

重庆大学学报  
*Journal of Chongqing University*  
ISSN 1000-582X, CN 50-1044/N

## 《重庆大学学报》网络首发论文

题目: 神经心理量表理解力检测的人体姿态特征识别方法  
作者: 房欣欣, 王兵凯, 孔航, 葛学人, 杨知方, 余娟, 吕洋, 陈晨曦, 李文沅  
收稿日期: 2021-05-21  
网络首发日期: 2021-07-01  
引用格式: 房欣欣, 王兵凯, 孔航, 葛学人, 杨知方, 余娟, 吕洋, 陈晨曦, 李文沅. 神经心理量表理解力检测的人体姿态特征识别方法. 重庆大学学报.  
<https://kns.cnki.net/kcms/detail/50.1044.N.20210701.1357.002.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 神经心理量表理解力检测的人体姿态特征识别方法

房欣欣<sup>1</sup>, 王兵凯<sup>1</sup>, 孔航<sup>1</sup>, 葛学人<sup>1</sup>, 杨知方<sup>1</sup>, 余娟<sup>1</sup>, 吕洋<sup>2</sup>, 陈晨曦<sup>2</sup>,  
李文沅<sup>1</sup>

(1.重庆大学 电气工程学院, 重庆 400044;  
2. 重庆医科大学附属第一医院 老年病科, 重庆 401122)

**摘要:** 神经心理测试可以对各认知域受损严重程度做出客观评价, 是检测疾病进展、评估药物疗效的有效手段。其中理解力测试部分通过判断受试者是否根据指令要求做出相应动作实现, 是老年人认知功能障碍评估的重要部分, 有利于痴呆的早预防早干预。提出了一套神经心理测试中理解力检测的人体姿态估计视频分析方法, 基于 Openpose 深度卷积网络提取人体关键点坐标, 然后基于图像形态学处理技术和 Faster R-CNN 等技术提出了纸张、牙刷等目标物体关键点二维坐标提取方法, 并以量表中动作要求建立人体姿态估计数学模型。最后, 通过实验对神经心理测试的 6 个动作进行识别, 结果表明, 所提姿态估计数学模型和交互动作识别方法能够有效检测人体姿态动作指令及人与纸张的交互指令。

**关键词:** 神经心理测试; 理解力检测; 人体姿态动态特征; 姿态识别模型; 目标检测  
**中图分类号:** TP391.41 **文献标志码:** A

## Human posture feature recognition method for neuropsychological comprehension test

FANG Xinxin<sup>1</sup>, WANG Bingkai<sup>1</sup>, KONG Hang<sup>1</sup>, GE Xueren<sup>1</sup>, YANG Zhifang<sup>1</sup>, YU Juan<sup>1</sup>,  
LÜ Yang<sup>2</sup>, CHEN Chenxi<sup>2</sup>, LI Wenyuan<sup>1</sup>

(1. School of Electrical Engineering, Chongqing University, Chongqing 400044, P. R. China; 2.  
Department of Geriatrics, the First Affiliated Hospital of Chongqing Medical University,  
Chongqing 401122, P. R. China)

**Abstract:** Neuropsychological tests can objectively evaluate the severity of cognitive impairment, which is an effective means to detect disease progression and evaluate drug efficacy. Comprehension test is an important part of cognitive impairment assessment of the elderly, which performs assessment by whether the subjects make accurate actions according to the instructions.

收稿日期: 2021-05-21

基金项目: 重庆市在渝院士牵头科技创新引导专项 (cstc2019yszx-jscxX0004); 重庆市教委科学技术研究项目 (KJQN201900109)。

Supported by Science Innovation Programs Led by the Academicians in Chongqing under Project (cstc2019yszx-jscxX0004), the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN201900109).

作者简介: 房欣欣 (1992—), 女, 博士研究生, 主要从事深度学习应用、电力市场研究,  
(E-mail) 1187919949@qq.com。

通讯作者: 杨知方, 男, 研究员, 博士生导师, 主要从事人工智能应用、能源电力优化等研究,  
(E-mail) yangzfang@126.com。

It is also conducive to the early prevention and early intervention of dementia. A video analysis method of human posture estimation for comprehension detection in neuropsychological testing is proposed. The coordinates of key points of human body are extracted based on OpenPose. Then, based on the image morphology processing technology and Fast R-CNN, a two-dimensional coordinate extraction method is proposed for the key points of the specified target objects such as paper and toothbrush. Also, the mathematical model of human posture estimation is established. Finally, six actions of neuropsychological test are tested to verify the effectiveness. The results show that the proposed mathematical model of posture estimation and interactive action recognition method can effectively detect human posture action commands and interactive instructions.

**Keywords:** neuropsychologic test; comprehension detection; human posture dynamic feature; posture recognition model; object detection

痴呆是一种增龄性综合征，其患病率随年龄增长而迅速上升。中国 65 岁以上的老年人痴呆的患病率 5.14%，85 岁以上增至 23.66%<sup>[1]</sup>，目前已经有超过 1 000 万痴呆老人（占全球 25%），成为继心血管疾病、脑卒中和恶性肿瘤之后威胁老年人群生命健康的第四大“杀手”。阿尔茨海默病（Alzheimer's disease, AD）作为痴呆的主要类型，是以进行性认知功能障碍为主要特征的神经系统退行性病变，占痴呆的 50%~75%。在临床实践中，AD 患者认知功能损害的严重程度、疗效随访和预后随访极大依赖神经心理测试评估。

理解力测试是神经心理测试的重要项目，通过判断受试者是否根据指令要求做出相应动作实现，是老年人认知功能障碍评估的重要部分。理解力测试包含多项人体动作测试以及和目标物体的互动测试。文献[2-4]分别基于 Kinect 传感器提出了站、蹲、坐、弯腰、躺和上肢运动的识别方法。当关节点发生重叠时，文献[5]提出了修复方法用以处理关节点在一个或者多个姿势中重叠的情况。文献[6-8]基于 OpenPose 算法研究老年人跌倒行为的检测方法。文献[9-11]将人体姿态识别方法运用到患者康复训练和康复效果评估中。文献[9]提出了一套基于 OpenPose 的上肢康复系统，通过三维重建获得三维人体姿态信息并利用交互软件指导患者进行家用康复训练。文献[10]提出一种基于 OpenPose 和 Kinect 的三维人体姿态估计方法，利用霍特双参数指数膨化方法对关节点运动轨迹进行平滑和预测。文献[11]结合 OpenPose 和门控循环单元网络提出一种人体康复动作识别算法，利用注意力机制构建融合三层时序特征的 GRU 网络进行康复动作二分类。文献[12-14]将人体姿态识别算法进行改进并应用在无人驾驶、太极动作、手势检测等场景。目前已有大量不同场景的人体姿态识别研究，但是很少针对神经心理测试的测试项进行研究，尤其是其涉及受试者与目标物体之间的互动。

神经心理理解力测试包含折纸、握拳、指屋顶地板、拍肩膀、移动牙刷等指令要求。针对其特定的人体姿态识别要求及受试者与目标物体互动等问题，本文提出一种基于 OpenPose 的神经心理量表理解力测试方法，结合 Faster R-CNN 考虑目标物体的形态转变，综合建立人体姿态和目标物体的互动识别模型，完成了整套动作的评分。实验结果表明，本方法能够有效实现神经心理理解力测试中特定人体姿态识别及受试者与目标物体互动识别。

# 1 OpenPose 的人体姿态估计

## 1.1 OpenPose 的坐标获取

OpenPose 是由卡耐基梅隆大学基于卷积神经网络和监督学习开发的 C++ 开源库, 它提供了基于 COCO 数据集<sup>[15]</sup>的 18 个身体关键点、左右手各 21 个关键点、脸部 70 个关键点的坐标, 以二维坐标和置信度 $(x, y, c)$ 的形式给出, 如图 1 所示。其工作原理是: 先把输入尺寸为  $W \times H$  的二维图片经过 10 层 VGG19 网络转化为特征图, 再利用双支路卷积网络进行处理, 双支路分别进行关键点置信度  $S$  预测和关键点亲和度向量场  $L$  预测, 最后经过关键点聚类进行骨架组装得到关键点坐标。OpenPose 可以实现人体骨骼关节、手指运动、面部表情等姿态估计, 借鉴卷积姿态机 (convolutional pose machines, CPM) 中采用大卷积核获得大感受野的思想, 可以较好处理遮挡情况下的人体姿态识别问题。

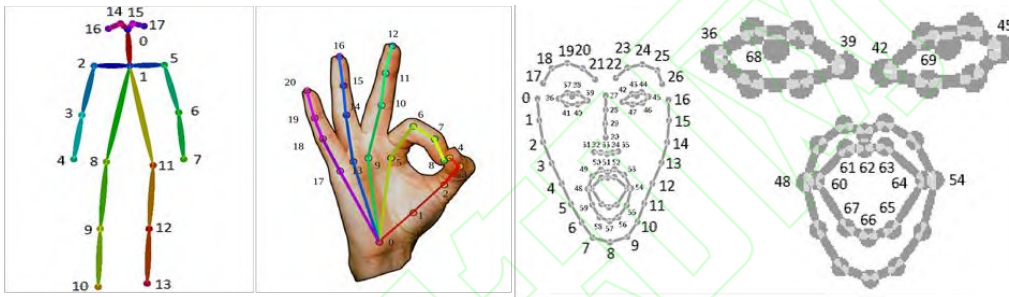


图 1 OpenPose 中身体、手部和面部关键点

Fig.1 Key points of body, hand and face detected by OpenPose

## 1.2 坐标预处理方法

在利用 OpenPose 采集坐标数据的过程中, 由于肢体交叉遮挡、拍摄角度、背景光线明暗等问题, 获得的原始坐标数据 $(x, y, c)$ 中包含有部分置信度  $c$  较低的数据和少量异常抖动数据, 对数据的准确性造成影响, 导致后续的姿态估计中极易出现误判的现象, 所以有效的数据预处理是非常必要的。

首先, 剔除置信度较低的数据, 以保证采集坐标数据整体的精度; 针对异常抖动数据, 通过中值滤波的方法消除其影响, 经过多组数据验证, 滤波窗口大小为 11 的中值滤波操作的去抖效果最好; 最后用插值的方法将剔除的数据补全<sup>[16]</sup>, 插值滤波的具体步骤如下:

由于被剔除的数据较少且不连续, 所以采用分段线性插值的方法进行补全: 分段线性插值函数记作  $I_n(x)$ , 它满足

$$I_n(x_i) = y_i \circ \quad (1)$$

$I_n(x)$  在每个区间  $[x_i, x_{i+1}]$  是线性函数,  $I_n(x)$  可表示为

$$I_n(x) = \sum_{i=0}^n y_i l_i(x), \quad (2)$$

$$l_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i] \\ 0, & x \in (-\infty, x_{i-1}) \cup (x_{i+1}, +\infty) \\ \frac{x - x_{i+1}}{x_i - x_{i+1}}, & x \in [x_i, x_{i+1}] \end{cases} \quad (3)$$

$y_i$ 为 $x_i$ 对应的实际值,  $l_i(x)$ 为插值基函数, 保证插入值的计算仅与相邻两个节点有关。 $l_n(x)$ 具有良好的收敛性,  $n$ 越大时, 与实际数值的误差越小, 能够较好地保证补全数据的可靠性。

### 1.3 姿态估计数学模型

由于人体姿态的变化都伴随着肢体的伸缩移动, 相应的关键点之间会产生距离、角度、相对位置的变化, 所以姿态识别可以通过分析人体关键点间的数学信息变化特征来实现。对预处理得到的关键点进行计算, 得到坐标数据间的欧式距离、余弦角、斜率、相对位置等信息, 通过分析神经心理测试理解力测试标准动作完成时以上信息的变化特征, 分别设置相应的阈值或区间条件, 当达到该阈值或者满足区间条件时视为实现该动作。

以完成屈肘动作为例, 说明姿态估计数学模型构建及判断流程:

1) 设 $a, b, c$ 分别为人体右肩、右肘、右腕关键点, 如图 2 所示。利用 OpenPose 获得受试者完成屈肘动作过程中各个关键点坐标值。

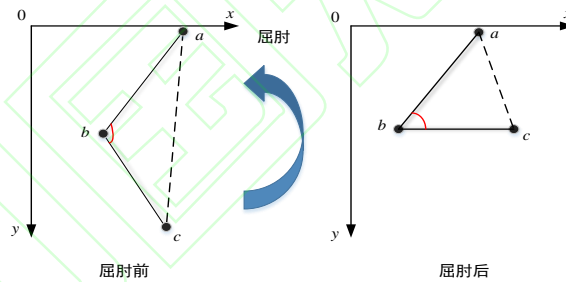


图 2 屈肘动作评分参考图

Fig.2 Elbow flexion score

通过式(3), 计算关键点之间的欧式距离 $d_{ac}$ 、余弦角 $\cos \theta_b$ 、斜率 $k_{ac}$ 相对位置 $x_{ac}, y_{ac}$ 等信息;

$$\begin{cases} d_{ac} = \sqrt{(x_a - x_c)^2 + (y_a - y_c)^2}, \cos \theta_b = \frac{d_{ab}^2 + d_{bc}^2 - d_{ac}^2}{2 \cdot d_{ab} \cdot d_{bc}} \\ k_{ac} = \frac{y_a - y_c}{x_a - x_c}, x_{ac} = x_a - x_c, y_{ac} = y_a - y_c \end{cases} \quad (4)$$

标准屈肘动作完成时 $d_{ac}$ 和 $\cos \theta_b$ 分别呈现明显的减小和增大趋势, 所以设置阈值条件 $d_{ac\_min}$ 和 $\cos \theta_{b\_max}$ ; 当受试者的 $d_{ac}$ 和 $\cos \theta_b$ 分别小于和大于上述相应阈值时, 视为完成屈肘动作。

### 1.4 交互动作的识别

在神经心理量表理解力测试中, 常常涉及到受试者和其他物体的交互, OpenPose 仅能获得人体的关键点坐标, 为了拓展 OpenPose 在交互动作识别中的应用, 需要对交互物体进行检测和定位, 结合交互物体的特点, 可以将交互物体分为以下两类: 第一类交互物体特征显著、动作过程中不涉及形态变化, 如牙刷、鼠标等, 可以使用深度学习目标检测技术对交互物体进行识别, 实现交互物体的坐标提取; 第二类交互物体特征提取困难、动作过程中发生形态变化, 如折纸动作中的纸张, 可以采用图像处理技术, 利用颜色空间变换、连通域提取等方法, 实现该类物体的定位和坐标提取。基于以上分析, 提出了以下两种提取物体坐标的方案。

#### 1.4.1 基于 Faster R-CNN 的目标检测

Faster R-CNN 是目标检测的主要框架之一<sup>[17]</sup>, 相比于 YOLO、SSD 等框架其精度更高<sup>[18]</sup>, Faster R-CNN 的整体结构如图 3 所示。

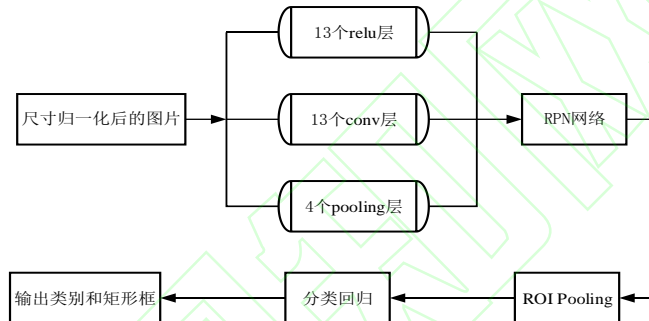


图 3 Faster R-CNN 整体结构

Fig.3 Whole structure of R-CNN

Faster R-CNN 由 4 部分结构组成: 首先, 卷积层提取输出图片特征, 得到特征图; 然后通过 RPN 网络输出多个推荐候选区域; 之后通过 ROI Pooling 将不同大小的输入转化为固定长度的输出; 最后分类回归, 输出候选区域所属的类及其在图像中的精确位置。

本 Tensorflow Object Detection API 中提供的 Faster R-CNN 框架和 COCO 数据集 (包含 80 个类别) 实现目标检测, 可以将待识别物体以矩形框加置信度标注的形式框选出来, 同时得到归一化后的矩形框左上角坐标 $(x_{min}, y_{min})$ 和右下角的坐标 $(x_{max}, y_{max})$ , 其参考坐标系与 OpenPose 的参考坐标系一致, 如图 4 所示。

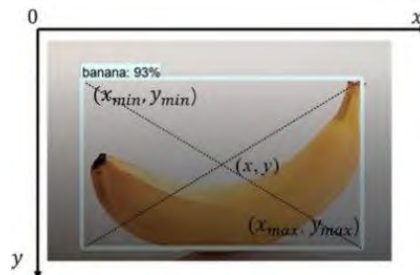


图 4 目标检测结果

Fig. 4 Target detection results

结合量表测试的需要,做出了以下两方面的改进:一方面,把归一化的像素坐标值转化为实际的像素坐标,同时用矩形框中心坐标代替物体坐标,简化分析过程,可以用式(4)计算矩形框的中心坐标 $(x, y)$ 来代替被检测物体:

$$\begin{cases} x = \frac{(x_{\max} + x_{\min})}{2} \cdot \text{img\_width} \\ y = \frac{(y_{\max} + y_{\min})}{2} \cdot \text{img\_height} \end{cases} \quad (5)$$

另一方面,AD量表涉及到的交互物体中的卡片虽然不包含在COCO数据集类别中,但是由于卡片在动作过程中不发生折叠等形态变化,可以通过在卡片上绘制香蕉等COCO数据集中的物体的图案,通过检测香蕉间接实现对卡片的检测和坐标提取。

#### 1.4.2 基于图像形态学处理技术的纸张坐标提取算法

对于纸张这类物体由于特征提取困难,难以使用目标检测方法进行坐标提取,所以通常从数字图像技术的角度进行处理。传统数字图像技术中的坐标提取方法是将纸张用固定形状的矩形框定位出来,但是当纸张的形状、位置、颜色等发生变化或受到遮挡时,无法准确提取出纸张边角坐标和面积变化情况,本文提出基于图像形态学处理技术的纸张坐标提取方法,如图5所示。

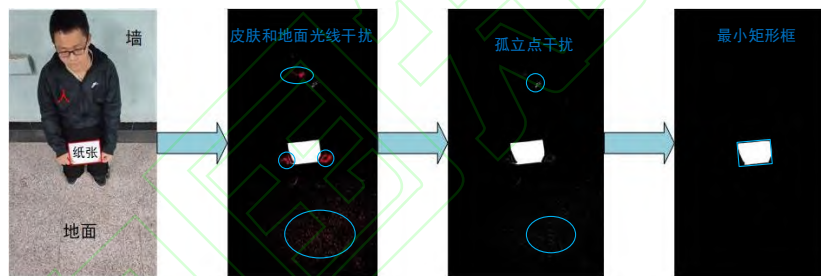


图5 纸张坐标提取流程图

Fig.5 Extraction process the paper coordinate

具体表述如下:

1) 二值化处理,由于纸张区域为明显的白色,像素值接近255,合理设置[210,255]的像素阈值,消除大部分的背景干扰,但仍有人体皮肤和地面反射光线干扰的存在。

2) 为了消除皮肤的干扰<sup>[19]</sup>,将图片像素信息从RGB映射到[Y Cb Cr]颜色空间,映射公式为

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + [K_r \quad K_g \quad K_b] \begin{bmatrix} R \\ G \\ B \end{bmatrix}, [K_r \quad K_g \quad K_b] = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \quad (6)$$

[Y Cb Cr]颜色空间是一种常用的肤色检测的色彩模型,其中Y表亮度,Cb代表光源中的蓝色分量,Cr代表光源中的红色分量。人的肤色在外观上的差异是由色度引起的,不同人的肤色分布集中在较小的区域内。肤色的[Y Cr Cb]颜色空间[Cr Cb]平面分布在近似的椭圆区域

内, 通过判断当前像素点的[Cr Cb]是否落在肤色分布的椭圆区域内, 就可以很容易地确认当前像素点是否属于肤色, 由此滤除皮肤的影响。

3) 在消除皮肤干扰后, 可以发现图像中还存在部分孤立点, 可以对图像进行开闭运算进行消除<sup>[20]</sup>。首先使用开运算, 对图像先腐蚀运算, 再膨胀运算, 消除孤立的小点; 由于手的遮挡导致纸张部分的图像不完整, 所以可以在开运算之后进行闭运算, 对图像先膨胀运算, 再腐蚀运算, 以填充凹角。

4) 用最小矩形框将图像中最大连通域, 即纸张区域框选出来<sup>[21]</sup>, 这样可以准确提取出纸张 4 个边角的坐标信息。

## 2 人体姿态估计在 AD 量表理解力智能化检测中的应用

### 2.1 神经心理测试中理解力测试指令

神经心理测试量表中通过要求受试者完成指定动作来评估受试者的理解能力, 主要要求受试者完成如下动作: 1) 用右手拿纸, 再用双手对折, 将纸放在大腿上; 2) 握拳; 3) 指指屋顶, 然后指指地板; 4) 用一只手的两个手指拍每个肩膀两次; 5) 将牙刷放在卡片上面, 然后再拿回来, 将鼠标放到卡片的另一边并且将卡片翻过来。

### 2.2 识别流程

本文基于第 1 章所提人体姿态识别及目标检测方法设计了如图 6 所示的姿态评估流程。

首先通过语音给出受试者动作指令, 然后调用 2 个摄像头从不同角度同时记录受试者的动作, 将记录的视频由 OpenPose 处理后得到人体关键点二维坐标, 对所得坐标进行滤波、插值等预处理以消除数据抖动可能带来的误判, 将记录的视频由图像形态学或 API 目标检测处理后得到纸张、牙刷等交互物体的二维坐标, 对所得坐标进行平移、翻转后使之与人体关键点坐标位于同一坐标系下, 最后将预处理后的坐标输入到姿态识别数学模型中, 得到该动作的评分情况。

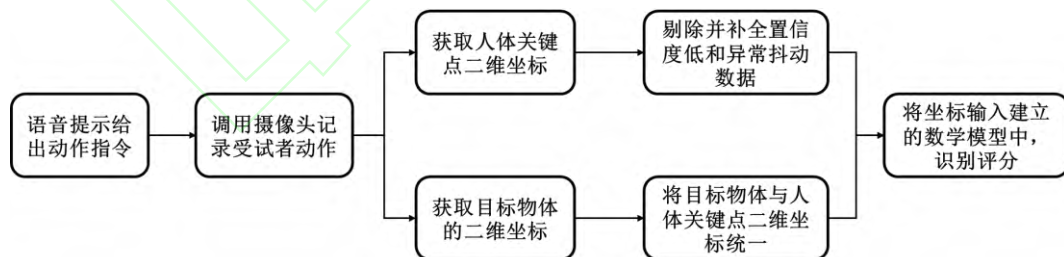


图 6 姿态识别流程

Fig. 6 Human posture recognition process

### 2.3 建立姿态识别模型

1) 用右手拿纸, 再用双手对折, 将纸放在大腿上: 需提取的关键点: 身体关键点 2, 3, 4, 5, 6, 7, 8; 纸张的 4 个顶点。右手拿纸: 当右手手腕与纸张顶点 1 的欧式距离 $d_1$ 持续 2 s (60 帧) 小于设定的距离阈值时, 判定受试者完成右手拿纸;



$$d_1 = \sqrt{(x_{b_4} - x_{p_1})^2 + (y_{b_4} - y_{p_1})^2}, \quad (7)$$

式中:  $(x_{b_i}, y_{b_i})$ 表示第*i*个身体关键点坐标;  $(x_{p_i}, y_{p_i})$ 表示第*i*个纸张顶点坐标。

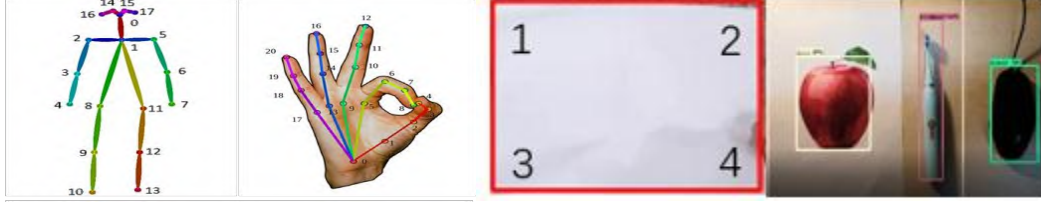


图7 人体姿态关键点、纸张、牙刷等物体关键点信息

Fig.7 Key point information of body, hand, paper, toothbrush etc

折纸: 当整个过程中纸张面积的最大值 $S_{\max}$ 和最小值 $S_{\min}$ 满足 $S_{\max} \geq 2S_{\min}$ 时, 判定受试者完成折纸。

$$a = \sqrt{(x_{p_1} - x_{p_2})^2 + (y_{p_1} - y_{p_2})^2}, \quad b = \sqrt{(x_{p_1} - x_{p_3})^2 + (y_{p_1} - y_{p_3})^2}, \quad S_{\text{paper}} = a \times b. \quad (8)$$

将纸放到大腿上: 当左右手肘夹角的余弦值 $\cos \angle 567$ ,  $\cos \angle 234$ 都出现了先增大再减小这一过程并且此后每一帧身体关键点8与纸张顶点1的欧式距离 $d_3$ 均小于设定的阈值距离时, 判定受试者完成将纸放在大腿上。

$$d_3 = \sqrt{(x_{b_8} - x_{p_1})^2 + (y_{b_8} - y_{p_1})^2}, \quad (9)$$

$$\cos \angle 567 = \frac{d_{56}^2 + d_{67}^2 - d_{57}^2}{2 \times d_{56} \times d_{67}}, \quad \cos \angle 234 = \frac{d_{23}^2 + d_{34}^2 - d_{24}^2}{2 \times d_{23} \times d_{34}}, \quad (10)$$

式中,  $d_{ij}$ 表示身体关键点*i,j*之间的欧式距离。

2) 握拳: 需提取的关键点: 右手关键点0、4、8、12、16、20。当某一帧右手所有指尖到手腕的欧式距离 $d_{i_0}$ 同时小于整个过程中最大欧式距离 $d_{i_0_{\max}}$ 的1/2 即同时满足 $d_{i_0} < 1/2d_{i_0_{\max}}$ 时, 判定受试者完成握拳。

$$d_{i_0} = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}, \quad i = 4, 8, 12, 16, 20. \quad (11)$$

3) 指指屋顶, 然后指指地板: 需提取的关键点: 右手关键点0、8。当 $\Delta y < 0$ 且 $|k_{80}| > 4$ 时判定受试者完成指指屋顶; 当 $\Delta y > 0$ 且 $|k_{80}| > 4$ 时判定受试者完成指指地板。

$$\Delta y = y_8 - y_0, \quad |k_{80}| = \left| \frac{y_8 - y_0}{x_8 - x_0} \right|, \quad (12)$$

式中:  $\Delta y$ 表示食指指尖与手腕在*y*方向的相对位置;  $k_{80}$ 表示食指指尖与手腕两点连线的斜率。

4) 用一只手的2个手指拍每个肩膀2次: 需提取的关键点: 身体关键点2, 3, 4, 5; 右手关键点0、4、8、12、16、20。拍左肩: 当某一帧右手手腕到左肩的欧式距离小于设定

的距离阈值并且右手肘、右手腕、左肩 3 点共线即满足 $|\Delta k_1| < 0.05$  且  $d_{45} < 100$ 时, 判断受试者完成拍左肩动作。拍右肩: 当右手肘和右手腕的连线近似垂直并且右手手腕到右肩的欧式距离小于设定的距离阈值即满足条件 $|k_{34}| > 10$  且  $d_{24} < 100$ 时判断受试者完成拍右肩动作。

$$d_{24} = \sqrt{(x_2 - x_4)^2 + (y_2 - y_4)^2}, \quad d_{45} = \sqrt{(x_4 - x_5)^2 + (y_4 - y_5)^2}, \quad (13)$$

$$k_{ij} = \frac{y_i - y_j}{x_i - x_j} (i, j = 2, 4; 3, 4; 4, 5), \quad \Delta k_1 = k_{34} - k_{45}。 \quad (14)$$

5) 将牙刷放在卡片上面, 然后再拿回来: 当牙刷与卡片的横坐标之差 $\Delta x_{c_t}$ 先减小后增大并且前后 $\Delta x_{c_t}$ 的符号相同时, 判定受试者完成该动作。将鼠标放到卡片的另一边并且将卡片翻过来: 当 $\Delta x_{c_m}$ 的正负发生变化情况即可判定受试者将鼠标放到了卡片的另一边。卡片正面和背面分别印有苹果和香蕉的图案。当苹果的坐标消失, 香蕉的坐标出现时, 判定受试者完成将卡片翻过来。

$$\Delta x_{c_t} = x_c - x_t, \quad \Delta x_{c_m} = x_c - x_m, \quad (15)$$

式中:  $x_c$ 为卡片的横坐标;  $x_t$ 为牙刷的横坐标;  $x_m$ 为鼠标的横坐标;  $\Delta x_{c_m}$ 代表鼠标与卡片的相对位置。

### 3 实验结果及分析

本实验的硬件系统主要由两个像素为 800 万, 最高分辨率为 3 264×2 448 的摄像头构成, 其中一个摄像头用于正面拍摄, 另一个摄像头用于俯视拍摄。软件系统主要由 OpenPose1.3.0 和 Opencv-Python3.4.5 构成。根据 AD 量表中指定受试者完成的动作的特点, 调用相应的摄像头记录受试者完成的动作, 将该视频交由 OpenPose1.3.0 和 Opencv-Python3.4.5 处理后分别得到每一帧(1 秒 30 帧)人体关键点的二维坐标和交互物体(如纸张等)的二维坐标。将得到的二维坐标经数据预处理后输入到姿态识别模型中, 输出受试者该动作的完成得分情况。

#### 3.1 数据预处理前后的影响

由于 OpenPose 在提取人体关键点二维坐标时可能出现的出现跟丢、跟错等情况, 得到的坐标置信度较低, 进而有可能造成姿态误判, 故需要对获取的坐标进行预处理操作。如图 8 所示, 在识别指指地板这一动作的过程中, 受试者已经完成了指指地板这一动作, 但是由于 OpenPose 对食指指尖的定位出现偏差(偏差点为点 8, 实际点为 8'), 通过式(15)计算出 8、0 两点连线斜率的绝对值 $|k_{80}|$ 小于实际 8'、0 两点连线斜率的绝对值 $|k_{8'0}|$ 从而误判为受试者未完成该动作。

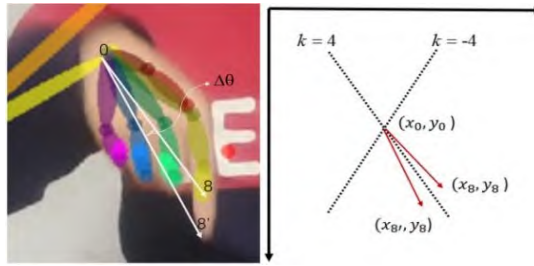


图 8 姿态误判示意图

Fig. 8 Attitude misjudgment

如表 1 所示, 通过滤波剔除第 43 帧的坐标, 再结合第 42、44 帧的坐标对其进行均值插值, 再由式(10)计算得到的  $|k_{80}|$  满足判断条件。通过滤波插值不仅提高了姿态识别的准确率, 也保证了动作的连贯性和完整性。

表 1 滤波前后的数据变化  
Table 1 Data changes before and after filtering

	帧	x	y	置信度	$ k_{80} $
滤波前	42	674.65	752.97	0.36	19.4
	43	652.19	691.78	0.09	2.58
	44	658.87	756.15	0.43	16.2
滤波后	42	674.65	752.75	—	19.4
	43	666.76	754.56	—	7.1
	44	658.87	756.15	—	16.2

### 3.2 姿态识别结果分析

为了验证所建姿态识别模型的准确度, 本部分测试了神经心理测试的所有指令。实验结果将对“用右手拿纸, 再用双手对折, 将纸放在大腿上”这一动作的识别结果进行详细分析, 图 9 为识别过程。

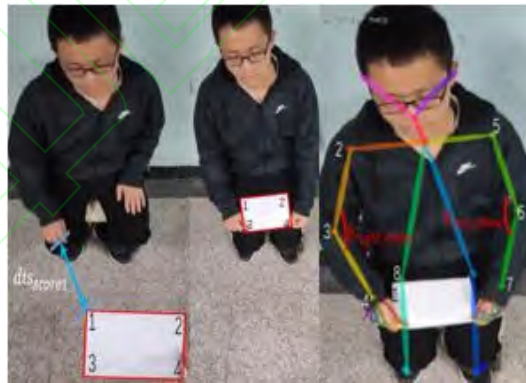


图 9 姿态识别过程

Fig.9 Human posture recognition

右手手腕到纸张的欧式距离( $d_1$ )如图 10 所示, 蓝色段曲线( $d_1$ )从 400 mm 左右快速下降到 170 mm 左右, 此过程对应受试者伸右手拿纸的过程, 橙色段曲线( $d_1$ )持续小于设定的距离阈值, 对应着受试者拿到纸后折纸和放纸的过程。受试者拿到纸之后,  $d_1$  出现了连续的波动, 这是因为受试者在折纸的过程中纸张顶点 1 的位置会随机跳动; 在后半段  $d_1$  再次出现了轻微的下降, 是因为在放纸的过程中, 手与纸张同俯摄像头之间距离增大。

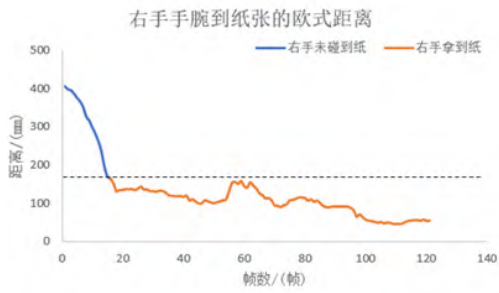


图 10 右手手腕到纸张的欧式距离

Fig.10 Distance from the wrist of the right hand to the paper

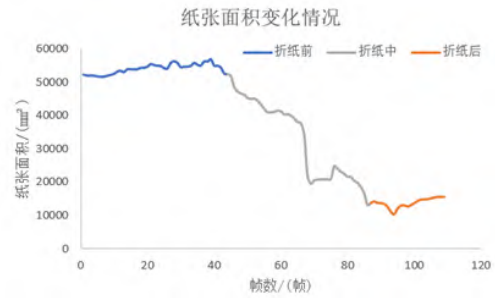


图 11 纸张面积变化

Fig.11 Changes of paper in size

纸张面积( $S$ )变化如图 11 所示, 蓝色段曲线纸张的面积 $S$ 维持在  $50\,000\text{ mm}^2$ 左右, 对应受试者取纸的过程, 在这个过程中纸张的面积基本不发生变化; 灰色段曲线纸张面积迅速减小, 对应着受试者折纸的过程; 橙色段曲线纸张的面积维持在  $15\,000\text{ mm}^2$ 左右, 对应受试者完成折纸后纸张的状态, 由于手部遮挡以及纸张与摄像头之间的距离增大等因素, 导致折纸后纸张面积小于折纸前纸张面积的一半。

图 12 中的橙色段曲线 $\cos \angle 567$ 和 $\cos \angle 234$ 增大, 则左右手肘夹角减小, 对应着受试者将纸张拿到胸前; 蓝色段曲线 $\cos \angle 567$ 和 $\cos \angle 234$ 减小, 则左右手肘夹角增大, 对应着受试者将纸张放下。此后即图 13 中红色段曲线 $d_3$ 仍小于设定的距离阈值, 判断受试者完成将纸放到大腿上。

实验结果出现了 $d_3$ 小于设定的距离阈值, 但手肘夹角余弦值未出现先变大再变小这一过程, 说明受试者完成折纸后, 纸张位于大腿的正上方, 此时还未将纸张放到大腿上。若手肘夹角余弦值出现先变大再变小这一过程后,  $d_3$ 大于设定的距离阈值, 则受试者可能将纸放在大腿的旁边或者其他位置, 此类情况判定受试者未完成该动作。

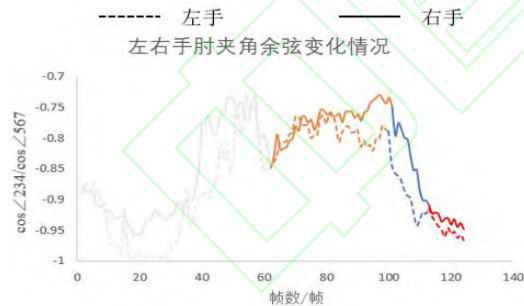


图 12 左右手肘夹角余弦变化情况

Fig.12 Cosine value of the wrist

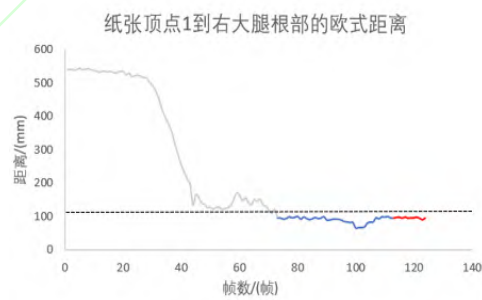


图 13 纸张到身体关键点的欧式距离

Fig.13 European distance from paper to body key

“握拳”、“指指屋顶然后指指地板”、“将牙刷放在卡片上面, 然后再拿回来”、“将鼠标放到卡片的另一边并且将卡片翻过来”“用一只手的 2 个手指拍每个肩膀 2 次”5 个动作的识别情况进行分析, 识别结果如表 2 所示。

表 2 其他动作的识别结果  
Table 2 Recognition results of other actions

识别动作	提取对应坐标	完成情况	得分
握拳			1
指指屋顶然后指指地板			1
将牙刷放在卡片上面, 然后再拿回来			1
将鼠标放到卡片的另一边并且将卡片翻过来			1
用一只手的两个手指拍每个肩膀两次			1

进一步, 选取老年受试者对本文所提方法进行了相关测试, 结果表明, 文章所提方法可以适应老年人神经心理测试量表的评估要求。

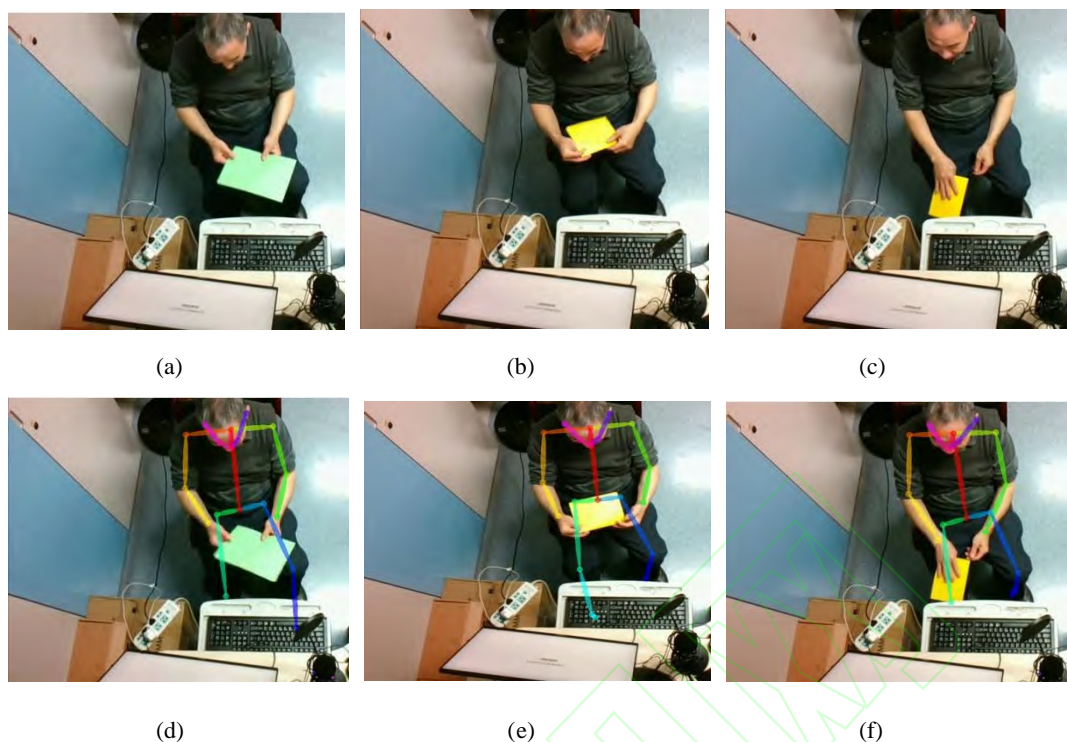


图 14 老年人姿态识别测试过程

Fig. 14 The test process for the elderly

## 4 结 论

本文针对量表中理解力测试部分提出一套基于 OpenPose 的人体姿态估计视频分析系统, 辅助传统人工 AD 诊断流程。首先利用 OpenPose 获取了人体关键点坐标, 然后基于图像形态学处理技术和 Faster R-CNN 等技术提出对纸张、牙刷等指定目标二维坐标提取方法, 接着针对量表中的具体动作建立人体姿态识别数学模型, 最后, 通过实验验证了模型的准确性和可靠性。

### 参考文献:

- [1] Jia J P, Wang F, Wei C B, et al. The prevalence of dementia in urban and rural areas of China[J]. *Alzheimer's & Dementia*, 2014, 10(1):1-9.
- [2] Wasenmüller O, Stricker D. Comparison of Kinect V1 and V2 depth images in terms of accuracy and precision[C]. *Asian Conference on Computer Vision Workshop*. Springer, 2016, 34-45.
- [3] Wang W J, Chang J W, Haung S F, et al. Human posture recognition based on images captured by the Kinect sensor[J]. *International Journal of Advanced Robotic Systems*, 2016, 13(2): 54.
- [4] Hsu S C, Huang J Y, Kao W C, et al. Human body motion parameters capturing using kinect[J]. *Machine Vision and Applications*, 2015, 26(7): 919-932.
- [5] 李昕迪, 王云龙, 何艳, 等. 基于 Kinect 的人体单关节修复算法研究[J]. *自动化技术与应用*, 2016, 35(4):100-102, 124.  
Li X D, Wang Y L, He Y, et al. Research on the algorithm of human single joint point repair based on Kinect[J]. *Techniques of Automation and Applications*, 2016, 35(4), 100-102, 124.(in Chinese)
- [6] 沈秉乾, 武志勇, 贺前华, 等. 人体姿势状态判决的跌倒检测方法[J]. *计算机应用*, 2014(Z1):223-227, 264.

- Shen B Q, Wu Z Y, He Q H, et al. Falling detection method based on human body posture judgment[J]. Journal of Computer Applications, 2014(Z1): 223-227, 264. (in Chinese)
- [7] 汪大峰. 视频监控中跌倒行为识别[J]. 电子设计工程, 2016, 24(22): 122-126.  
Wang D F. Abnormal behavior recognition of fall in surveillance video[J]. Electronic Design Engineering, 2016, 24(22):122-126. (in Chinese)
- [8] 张程, 祝凯, 赵德鹏, 等. 基于人体骨架的跌倒行为识别研究[J]. 电子技术与软件工程, 2020(23): 85-86.  
Zhang C, Zhu K, Zhao D P, et al. Fall behavior recognition based on human skeleton[J]. Electronic Technology & Software Engineering, 2020(23): 85-86. (in Chinese)
- [9] 唐心宇, 宋爱国. 人体姿态估计及在康复训练情景交互中的应用[J]. 仪器仪表学报, 2018, 39(11): 195-203.  
Tang X Y, Song A G. Human pose estimation and its implementation in scenario interaction system of rehabilitation training[J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 195-203. (in Chinese)
- [10] 王怀宇, 林艳萍, 汪方. 基于 OPENPOSE 的三维上肢康复系统[J]. 机电一体化, 2018, 24(9):31-37.  
Wang H Y, Lin Y P, Wang F. Three-dimensional upper limb rehabilitation system based on OPENPOSE[J]. Mechatronics, 2018, 24(9):31-37. (in Chinese)
- [11] 白敬, 宋爱国, 李会军, 等. 基于工作空间测量的居家脑卒中患者上肢康复训练评估系统[J]. 仪器仪表学报, 2018, 39(4): 74-81  
Bai J, Song A G, Li H J, et al. Upper extremity rehabilitation training and assessment system for home stroke patients based on workplace measurement[J]. Chinese Journal of Scientific Instrument, 2018, 39(4): 74-81. (in Chinese)
- [12] 徐彬, 郑燕萍, 曹高兴. 面向无人驾驶的行人三维姿态估计方法[J]. 科学技术与工程, 2018, 18(34): 85-91.  
Xu B, Zheng Y P, Cao G X. 3D pose estimation method of pedestrians for driverless vehicles[J]. Science Technology and Engineering, 2018,18(34):85-91. (in Chinese)
- [13] Qiao S, Wang Y, Li J. Real-time human gesture grading based on OpenPose[C]. International Congress on Image and Signal Processing. IEEE, 2017:1-6.
- [14] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017:7291-7299.
- [15] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]. In European Conference on Computer Vision. Springer, Cham: 740-755.
- [16] 司守奎, 孙玺菁. 数学建模算法与应用[M]. 北京: 国防工业出版社, 2011.  
Si S K, Sun X J. Mathematical modeling algorithms and applications[M]. Beijing: National Defense Industry Press, 2011. (in Chinese)
- [17] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [18] Rajendran S P, Shine L, Pradeep R, et al. Fast and accurate traffic sign recognition for self driving cars using retinanet based detector[C]//2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019: 784-790.
- [19] 曹建秋, 王华清, 蓝章礼. 基于改进 YCrCb 颜色空间的肤色分割[J]. 重庆交通大学学报(自然科学版), 2010, 29(3): 488-492.  
Cao J Q, Wang H Q, Lan Z L. Skin color division base on modified YCrCb color space[J]. Journal of

- Chongqing Jiaotong University (Natural Science), 2010, 29(3): 488-492. (in Chinese)
- [20] 张莹. 开闭运算在消除图象噪声中的应用研究[J]. 潍坊学院学报, 2002(2): 65-66.  
Zhang Y. The application of opening-closing operation to eliminate image noise[J]. Journal of Weifang University, 2002(2):65-66. (in Chinese)
- [21] 陈柏生. 一种二值图像连通区域标记的新方法[J]. 计算机工程与应用, 2006, 42(25):46-47.  
Chen B. A new algorithm for binary connected components labeling[J]. Computer Engineering and Applications, 2006(25):50-51. (in Chinese)

(编辑 詹燕平)

