

赵家玮, 董玉华. 基于 OpenCV 的静态手势识别研究[J]. 智能计算机与应用, 2024, 14(12): 190-194. DOI: 10.20169/j.issn.2095-2163.241228

基于 OpenCV 的静态手势识别研究

赵家玮, 董玉华

(大连民族大学 信息与通信工程学院, 辽宁 大连 116605)

摘要: 搭载手势识别的高科技产品在人们日常生活中发挥着重要作用。针对手势识别过程中背景繁杂等问题, 本文设计了基于 OpenCV 的手势识别系统。该系统采用基于 Y_C, C_b 空间 C_r 分量的 OTSU 法选取阈值对手势进行分割及二值化, 去除手势的背景信息; 在使用 Canny 边缘检测算子的基础上, 使用傅里叶描述子作为手势的特征向量, 把手势轮廓数字化; 将提取的手势特征通过 SVM 分类器构建手势模型, 从而识别手势动作所代表的数字信息。实验结果表明这一手势识别系统能够有效提取手势轮廓并且具有较高的准确率, 有助于实现人机交互的自然化和人性化。

关键词: OpenCV; 图像处理; 特征提取; SVM

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2024)12-0190-05

Research on static gesture recognition based on OpenCV

ZHAO Jiawei, DONG Yuhua

(School of Information and Communication Engineering, Dalian Minzu University, Dalian 116605, Liaoning, China)

Abstract: High-tech products equipped with gesture recognition play an important role in people's daily life. A gesture recognition system based on OpenCV is designed to solve the problem of complicated background in gesture recognition. The OTSU method based on the C_r component of Y_C, C_b space is used to select the threshold to segment and binarize the gesture, and remove the background information of the gesture. On the basis of using Canny edge detection operator, the Fourier descriptor is used as the feature vector of gesture, and the gesture contour is digitized. The extracted gesture features are used to construct a gesture model by SVM classifier, so as to identify the digital information represented by gesture actions. The experimental results show that this gesture recognition system can effectively extract gesture contours and has a high accuracy, which is helpful to realize the naturalization and humanization of human-computer interaction.

Key words: OpenCV; image processing; feature extraction; SVM

0 引言

伴随着人工智能时代的到来, 传统的人机交互已经难以满足人们日常生活的需求, 基于计算机视觉技术的人机交互方式成为研究的热点话题。手作为人体最重要和最灵活的一个部位, 能够表达意图、想法、情绪等信息, 故以手势作为交互信号, 不仅可以使得人机交互不再局限于命令行或图形界面, 还可以在很大程度上提高人机交互的效率。

许多研究学者对手势识别技术展开了研究, Kane 等学者^[1]在 2019 年提出了基于深度的形状矩阵来描述手的轮廓, 系统能够识别以短持续时间的姿态序列为特征的连续动态手势。Santos 等学者^[2]

在 2020 年提出了星 RGB 技术, 仅用颜色实现了对动态手势时空特征的表述。Mujahid 等学者^[3]在 2021 年提出了一种基于 YOLOv3 和 Dark Net-53 卷积神经网络的轻量级手势识别模型。Vladica 等学者在 2022 年提出了一种基于调频连续波 (FMCW) 雷达和合成手势特征生成器的低计算复杂度手势识别算法^[4]。李占利等学者^[5]在 2023 年提出一种 Gabor-C3D 手势识别算法, 采用多方向多尺度的 Gabor 滤波器组提取手势图像中空域和频域的纹理特征。

经过不断的研究, 手势识别技术在残障辅助、汽车行业、智能家居等方面均有广泛应用^[6]。随着该技术的不断成熟和完善, 未来会被应用于更多的行

作者简介: 赵家玮(2001—), 女, 硕士研究生, 主要研究方向: 图像信号处理。

通信作者: 董玉华(1978—), 女, 副教授, 主要研究方向: 智能信息处理与应用。Email: dongyuhua@dlmu.edu.cn。

收稿日期: 2023-07-03

业领域中,因此手势识别技术具有重要的研究价值。

1 系统框架

手势识别系统框架如图 1 所示。首先,通过 OpenCV 调用本地摄像头获取视频数据流作为系统的输入,对手势信息进行采集,采集信息包括手势的位置、大小、方向等;再对采集到的手势图像依次进

行滤波、手势分割和形态学处理等图像预处理操作^[7],去除复杂的背景环境;然后,使用傅里叶描述子作为描述手势轮廓的特征向量,提取所有手势的特征值;通过支持向量机算法对数据进行训练,以构建手势模型;最后,将待识别手势图像的傅里叶特征通过训练好的手势模型转化为系统已知的手势,并输出识别结果。

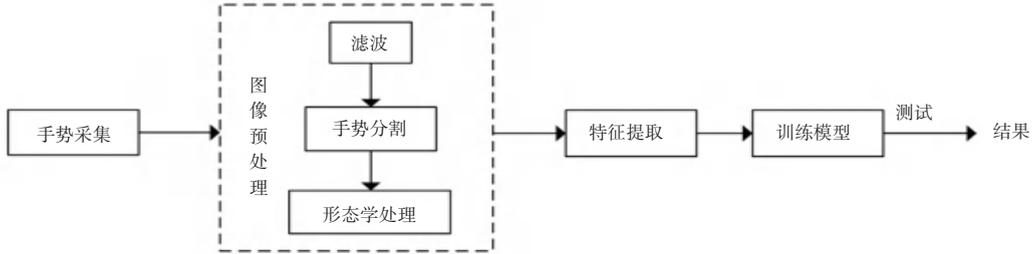


图 1 手势识别系统框架

Fig. 1 Framework of gesture recognition system

2 图像预处理

由摄像头采集到的手势图像中大部分都存在噪声和复杂的背景环境,进行图像预处理可以消除图像中无用的信息,同时最大化地保留图像的边缘和细节以提高识别的可靠性。

首先采用双边滤波去除手势图像中存在的噪声。由于采集到的视频图像序列在 RGB 颜色空间下对光照变化比较敏感,故需要进行色彩空间转换。在诸多颜色空间中, $Y C_b C_r$ 颜色空间具有很好的聚类性,并且可以拆分亮度和色度,从而忽略亮度对手势识别结果的影响。 $Y C_b C_r$ 颜色空间由 3 个参数组成,其中 Y 表示色彩的亮度, C_r 表示色调, C_b 表示色度^[8]。 $Y C_b C_r$ 与 RGB 颜色空间是线性关系,可以进行转换,具体表达式如下:

$$\begin{pmatrix} \hat{e}^Y \\ \hat{e}^{C_b} \\ \hat{e}^{C_r} \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.331 & 0.500 \\ 0.500 & -0.418 & 0.081 \end{pmatrix} \begin{pmatrix} \hat{e}^R \\ \hat{e}^G \\ \hat{e}^B \end{pmatrix} \quad (1)$$

转换至 $Y C_b C_r$ 颜色空间后提取 C_r 分量,对其采用 OTSU 算法自动选取最佳阈值进行手势分割^[9]。OTSU 算法的核心思路就是通过阈值 T 将图像的灰度数按照灰度级分为前景和背景两部分。其中,前景区域就是要提取的目标区域,该区域的灰度值均大于阈值 T ,反之则为背景区域。设定每个像素点的灰度值为 0 或者 255,遍历所有的灰度值找到最佳类间分割阈值,以此提取手势区域。手势分割过程的实现效果如图 2 所示。从图 2 中可以看出,基于 $Y C_b C_r$ 空间 C_r 分量的 OTSU 法可以消除亮度干扰信息的色彩空间,基本没有误分割和漏分割的情况,能比较完整地保留手势所存在的区域,去除背景等无关信息。

$$\begin{pmatrix} \hat{e}^Y \\ \hat{e}^{C_b} \\ \hat{e}^{C_r} \end{pmatrix} = \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} + \begin{pmatrix} \hat{e} \\ \hat{e} \\ \hat{e} \end{pmatrix}$$



(a) 原图 (b) 转换至 $Y C_b C_r$ 颜色空间 (c) 提取 C_r 分量 (d) 手势分割

图 2 手势分割过程的实现效果

Fig. 2 The effect of gesture segmentation process

当提取的手势区域不理想时,则需要通过形态学处理对该区域进行填充、细化和修剪毛刺,使提取的手势是一个完整的连通区域,从而保证后续工作的准确性。形态学处理主要是腐蚀和膨胀两个操作。腐蚀会消除图像中的边缘和杂点,减少区域的像素,膨胀会使图像中的边界扩张,增加区域的像素^[10]。这2种操作的组合使用即为开运算与闭运算,开运算是先腐蚀后膨胀,可以去除图像中细小、孤立的点,有一种“去毛刺”的效果。闭运算是先膨胀后腐蚀,能让内部的裂缝或外部孤立的点连成一体,达到“填充隙”的效果。形态学处理前后对比如图3所示。从图3中可以看出,经过形态学处理后的手势图像变得很平滑,并且空洞区域明显消失。

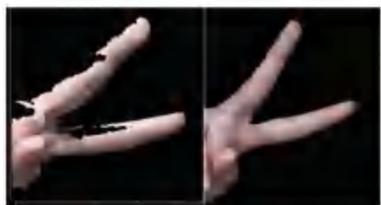


图3 形态学处理前后对比图

Fig. 3 Comparison before and after morphological treatment

3 特征提取

特征值的选取直接决定着系统识别结果的准确度,优异的特征值能够很好地描述手势轮廓的形状特征,有利于后续的手势分类与识别。本文采用傅里叶描述子作为手势的特征值,使用傅里叶描述子进行特征提取时,首先需要通过边缘检测寻找手势的边界轮廓,然后再计算边界的傅里叶特征,最后提取边界点坐标作为该手势的特征向量。

3.1 边缘检测

本文采用Canny边缘检测算子寻找手势轮廓。首先,需要对预处理后的手势图像进行高斯滤波来降噪;再计算图像的梯度幅度和方向,来估计手势边缘的强度和方向;然后,根据梯度方向,利用非极大抑制方法降低错误率;最后,创建黑色背景,使用双阈值方法筛选出边界线,选用白色线对手势边界轮廓进行绘制^[11]。经Canny边缘检测提取并绘制的手势轮廓如图4所示。从图4中可以看出,绘制的手势轮廓是一个连通的区域,拟合效果较好,没有毛刺等现象,并且形状区分度很强。

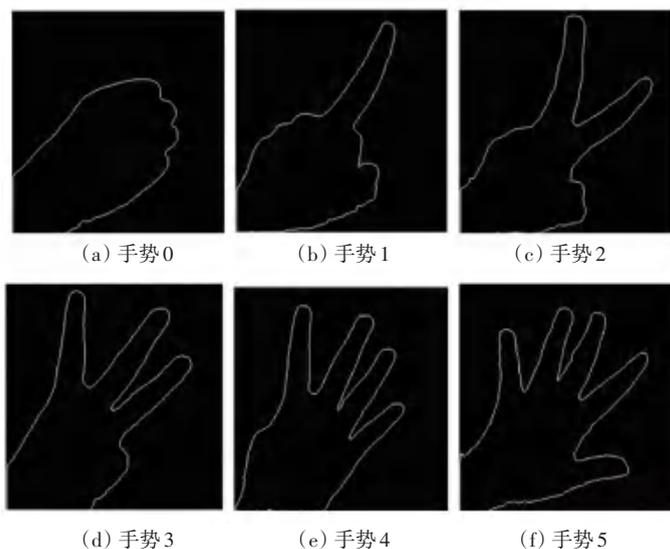


图4 绘制的手势轮廓图

Fig. 4 Gesture contour results diagram

3.2 傅里叶描述子

傅里叶描述子是一个用来描述轮廓的特征参数,只需少量的描述子就可以表示整个手势轮廓,是一个鲁棒性较好的形状特征。其基本思想是:假定物体的形状是一条封闭的曲线,这条边界曲线上动点坐标变化是一个以形状边界为周期的函数^[12],这个连续周期函数可以用傅里叶级数展开表示,表达式如下:

$$\phi(t) = \sum_{k=0}^{n-1} a_k \exp(-jkt) \quad (2)$$

其中,傅里叶级数中的系数 a_k 是与边界曲线的形状有关的,称为傅里叶描述子。对于同一种手势图像,识别结果不应因手势轮廓的位置、角度及缩放的变化而发生改变。故为了使描述子具有旋转、平移和尺度不变性,需要对傅里叶描述子进行归一化^[13],公式定义如下:

$$d(v) = \frac{|\phi(t)|}{|\phi(1)|} \quad (3)$$

其中, $t = 0, 1, \dots, n - 1; v = 0, 1, \dots, n - 1$ 。

$d(v)$ 的向量形式即为手势轮廓的傅里叶描述子特征向量。

通过对比发现, 选取少于 15 个傅里叶描述子时鲁棒性较好, 但是手势轮廓边界的拟合效果不是很

好, 无法区分不同手势的形状。选取多于 15 个傅里叶描述子时, 虽然形状区分能力较强, 但同时噪声比较敏感, 并且算法效率较低。故综合考虑, 本文提取 15 个傅里叶描述子分别来描述手势数字 0~5 的形状特征, 具体结果见表 1。将表 1 中提取的特征值经过傅里叶反变换, 可重构手势边缘, 所以该特征值对手势的描述比较准确。

表 1 15 个傅里叶描述子特征值

Table 1 Fourier descriptor eigenvalues

手势	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1 716	96	108	77	48	25	18	16	16	15	5	5	4	7	4
1	1 455	302	131	128	49	58	20	15	29	22	19	4	5	10	8
2	1 552	449	275	235	147	43	49	27	20	22	23	40	11	13	11
3	1 671	642	203	306	279	99	43	10	34	39	25	13	10	44	17
4	1 744	700	366	158	281	309	82	56	6	38	28	31	26	4	24
5	1 912	677	310	232	263	342	258	74	33	31	42	20	57	17	20

4 训练模型

4.1 创建训练集与测试集

首先需要建立手势库。由本地摄像头对手势数字 0~5 进行采集, 每种手势采集 20 张图像, 共得到 120 张图片, 然后对其进行图像预处理后, 得到只有手势区域的图像。再通过镜像和旋转两种数据增强方式对样本库中的手势进行扩充, 把每种手势图片扩充为 500 张, 总共生成了 3 000 张手势图像。提取上述图像增强后的手势库中的傅里叶描述子, 并以文本格式保存。规定测试集与训练集数据按照 1:4 的比例分配, 即有 2 700 张手势图像用于模型训练, 300 张图像用于测试, 每种手势有 50 张。

4.2 训练手势模型

本文通过支持向量机算法构建手势模型, 从而实现手势识别。支持向量机(SVM)是一种二分类模型, 可以解决小样本下的机器学习问题, 泛化能力较强。该方法是通过将实例的特征向量映射到空间中的一些点, 从而对特征向量进行区分^[14]。而数据集在低维空间中对应的向量无法被一个超平面区分开时, 通过核函数可以将其转化到更高维度的空间中实现分类。

将提取的 15 个傅里叶描述子特征值引入 SVM 分类器中, 训练分类器生成手势模型以识别手势所代表的数字^[15]。该过程首先设置存放手势类别标签, 并创建 SVM 对象, 然后通过网格搜索法选取最优参数, 利用此参数对训练集数据进行交叉验证以训练模型, 最后保存训练最好的模型, 利用测试集数据对该模型进行测试。

4.3 测试结果及分析

在完成模型训练后, 利用分割出的 300 张测试集对训练出的手势模型进行测试, 每个手势样本测试 50 张, 6 种数字手势 SVM 模型测试结果见表 2。由表 2 中可以看出, SVM 模型对测试的 6 种手势的识别率有所不同, 这是因为不同的手势其手势特征不同, 手势特征明显时, 模型的训练效果较好, 所以比较容易识别。相对而言, 有些手势特征不明显, 导致其识别效果相对较差, 所以识别准确率较低。但总体来看, 本文训练出的 SVM 模型对手势的识别效果十分不错, 6 次测试的识别正确率约 94.83%。

表 2 数字手势 SVM 模型测试结果

Table 2 Test results of digital gesture SVM model

项目	测试样本数	识别正确数	识别错误数	识别率/%
手势 0	50	46	4	92.00
手势 1	50	45	5	90.00
手势 2	50	49	1	98.00
手势 3	50	47	3	94.00
手势 4	50	50	0	100.00
手势 5	50	48	2	96.00
总计	300	285	15	94.83

4.4 结果展示

本文使用界面程序 Pycharm 作为开发平台, 使用 PyQt5 绘制可视化界面, 该界面包括打开相机和开始预测两个按钮、由摄像头获取的用户手势图像、边缘检测后绘制的手势轮廓、以及预测结果。对手势数字 0~5 进行测试, 系统界面显示结果如图 5 所示。从图 5 中可以看到, 本文建立的手势识别系统能够有效分割手势、绘制手势轮廓, 并且经过模型训练后可以准确识别 6 种手势, 正确地显示了对应结果。

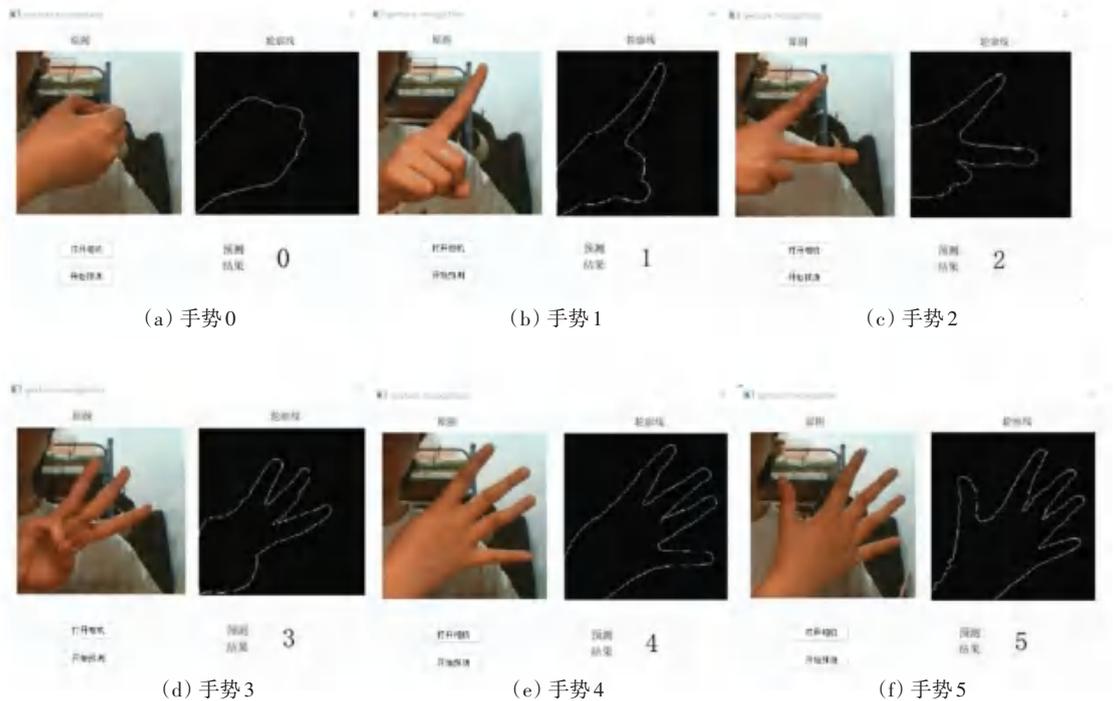


图5 系统界面和显示结果

Fig. 5 System interface and the display results

5 结束语

手势识别在人机交互领域发挥着越来越重要的作用,为了更好地与高科技产品结合以提高人们工作生活的效率,本文对其展开了研究。通过基于 YC, C_b 空间 C_r 分量的 OTSU 法实现肤色分割,减少了背景复杂问题对手势识别的干扰。选择傅里叶描述子描述手势的形状特征,使用支持向量机来训练手势模型,仿真测试结果表明手势识别的准确率为 94.83%,验证了该模型的正确性。但本文研究的手势识别系统仍存在一些局限,例如无法克服光线的影响。未来可以考虑结合其他算法,从而提高系统对光照等外界因素的适应性。其次,识别率还有待提高,下一步应对其它手势识别算法进行深入研究,以提高系统识别准确率。

参考文献

- [1] KANE L, KHANNA P. Depth matrix and adaptive Bayes classifier based dynamic hand gesture recognition[J]. Pattern Recognition Letters, 2019, 120: 24-30.
- [2] SANTOS C, SAMATELO J, VASSALLO R F. Dynamic gesture recognition by Using CNNs and star RGB: A temporal information condensation[J]. Neurocomputing, 2020, 400: 238-254.
- [3] MUJAHID A, AWAN M J, YASIN A, et al. Real-time hand gesture recognition based on deep learning YOLOv3 model[J].

- Applied Sciences, 2021, 11(9): 4164.
- [4] ZHAO Y, SARK V, KRSTIC M, Grass E. Low complexity radar gesture recognition using synthetic training data [J]. Sensors, 2023, 23(1): 308.
- [5] 李占利, 刘博宇, 靳红梅, 等. Gabor-C3D 手势识别算法[J]. 计算机工程与设计, 2023, 44(5): 1405-1411.
- [6] TAO Y F, HU P F, YANG W M. Computer vision technology in smart home [J]. Chinese Journal of Artificial Intelligence, 2020(5): 30-38.
- [7] 周丽娜. 基于视觉的手势识别系统研究[D]. 包头: 内蒙古科技大学, 2014.
- [8] 包宏, 邢邓鸿, 吴庚明, 等. 一种基于肤色模型的手势分割优化方法[J]. 电子技术与软件工程, 2022(10): 179-182.
- [9] ZHOU Kai, WAN Yi, HUANG Hailong. Research on recognition and application of hand gesture based on skin color and SVM[J]. Journal of Computational Methods in Sciences and Engineering, 2020, 20(1): 269-278.
- [10] 安航. 基于视觉的智能车交叉路口交通灯与车辆识别[D]. 重庆: 重庆交通大学, 2018.
- [11] 李军, 钟鹏. OpenCV 的车道线检测方法[J]. 华侨大学学报(自然科学版), 2021, 42(4): 421-424.
- [12] 李光, 王长元, 薛鹏翔. 基于手部关键点和肤色的手势识别[J]. 计算机系统应用, 2021, 30(1): 180-185.
- [13] 陈影柔, 田秋红, 杨慧敏, 等. 基于多特征加权融合的静态手势识别[J]. 计算机系统应用, 2021, 30(2): 20-27.
- [14] 洪期望, 李捍东. 基于支持向量机的手势识别研究[J]. 微处理机, 2022, 43(2): 47-50.
- [15] 徐玥, 周辉. 简单背景下基于 OpenCV 的静态手势识别[J]. 计算机科学, 2022, 49(S2): 393-398.