计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(9):159−166 [doi: 10.15888/j.cnki.csa.008718]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合手势识别和终端用户评估的交互控制①

张　宁[1], 王卫星[2], 胡宁峰[2]

[1](贵州大学 现代制造技术教育部重点实验室, 贵阳 550025)
[2](贵州大学 机械工程学院, 贵阳 550025)
通信作者: 王卫星, E-mail: wxwang1@gzu.edu.cn

摘　要: 基于 RGB 图像的手势识别因其对设备要求低、采集数据方便等在人机交互领域得到广泛的应用. 在 RGB 图像的手势识别与交互过程中, 一方面由于 RGB 的手势图像在采集过程中存在光照影响导致利用肤色信息进行手势分割的效率较低, 另一方面用户对交互手势的认知与设计师设计的手势有差异, 导致用户交互体验反馈较差. 针对这两个问题我们进行了系统性的优化: 首先把用户的认知与交互手势设计原则联系起来建立手势共识集; 其次进行手势图像的色彩平衡处理, 利用椭圆肤色模型分割手势区域; 然后将二值化手势图像输入到 MobileNet-V2 轻量化卷积神经网络进行手势识别率的计算. 手势的终端用户主观评价与手势识别技术结合可以较系统地为交互任务进行手势设计, 减少用户在实际交互过程中的认知偏差, 提高交互系统的可用性和效率.

关键词: 手势诱导; 肤色分割模型; MobileNet-V2; 手势识别; 人机交互

引用格式: 张宁,王卫星,胡宁峰.融合手势识别和终端用户评估的交互控制.计算机系统应用,2022,31(9):159–166. http://www.c-s-a.org.cn/1003-3254/8718.html

Interactive Control Based on Gesture Recognition and End User Evaluation

ZHANG Ning[1], WANG Wei-Xing[2], HU Ning-Feng[2]

[1](Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, Guizhou University, Guiyang 550025, China)
[2](School of Mechanical Engineering, Guizhou University, Guiyang 550025, China)

Abstract: Gesture recognition based on RGB images is widely used in the field of human-computer interaction because of its low requirements for equipment and convenient data collection. In the process of gesture recognition and interaction of RGB images, on the one hand, the efficiency of gesture segmentation based on skin color information is low due to the illumination influence of RGB gesture images during collection; on the other hand, the interactive gestures cognized by users are different from those designed by designers, which leads to poor feedback of users' interaction experience. In this study, we systematically optimize the above two problems. Firstly, users' cognition is linked with the interactive gesture design principles to establish a gesture consensus set. Secondly, the gesture image is subjected to color balancing, and an elliptical skin color model is used to segment the gesture area. Then, the binarized gesture images are input into a MobileNet-V2 lightweight convolutional neural network to calculate the gesture recognition rate. The combination of end-user subjective evaluation of gestures and gesture recognition technology can systematically design gestures for interactive tasks, reduce the cognitive deviation of users in the actual interaction process, and improve the usability and efficiency of interactive systems.

Key words: gesture elicitation; skin-color segmentation model; MobileNet-V2; gesture recognition; human-computer interaction (HCI)

① 基金项目: 贵州省科学技术基金 (黔科合基础 [2020]1Y262); 贵州省教育厅青年科技人才成长项目 (黔教合 KY 字 [2018]112)
收稿时间: 2021-12-26; 修改时间: 2022-01-29; 采用时间: 2022-02-18; csa 在线出版时间: 2022-06-16

© 中国科学院软件研究所 http://www.c-s-a.org.cn

人机交互 (human-computer interaction, HCI) 是关于设计、评价和供人们使用的交互式计算机系统. 在人机交互的发展过程中, 经历了从命令行界面交互阶段、图形用户界面 (GUI) 交互阶段到自然和谐的人机交互阶段. 目前基于穿戴设备的手势识别需要用户佩戴或触摸设备以感知人的手势变化[1,2], 其约束性使得人们对更友好的人机交互方式提出需求. 随着人工智能及计算机技术的快速发展, 基于视觉识别的人机交互是利用计算机输入命令形成成像系统, 替代人体视觉感知进行图像处理和分析任务. 该技术对环境的适应性较强, 也为手势识别提供了一个更加自然的人机交互接口.

手势是人与人之间最为自然的交流方式, 作为人机交互的手段具有灵活、直观的特点. Karam 等人[3] 通过对手势交互系统的研究将手势划分为了 5 类: 即指示性手势、示意性手势、操作性手势、信号性手势和手语性手势, 让手势交互系统的交互手势有了意义上的归属. 虽然手势识别在技术和方法上不断提升, 但在设计基于手势识别的交互系统时, 大多数研究是优先考虑实现交互系统的方便性, 即提高手势识别率而忽略了实际的用户需求和偏好, 导致最后交互系统可用性较低. 本文研究的目的是从优化手势的肤色分割和手势识别过程出发, 融合用户认知偏好, 为设计一个更有效的手势识别交互系统奠定基础.

1 交互手势设计理论

如今, 人机交互设计已经不仅仅是以完成交互任务为目的, 而是在这基础上更多地考虑用户的交互过程及交互体验. 娄泽华等人[4] 采用自顶向下的方法从用户、交互过程和交互系统 3 个方面总结出了 12 条手势设计原则, 为手势交互系统中的手势设计提供了一个立足点, 并为已设计的手势集提供一个综合评估标准.

虽然在现有的研究中总结出了一系列手势设计原则, 但交互设计师在设计交互手势时缺乏系统性, 用户体验感较差. 而在交互设计的早期通过了解用户的偏好和行为, 可以为设计者提供更有价值的信息, 可以实现交互系统的有用性和高效性. 因此, 用户参与式的研究被引入到手势设计之中, 减少设计师与最终用户的认知差距. 在手势设计阶段, 我们通过用户参与手势的引出来建立手势的共识集. Nacenta 等人[5] 研究发现用户定义的手势更容易被用户所记忆. Nielsen 等人[6] 提出了一种引出和开发用户定义手势的程序, 使用用户引导的程序对所研究的手势设计和整个交互设计过程提供了贡献. 根据手势设计的原则, 本文构建了 8 种经常使用手势的交互任务如表 1: 开始、暂停、完成、放大、缩小、上一步、下一步和旋转. 这 8 类交互命令通常应用在智能家居环境、智能机器人/无人机操作的移动设备、车载辅助驾驶任务等领域. 对于每个交互任务, 通过大量的用户参与进行手势设计. 最终, 相同的手势被分为一组, 选择手势相同的最大组来建立手势的共识集, 该方法的流程如图 1 所示.

表 1 交互任务类别

命令类别	手势意义
开始	让物体开始任务
暂停	让物体暂停任务
完成	完成当前所有任务
放大	实现对物体的放大任务
缩小	实现对物体的缩小任务
上一步	实现上一步骤的任务操作
下一步	实现下一步骤的任务操作
旋转	实现物体的旋转

2 RGB 图像的手势分割和识别方法

基于计算机视觉计算的图像处理在手势识别过程中有两个阶段: 手势分割和手势识别[7]. 目前, 手势分割是手势分析的第 1 步. 对于手势分割, 主要目标是将手部区域与背景分开. RGB 图像的手势估计对设备要求低且成本低廉, 在人机交互领域得到了广泛应用[8]. 王龙等人[9] 提出了一种结合肤色模型和卷积神经网络的手势识别方法. 袁敏等人[10] 采用阈值分割模型来提取手势的肤色, 通过直接选定 Cr, Cb 的阈值范围 (133≤Cr≤173, 77≤Cb≤127) 确定肤色范围, 这种提取方式简单快速, 但误检率较高. 杨红玲等人[11] 提出一种基于 YCrCb 颜色空间的肤色分割法, 去除背景后结合卷积神经网络进行手势识别, 但大多针对简单的场景且网络运算量较大. 当手势图像受到交互环境光干扰时, 简单的肤色分割模型并不能准确地分割出手势的区域. 在手势识别阶段, 随着深度学习的不断发展, 大多数目标识别网络, 诸如 Fast R-CNN[12]、Faster R-CNN[13] 和 Mask R-CNN[14] 等都是在全卷积神经网络 (FCN) 基础上更新发展而来的网络模型. 由于这些卷积神经网络在进行目标检测时会伴随着生成大量的候选区域, 最后对候选区域进行分类和回归, 计算量较大, 对数据集及设备的要求较高.

为了提高实际手势交互系统, 本文提出了椭圆肤

© 中国科学院软件研究所 http://www.c-s-a.org.cn

色建模结合轻量化卷积神经网络 (MobileNet-V2) 的方法来识别交互手势. 以手势区域的二值化灰度信息为输入, 降低原始 RGB 图像信号噪声的同时利用轻量化卷积神经网络的深度可分离卷积参数少、计算量小的特点减少手势识别模型的运算量. 手势分割和识别框架如图 2 所示.

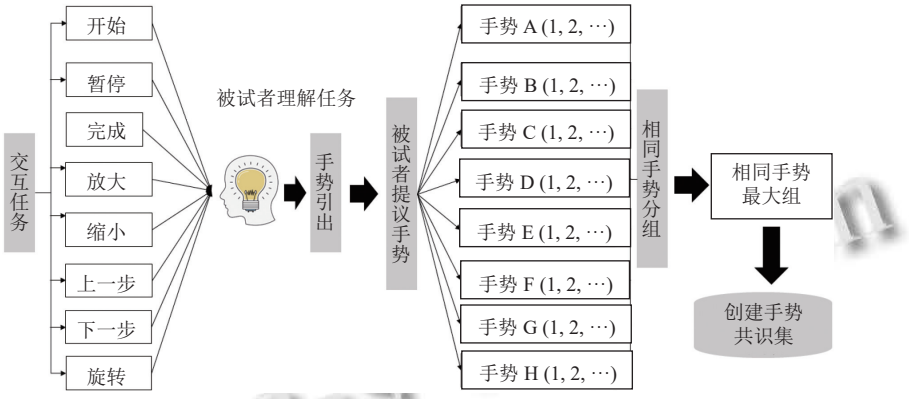

图 1 交互任务的共识手势设计

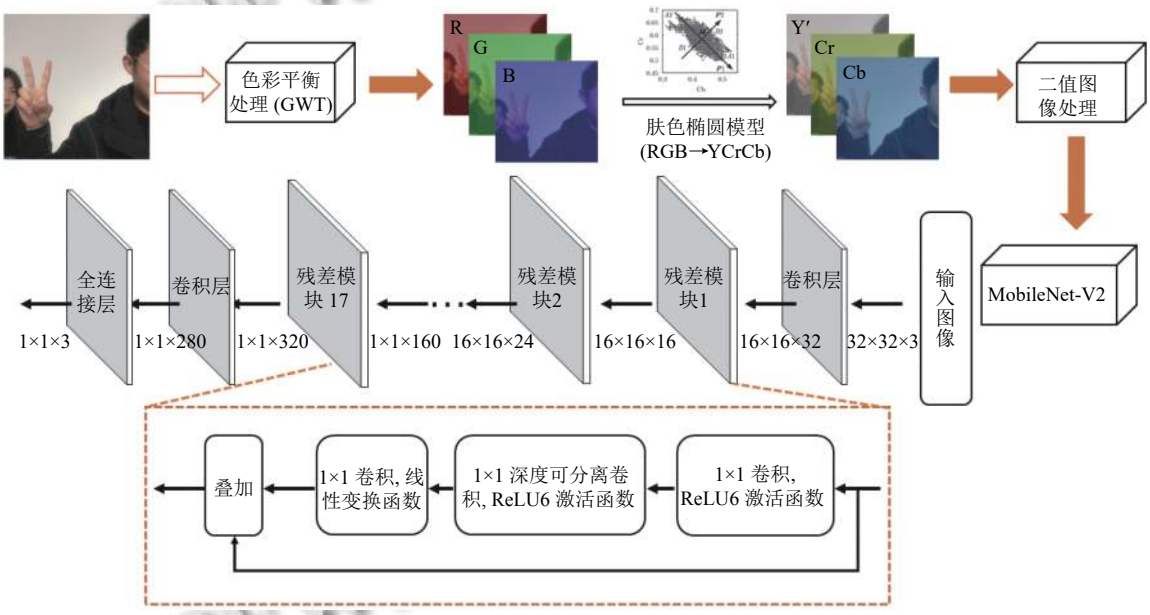

图 2 手势分割和识别框架图

2.1 RGB 图像色彩平衡处理

普通相机采集图像过程中, 由于成像设备不具备从变换的环境光照下维持图像原本颜色的功能, 导致采集的 RGB 图像颜色与真实颜色存在一定程度的色彩偏差, 选择合适的色彩平衡校正算法将有效避免光照环境对 RGB 图像颜色偏差的影响. 本文选用灰度世界算法 (gray world theory, GWT)[9] 平衡 RGB 手势图像. GWT 算法将 R、G、B 三个分量对光线平均反射的均值趋于同一灰度值, 从而可以减弱环境光的影响, 达到 RGB 图像色彩的平衡. GWT 表达公式如下:

$$R = R \times \left[\frac{K}{R_{AVG}}\right] \tag{1}$$

$$G = G \times \left[\frac{K}{G_{AVG}}\right] \tag{2}$$

$$B = B \times \left[\frac{K}{B_{AVG}}\right] \tag{3}$$

$$K = \frac{R_{AVG} + G_{AVG} + B_{AVG}}{3} \tag{4}$$

© 中国科学院软件研究所 http://www.c-s-a.org.cn

其中, R、G、B 分别为 RGB 图像的 3 个通道, R_{AVG}、G_{AVG}、B_{AVG} 为 R、G、B 三通道平均值, K 为 R、G、B 三个通道的增益系数.

2.2 椭圆肤色分割模型

肤色分割是利用颜色空间的转换分离肤色区域与背景区域的方法. YCrCb 颜色空间是一种正交颜色空间, 可以减小 RGB 颜色空间中的冗余信息. 由于 YCrCb 颜色空间可以独立统计各个颜色成分, 从而用于优化彩色信号的传输. RGB 图像的肤色信息转化到 YCrCb 颜色空间时, 肤色像素点映射在 CrCb 的二维空间中近似呈现椭圆分布, 且肤色区域的色度与亮度成非线性函数关系. 因此, 为了使肤色聚类时不受亮度 Y 的影响, 从 RGB 颜色空间转换到 YCrCb 颜色空间时去除了高光和阴影部分 (即 Y 的最大值和最小值), 将色度 Cr、Cb 进行非线性变换, 最终建立 CrCb 的椭圆肤色统计模型. 如果该模型判断 RGB 图像对应的每个像素点值在椭圆内则为肤色像素点, 否则为非肤色像素点[15]. 通过椭圆肤色模型分割手势区域, 我们得到手势的二值图像. 椭圆模型计算公式为式 (5) 和式 (6), 肤色判别公式为式 (7).

$$\frac{(x-eC_x)^2}{a^2}+\frac{(y-eC_y)^2}{b^2}=1 \tag{5}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} C'_b-C_x \\ C'_r-C_y \end{bmatrix} \tag{6}$$

$$D\left(C'_b,C'_r\right)=\begin{cases} 1, & \dfrac{(x-ec_x)^2}{a^2}+\dfrac{\left(y-ec_y\right)^2}{b^2}\leqslant 1 \\ 0, & \text{其他} \end{cases} \tag{7}$$

其中, C'_r、C'_b 分别代表 RGB 输入信号的红色部分和蓝色部分与 RGB 信号亮度值之间的差异量; C_x、C_y 代表椭圆内肤色像素点的坐标; a、b 分别表示椭圆的长半轴和短半轴; θ 表示椭圆在水平方向上的旋转角度.

2.3 手势轮廓的重建

经过上述肤色分割过程得到手势的二值图像, 但手势区域边缘可能会存在大小不一的空洞或者是其他非手势区域的现象, 这将会影响手势特征地提取, 因此我们重建手势肤色区域. 由于现实中使用相机采集手势时, 用户做出的手势常常位于身体前方, 且靠近相机的摄像头, 因此手势区域在整个图像中占比较大. 首先, 采用 OpenCV 中 cv2.findContours 函数查找二值图像中所有轮廓; 其次, 用 lambda 表达式对轮廓区域面积大小进行排序, 以轮廓面积为参数找到最大轮廓并绘制最大轮廓的外接矩形; 最后, 利用形态学的闭运算, 设置结构为 3×3 的全 1 矩阵对二值图像进行先膨胀后腐蚀的过程, 将最大轮廓区域填充为白色, 其余区域为黑色, 并平滑区域边界.

2.4 基于二值化手势图像的手势识别网络

在手势识别阶段, 采用轻量级卷积神经网络 MobileNet-V2 对分割后的手势进行特征提取和识别, 相比于传统的卷积神经网络在确保正确率的前提下大大减少了模型的参数和运算量, 可以实现在移动设备或者嵌入式设备上运行手势识别模型. MobileNet-V2 网络是在 MobileNet-V1 网络的基础上改进的[16]. 它的主要优点是可以有效地降低参数和计算复杂度. 除了中等深度的可分离结构, 它还包括一个扩展层 (expansion layer) 和一个投影层 (projection layer). 因此, MobileNet-V2 网络的整个过程由“投影层→特征提取卷积层→扩展层”变为“扩展层→特征提取卷积层→投影层”. MobileNet-V2 网络选择了 Linear Bottleneck 代替 non-linearities 非线性激活函数直接输出, 进一步防止提取的特征被 ReLU 函数压缩破坏. MobileNet-V2 的网络模块如表 2 所示.

表 2 MobileNet-V2 网络模块

输入	操作符	输出
$H\times W\times K$	扩张层, 1×1卷积结构, ReLU6激活函数	$H\times W\times(tK)$
$H\times W\times tK$	深度可分离层, 3×3卷积结构, ReLU6激活函数	$H/s\times W/s\times(tK)$
$H/s\times W/s\times tK$	投影层, 1×1卷积结构, 线性变换函数	$H/s\times W/s\times K'$

表 2 中, $H\times W$ 为通道大小, t 为通道扩张倍数, K 为卷积核大小, s 为步长. K'表示输出通道数.

3 实验设计

3.1 以用户为中心的共识手势调研

实验选取了平均年龄在 26 岁的 50 名被试 (25 男, 25 女) 进行交互手势的设计实验. 被试的惯用手为右手, 其中 35 位被试使用过手势交互系统的相关经验, 另外 15 位仅仅了解过这方面的知识. 通过向被试口头描述本实验的流程及构建的 8 个任务命令, 让被试为其设计 1 个可执行的手势. 最后我们通过拍摄视频方式收集 50 名被试引出的手势.

对于每个交互任务, 将用户引出的相同手势分为

© 中国科学院软件研究所 http://www.c-s-a.org.cn

一组, 提取相同手势数量的最大组为该交互任务的共识手势, 最后 8 类交互任务得到了 8 个共识手势. 在得到共识手势后, 50 名被试还需要对该 8 类共识手势进行主观评价进一步分析手势的可用性. 首先, 本文确定了 4 个手势的评价指标及相关描述: (1) 记忆性: 一个手势被设计一段时间后, 用户再次使用时的记忆率[5]; (2) 舒适度: 执行一个手势后, 用户身体上的难易程度感知, 以及是否会产生疲惫感[17]; (3) 直观性: 看到这个命令是否会很容易地想到与之对应的手势[6]; (4) 匹配度: 当引出一个手势时, 人们认为手势与交互任务匹配度的主观评价[18]. 其次, 进行问卷调查. 问卷包含 8 个题目, 每个题目有 5 个选项, 每项采用 Likert 量表 (非常不符合 (1 分)、不太符合 (2 分)、不确定 (3 分)、符合 (4 分) 以及非常符合 (5 分)) 收集被试对手势可用性的主观评价.

3.2 模型验证

3.2.1 数据采集

本文的实验数据采用 500 万像素的微软 LifeCam 相机采集得到, 相机模拟交互平台中的摄像头, 采集面积设置为 [900, 900].实验数据处理采用 TensorFlow 1.14.0 环境和 OpenCV-Python 3.4.2.16 的开发平台.

50 名 (25 名男性和 25 名女性) 被试者再次被邀请参加了手势的采集实验, 实验中, 被试者需要执行 8 类共识手势, 每位被试要求在操作每一种手势时变换 2 个不同角度, 每个角度的手势需采集 2 张, 共计采集的图像数据 800 张. 在收集手势实验中, 摄像机被放置在实验者的正前方 150 cm, 与视平线夹角 15°位置, 采集过程如图 3 所示. 实验前, 所有被试者被要求抬头挺胸正视前方, 胳膊放松平放在桌面上. 实验中, 被试者根据工作人员对手势特征的口头描述做出相应的手势. 被试者做每一个手势时, 手臂需要从桌面抬起 60°对准相机摄像头. 然后, 保持手臂不动, 相同的手势要求以舒适的姿势变换两个角度. 每个手势角度持续 2 s, 变换不同手势动作操作间隙休息 1 min, 以避免肌肉疲劳. 身体其他部位尽量保持静止, 便于相机能够准确采集手部姿势.

3.2.2 数据处理

大规模数据集是卷积神经网络成功应用的前提[19]. 在训练模型之前, 需要对手势数据进行扩增, 以提高网络的泛化能力. 扩增阶段采用了 Keras 的图像数据生成器策略 (image data generator). 图像生成器对每类训练的手势图片适时地进行数据增强处理, 本文采用了 8 种数据扩增处理方法: 分别是旋转变换、翻转变换、随机裁剪、偏移变换、随机噪声、调整对比度变换、USM 锐化增强和直方图均衡化变换.

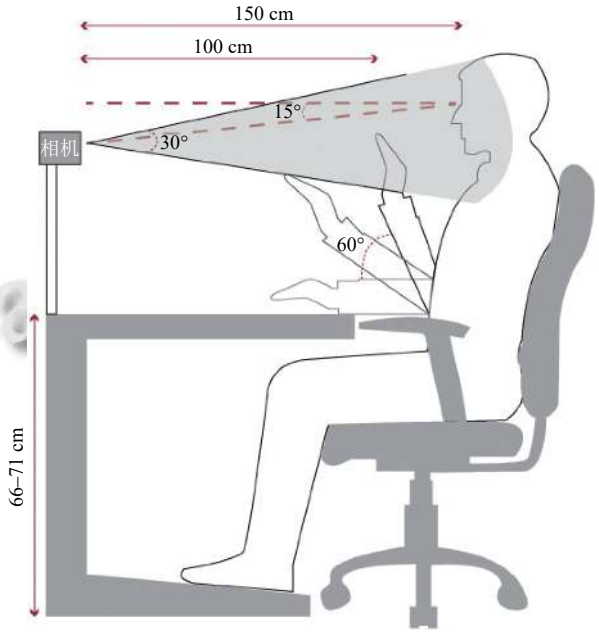

图 3 手势数据采集图

对 50 名被试采集的 800 (50×8×2) 张手势图像进行数据扩增得到 6 400 个手势图像. 其中 6 000 张图像用于模型的训练, 400 张图像用于模型的测试. 为了保证后期算法模型处理的速度以及不过度影响图像失真, 在手势图像从 RGB 颜色空间转换到 YCrCb 颜色空间时, 将手势图像统一归一化到 32×32 尺寸大小.

4 结果与分析

4.1 共识手势及主观评价

对 50 位被试进行手势设计调研后, 通过统计相同手势的分组情况, 得到最大组的手势即为该任务命令的共识手势, 本文构建的 8 个任务命令的共识手势分别如表 3 所示.

在对手势进行主观评价调查中我们共发放 50 份调查问卷, 得到有效问卷 50 份, 调查统计结果如图 4 所示. 根据样本量我们统计了交互手势的记忆性、舒适度、直观性及匹配度的平均分, 采用 Jarque-Bera 检验法结合 P 值计算对样本数据进行正态性检验, 结果如表 4 所示.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

表 3 交互任务与共识手势

交互任务	共识手势	手势特征
A放大		两指张开
B旋转		类似打电话手势
C下一步		拇指指向右
D开始执行		胜利的手势
E暂停		五指张开露出手掌
F完成		OK的手势
G缩小		手指闭合成拳头
H上一步		拇指指向左

从图 4 中可以看出, 被试者对于手势 4 个指标符合程度的接受度较好. 从表 4 中可以更具体的看出: 记忆性指标中“开始执行”手势 D 与“完成”手势 F ($P<0.05$) 呈现出显著性, 不具有正态分布特质, 原因是胜利的手势与 OK 的手势在日常生活中有相似的语义, 因此被试者认为这两个手势不易区分记忆; 匹配度指标中“完成”手势 F 呈现出显著性, 不具有正态分布特质, 原因是“开始执行”手势 D 也可以匹配“完成”命令的交互任务; 舒适度指标和直观性指标中所有手势都没有呈现出显著性, 表明被试者对手势的符合度评价较为一致. 综合来看, 用户引出的手势在实际交互任务中具有较高的符合度.

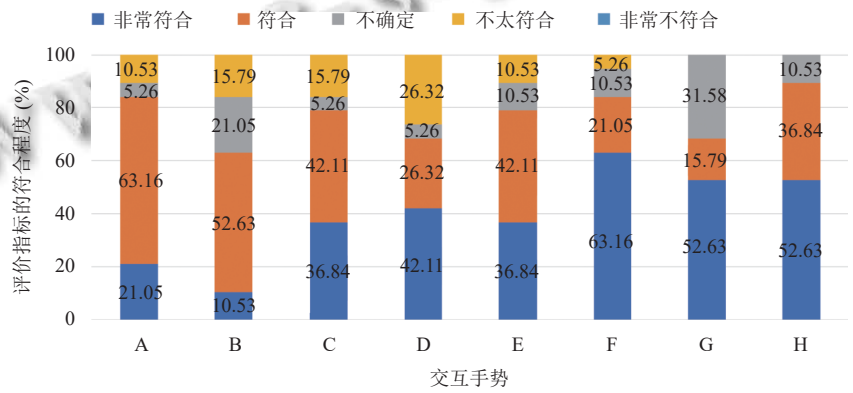

图 4 交互手势的可用性评价

表 4 交互手势主观评价及 P 值

交互任务	指标符合度平均分				P值			
	记忆性	舒适度	直观性	匹配度	记忆性	舒适度	直观性	匹配度
放大	4.05	4.53	4.16	3.95	0.281	0.205	0.117	0.137
旋转	3.47	3.95	3.63	3.58	0.636	0.090	0.649	0.620
下一步	4.05	4.47	3.79	4.00	0.179	0.167	0.336	0.282
开始执行	4.11	4.37	3.89	3.84	0.001	0.103	0.339	0.312
暂停	4.26	4.58	4.11	4.05	0.926	0.160	0.286	0.313
完成	4.84	4.58	4.63	4.42	0.002	0.204	0.291	0.033
缩小	4.05	4.63	4.11	4.21	0.372	0.104	0.311	0.268
上一步	4.16	4.68	4.11	4.42	0.191	0.135	0.145	0.366

4.2 肤色分割与手势识别结果

图 5 展示了手势的肤色检测和分割过程. 经过色彩均衡处理和椭圆肤色模型, 原始 RGB 图像图 5(a) 转换到 YCrCb 颜色空间, 得到二值化图像图 5(b), 原始图像中的肤色区域被很好地提取出来; 通过计算二值化图像中的所有轮廓的面积, 为最大轮廓绘制矩形框进一步定位手势区域如图 5(c); 通过形态学闭运算处理将最大轮廓填充为白色, 其余区域填充为黑色, 平滑手势轮廓后结果如图 5(d) 所示. 图 6 展示了本文经过处理后的 8 类共识手势的二值化手势图像.

经过肤色分割处理后, 将 6 000 张手势图像输入到 MobileNet-V2 网络中进行训练, 400 张手势数据被用于测试. 本文使用 0.000 1 学习率、Adam 优化器和交叉熵损失函数来分析训练集中手势的准确性和损失

© 中国科学院软件研究所 http://www.c-s-a.org.cn

率, 并且采用精确率、召回率、调和平均数来表示模型对 8 种不同手势的识别性能, 如图 7 所示.

(a) 原始图像 (RGB) (b) 椭圆肤色模型 (RGB→YCrCb) (c) 最大轮廓绘制 (d) 形态学处理

图 5 手势肤色分割过程

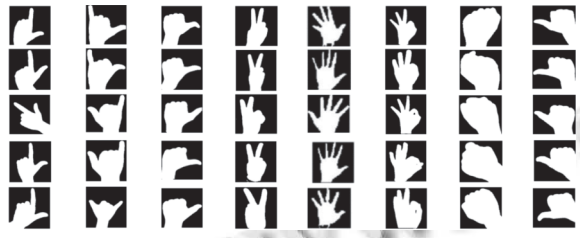

图 6 分割后的二值化手势图像

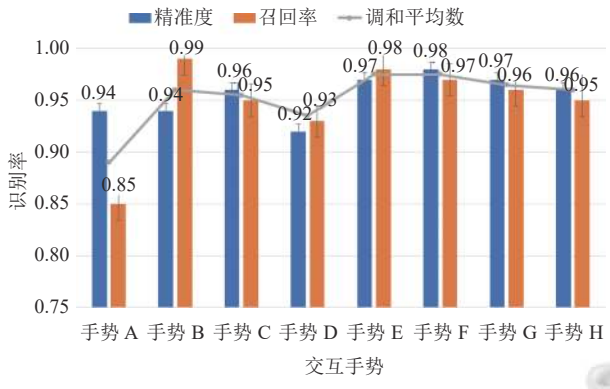

图 7 共识手势识别率

4.3 方法对比结果

将本文模型与另外 4 种方法进行比较, 第 1 种是王龙等人[9] 提出的 CNN 结合肤色模型的方法. 该方法在 Marcel-Train 标准手势数据库中对模型进行测试, Marcel-Train 数据库共有 7 种不同的手势, 每种手势包含了 3 种不同背景 (亮、暗、复杂背景) 下的图像. 首先利用肤色的高斯模型分割手势区域, 然后利用 CNN 识别手势. 第 2 种方法是杨波等人[20] 提出的复杂背景下基于手势空间分布特征的手势识别算法 HDF. 该方法利用高斯模型的分割手势的肤色区域, 综合手势特征向量的相似性来识别手势, 在环境光相对稳定下进行了手势识别实验. 其次, 我们还增加了单独 MobileNet-V2 网络和 VGG16 网络的识别率对比. 图 8 显示了上述方法在本文构建的 8 种手势数据集中识别率的比较. 从对比结果看出, 文本的椭圆肤色模型结合 MobileNet-V2 网络在基于 RGB 彩色图像的手势识别中有较好的识别率和稳定性.

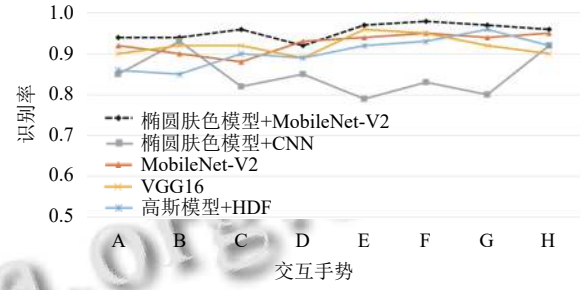

图 8 5 种方法的识别率曲线

5 结论

针对基于手势识别的人机交互系统, 存在用户对交互手势容易产生的认知偏差导致交互系统的有用性低, 本文的贡献在于从终端用户 (参与者) 出发引出交互的共识手势, 根据记忆性、直观性、舒适度和匹配度指标得到用户对交互手势的主观评价结果. 并将椭圆肤色分割模型和 MobileNet-V2 手势识别网络创新性结合, 优化了手势分割和识别过程. 从结果看出优化后的方法对 RGB 彩色手势图像有较好的分割和识别效果, 在确保识别率的前提下大大减少了网络参数和运算量, 这也为基于 RGB 手势识别的交互系统更广泛地走入人们生活提供技术和理论基础.

下一步的工作包含两个方面: (1) 本文构建的交互任务较少, 因此, 在未来的工作中将进一步扩展交互任务和手势的类别; (2) 使用本文引出的共识手势来实际开发交互界面系统, 并以实际交互任务模拟实验进行基于手势识别的交互控制系统评估.

参考文献

1 Jiang S, Gao QH, Liu HY, *et al*. A novel, co-located EMG-FMG-sensing wearable armband for hand gesture recognition. Sensors and Actuators A: Physical, 2020, 301: 111738. [doi: 10.1016/j.sna.2019.111738]

2 Xie RQ, Cao JC. Accelerometer-based hand gesture recognition by neural network and similarity matching. IEEE Sensors Journal, 2016, 16(11): 4537–4545. [doi: 10.1109/JSEN.2016.2546942]

3 Karam M, Schraefel MC. A study on the use of semaphoric

© 中国科学院软件研究所 http://www.c-s-a.org.cn

gestures to support secondary task interactions. Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems. Portland: ACM, 2005. 1961–1964.
4 娄泽华, 殷继彬. 人机交互中的手势设计原则分析. 软件导刊, 2018, 17(4): 19–24, 35.
5 Nacenta MA, Kamber Y, Qiang YZ, *et al*. Memorability of pre-designed and user-defined gesture sets. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Paris: ACM, 2013. 1099–1108.
6 Nielsen M, Störring M, Moeslund TB, *et al*. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. Proceedings of the 5th International Gesture Workshop. Genova: Springer, 2003. 409–420.
7 易靖国, 程江华, 库锡树. 视觉手势识别综述. 计算机科学, 2016, 43(S1): 103–108.
8 邢予权, 潘今一, 王伟, 等. 基于语义分割与迁移学习的手势识别. 计算机测量与控制, 2020, 28(4): 196–199, 204.
9 王龙, 刘辉, 王彬, 等. 结合肤色模型和卷积神经网络的手势识别方法. 计算机工程与应用, 2017, 53(6): 209–214. [doi: 10.3778/j.issn.1002-8331.1508-0251]
10 袁敏, 姚恒, 刘垒. 结合三帧差分和肤色椭圆模型的动态手势分割. 光电工程, 2016, 43(6): 51–56.
11 杨红玲, 宣士斌, 梁竣程, 等. 基于肤色检测与卷积神经网络的手势识别. 微型机与应用, 2017, 36(22): 58–61.
12 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1440–1448.
13 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
14 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386–397. [doi: 10.1109/TPAMI.2018.2844175]
15 Zhao SY, Song XM, Tan WJ, *et al*. A novel approach to hand gesture contour detection based on GVF snake model and skin color elliptical model. Proceedings of 2010 International Conference on Computer Application and System Modeling. Taiyuan: IEEE, 2010. V5-381−V5-384.
16 Sandler M, Howard A, Zhu ML, *et al*. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
17 Zaiti IA, Pentiuc ŞG, Vatavu RD. On free-hand TV control: Experimental results on user-elicited gestures with Leap Motion. Personal and Ubiquitous Computing, 2015, 19(5): 821–838.
18 Kühnel C, Westermann T, Hemmert F, *et al*. I'm home: Defining and evaluating a gesture set for smart-home control. International Journal of Human-computer Studies, 2011, 69(11): 693–704. [doi: 10.1016/j.ijhcs.2011.04.005]
19 杨丽梅, 李致豪. 面向人机交互的手势识别系统设计. 工业控制计算机, 2020, 33(3): 18–20, 22. [doi: 10.3969/j.issn.1001-182X.2020.03.007]
20 杨波, 宋晓娜, 冯志全, 等. 复杂背景下基于空间分布特征的手势识别算法. 计算机辅助设计与图形学学报, 2010, 22(10): 1841–1848.

(校对责编: 牛欣悦)

© 中国科学院软件研究所 http://www.c-s-a.org.cn