对应的焦距与像距以形成查找表格。实际使用时,控制模块先通过 ToF 传感器测得当前粗略物距 μ ,然后通过查找表将当前变焦刻度换算为焦距 f ;进而根据

公式计算出目标像距 v , 利用查找表将其转变为对应对焦环刻度。基于查找表的粗调基本上仅需 1 次操作便可完成。细调算法首先通过 Tenengrad 函数得出图像的清晰度^[9]。Tenengrad 函数原理为采用 Sobel 算子提取图像的水平方向和垂直方向梯度后求平方和作为清晰度数值。对图像 $I(i, j)$ 的 Tenengrad 函数清晰度评价表达式如下:

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * I(i, j) \quad (5)$$

$$G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * I(i, j) \quad (6)$$

公式(5)(6)中的“*”表示矩阵间的卷积。

$$T(i, j) = \sqrt{G_x^2 + G_y^2} \quad (7)$$

$T(i, j)$ 为卷积结果的平方和。

$$s = \sum_i \sum_j T(i, j) \quad (8)$$

公式(8)中的 s 即为图像清晰度值。根据 s 值可使用爬坡算法对像距 v 进行微调来对焦。

采用基于查找表的粗调和基于爬坡算法的细调相结合的方式, 大大地减少对焦操作次数, 有效降低了算力消耗; 同时, 利用查找表粗调可将后续细调的范围缩减至极小范围, 有效的避免了爬坡算法陷入局部最优解的弊端, 提升了细调成功率。

2.3 基于 Haar 级联分类器的舌体检测算法

当前业界成熟的舌体检测算法较为鲜见, 因此进行了舌体检测算法的开发以完成后期舌像检测、裁剪分割等图像处理。当前图像识别技术发展日新月异, 支持向量机(support vector machines, SVM)、AdaBoost 以及基于深度学习的 R-CNN、SSD、YOLO^[10]等图像识别算法层出不穷, 功能和性能也不断提升。然而, 从便携式装备角度出发, 图像识别算法应当综合考虑设备算力、成本、功耗和体积等因素以选择最佳方案。基于深度学习的目标检测算法往往对控制模块性能要求较高, 需要图像处理单元(GPU)实现加速来将单帧检测时间缩短至 1s 以内^[11]。综合考虑多种因素后, 选择了基于 Haar 级联分类器的舌部检测算法。Haar 级联分类器是一种基于 Haar-like 特征的算法, 其运用积分图加速计算和 Adaboost 训练的强分类器级联而实现目标检测^[12], 对硬件算力要求相对较小。图 5 为基于 Haar 级联分类器舌体检测算法的训练和检测流程。

舌体检测算法模型的训练采用 969 张 40×40 分辨率归一化灰度舌头图片作为正样本。由于具有与舌头器官类似“U”形特征的物体在实际生活中较为易见, 不胜枚举; 考虑到本装置主要应用场合为医院、社区、家居等环境, 按照与被检测目标相似但不相同的原则, 挑选了 2559 张非舌头图片作为负样本。最终以 20 级训练级数、最小

命中率 0.9999 和最大虚警率 0.5 的方式完成了训练。

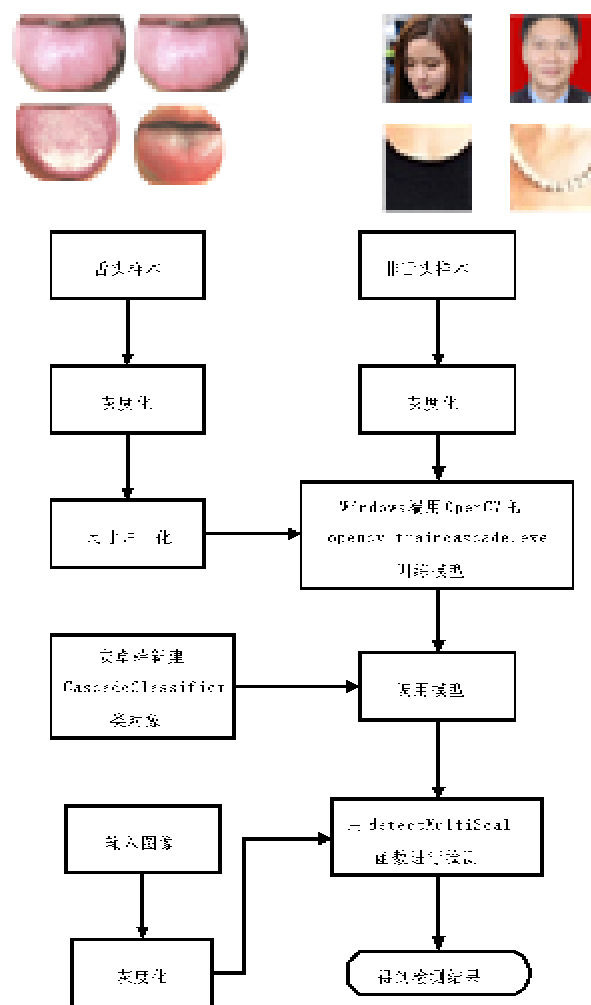


图 5 基于 Haar 级联分类器舌体检测算法的训练和检测流程

3 结果

本装置控制模块核心主要由瑞芯微(Rockchip)公司的一款主频 1.5GHz 算力 2280 DMIPS 的处理器、1GB DDR3 内存构成。为了对比验证本文所述舌部目标检测算法, 首先利用 400 张舌头图片构成训练集, 完成了一款 YOLOv5 模型的训练; 然后选取了 100 张舌头图片对本文所述检测算法和 YOLOv5 模型进行了对比测试, 具体测试结果如表 1 所示。

表 1 舌部目标检测算法性能测试结果

算法	正确/次	误检/次	漏检/次	单帧检测时间/s
本算法	82	4	14	0.058
YOLOv5	88	0	12	1.36

如表所示, 本文提出的算法与 YOLOv5 算法在正确率和漏检率上基本相当而误检率上有一定差距, 但本算法的单帧检测时间仅为 58ms, 明显优于 YOLOv5 的 1.36s, 证明了本算法在资源消耗量上优势明显。

为测试本装置的总体功能与性能, 召集了 30 名身

高介于 165 ~ 185cm 之间的志愿者参与面舌像图像采集功能测试。测试前未对志愿者进行使用培训,其初始坐姿和纵深位置随意选择。测试过程中,全程无他人干预而只由装置通过语音进行引导。实验结果显示,通过位姿引导,志愿者平均只需 3 次便完成位姿调整并达到要求。进而,完成了面舌像的图像采集操作;30 位受试者均成功完成了面舌像图像采集,采集时间在 28 ~ 57s 之间,平均为 33s。

图 6 显示了测试过程中 3 组代表性图片采集效果。其中第一行为初始状态下对 3 位志愿者的图像采

集结果,其中图片中黑框内显示的为检测到的纵深距离(单位为 cm);可以看到 3 组照片中志愿者身高、位置和初始坐姿均不同。第二行为完成位姿纠正后的面像采集结果,其图内小红框表示中心预设像素框,大红框内即为人脸轮廓像素框;可以看出,经过位姿引导和摄像模块自适应调整后,最终志愿者纵深距离均调整在规定范围(50 ~ 70cm)内,其鼻尖均与中心预设像素框区域重叠,人脸轮廓大致均等。第三行为最终舌像采集结果,由图可见,志愿者舌像清晰且舌体轮廓居中,达到了标准化采像要求。

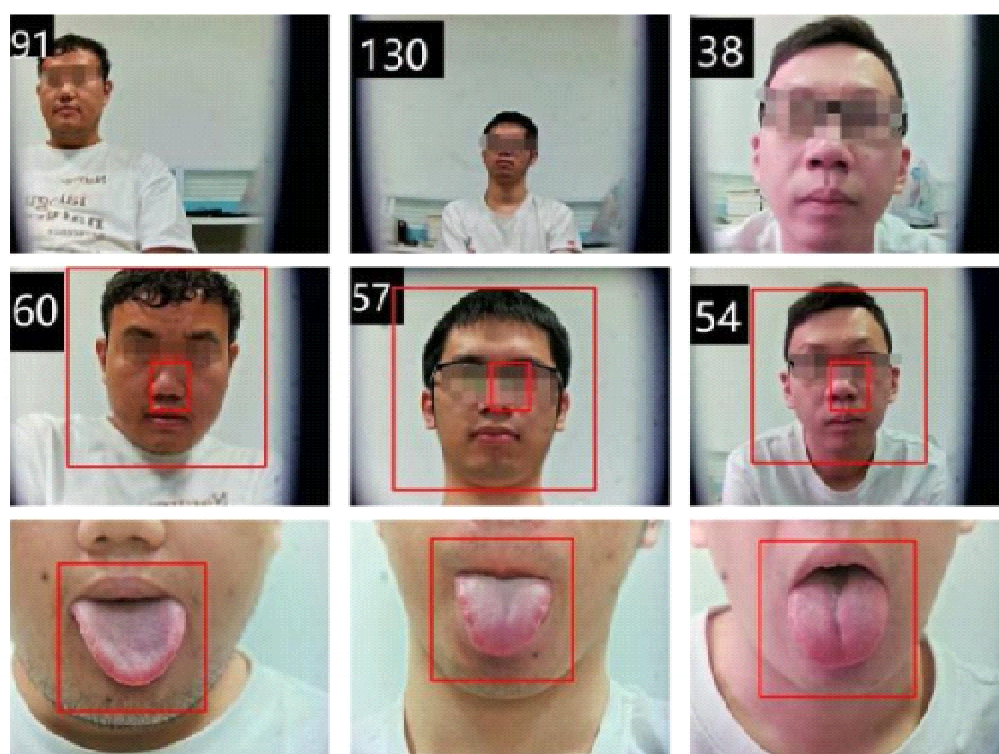


图 6 三种典型情况下的实验结果展示

4 结论

文章提出了一种新型的便携式面舌望诊设备,能让使用者以较为宽松的坐姿完成标准化高清晰度面舌诊照的采集,为中医望诊面舌图像采集装置的推广提供了一种新思路。

参考文献:

- [1] 邓佳豪,王丽珍. 基于中医切脉的远程脉诊仪设备研究与开发[J]. 医用生物力学,2021,36(81):463.
- [2] 李非,于琦,王映辉,等. 中医脉象仪应用技术研究[J]. 中华中医药杂志,2021,36(11):6839.
- [3] 王静,常佩芬,关静,等. 中医舌诊客观化研究[J]. 中国中医基础医学杂志,2022,28(9):1531.
- [4] 魏翔宇,詹松华,周华,等. 中医舌诊新技术研究进展[J]. 中国中

医基础医学杂志,2021,27(9):1519.

- [5] 吴元庆,王婷,李峰煜,等. 高转移效率时间飞行(ToF)图像传感器的像素单元优化[J]. 电子元件与材料,2022,41(8):810.
- [6] MediaPipe | Google for Developers [EB/OL]. <https://developers.google.cn/mediapipe> [Online; accessed November 23, 2023]
- [7] GB/T 2428 - 1998, 成年人头面部尺寸[S].
- [8] 车一峰,杨国舜,陈立国,等. 用于影像测量仪垂直尺寸测量的自动对焦技术[J]. 计量学报,2023,44(4):521.
- [9] 杨海,冯选璋,刘家屹等. 基于图像评价的显微镜自动对焦方法[J]. 激光与光电子学进展,2023,60(4):305.
- [10] 曹家乐,李亚利,孙汉卿,等. 基于深度学习的视觉目标检测技术综述[J]. 中国图象图形学报,2022,27(6):1697.
- [11] 李彦峰. 深度学习在 GPU 上的计算优化[J]. 信息与电脑(理论版),2018,(13):16.
- [12] Viola P, Jones M J. Robust real-time face detection [J]. Int J Comput Vis, 2004, 57(2):137.

(责任编辑:吴慧康)