

# 基于 OpenCV 的多特征实时手势识别

孙 红, 廖 蕾

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘 要** 基于视觉的手势识别是实现新一代人机交互的关键技术。从手势分割和手势表示两方面入手, 提出了一种基于 OpenCV 的多特征实时手势识别方法。利用 HSV 颜色空间肤色分割算法分割出肤色区域, 根据手势的几何特征分离出手势区域, 然后运用凸包算法检测指尖, 结合指尖个数、手指间角度特征和轮廓长宽比特征建立决策树对本定义的 12 种不同手势进行分类。实验结果表明, 本方法具有较好的鲁棒性、实时性好、识别率高。

**关键词** 手势识别; 多特征; 肤色分割; 凸包算法; 决策树

**中图分类号** TP391.41 **文献标识码** A **文章编号** 1007-7820(2015)08-145-04

## Real-time Gesture Recognition with Multiple Cues Based on OpenCV

SUN Hong, LIAO Lei

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract** Computer vision-based hand gesture recognition has become a key technique in the realization of new human-computer interaction. This paper proposes a method in the recognition of real-time hand gestures with multiple cues based on OpenCV. The skin area is detected with HSV, and the hand is segmented with gesture geometric characteristics. Then the fingertips are detected by the convex hull algorithm. For distinguishing and recognizing the 12 different gestures defined, a decision tree is adopted combined with the number of fingers, the angle between your fingers aspect ratio features and contour features. Experimental results demonstrate that the proposed method is robust and real-time, and high recognition rate and various gestures are achieved.

**Keywords** gesture recognition; multiple cues; skin segmentation; convex hull algorithm; decision tree

随着计算机技术的不断发展和普及, 与计算机的交互方式也不断在变化, 而符合人际交流习惯的人机交互技术, 建立和谐自然的人机交互环境成为当前的发展趋势<sup>[1]</sup>。手势作为一种自然直观、表达意义丰富的肢体语言, 越来越多地被运用到人机交互技术。手势识别技术也因此成为了当前人机交互研究的热点。

早期的手势识别要求用户佩戴特殊的数据手套<sup>[2]</sup>, 这种方法虽能精确的识别手势, 但掩盖了手势自然的表达方式, 且设备昂贵。为此, 基于视觉的手势识别应运而生。基于视觉的手势识别输入简单, 对设备要求低, 更加自然直接。在此基础上, 有学者提出了众

多的识别方法。张良国<sup>[3]</sup>等人首次利用 Hausdorff 距离模板匹配的思想, 实现了基于单目视觉的 30 个手指字母的手势识别, 但没有考虑 Model 的旋转与缩放; 黄国范<sup>[4]</sup>等人采用模板匹配的方法, 用改进后的 Hu 矩作为特征向量进行测试样本与模板之间的相似性度量, 但识别率仍然差强人意; 杨波等人<sup>[5]</sup>提取空间相对密度和指节相对间距特征的手势空间分布特征, 计算总的相似性来识别手势, 虽引入随机采样机制提高了识别速度, 但只有少数手势识别率高; 林水强<sup>[6]</sup>等人通过分析手势的几何特征计算特征参数, 采用决策树分类方法进行手势识别, 该方法识别率达到 94% 以上, 但能识别的手势种类不多。

针对上述问题, 本文给出一种基于手势表现特征的手势识别方法。采用 HSV 颜色空间肤色分割算法分割出肤色区域, 然后根据手势的几何特征排除非手势区域; 继而检测指尖, 结合指尖个数, 手指间角度特征和轮廓长宽比特征来进行手势识别。实验结果表明, 本文算法在保证旋转、平移和缩放的不变性的前提下, 既能识别较多的手势, 又有较高的识别率。

收稿日期: 2014-11-19

基金项目: 国家自然科学基金资助项目 (61472256; 61170277); 上海市教委科研创新重点基金资助项目 (12zz137); 沪江基金资助项目 (C14002)

作者简介: 孙红 (1964—), 女, 副教授。研究方向: 计算机网络通信与云计算, 控制科学与工程, 模式识别与智能系统。廖蕾 (1991—), 女, 硕士研究生。研究方向: 图像处理, 机器视觉。E-mail: doublel7830469@sina.com

## 1 手势分割及指尖检测

### 1.1 手势分割

由于 HSV (Hue - Saturation - Value) 色彩空间是从人的视觉系统出发,用色调、饱和度和亮度来描述色彩,比 RGB 空间更符合人的视觉特性<sup>[7]</sup>。本文采用肤色模型,把获取的图像从 RGB 空间转化到 HSV 空间,肤色在 HSV 空间的取值范围为 Hue  $\in [30, 45]$ , Saturation  $\in [35, 200]$ , Value  $\in [20, 255]$ ,由此转换可得到肤色二值图。再运用去噪、膨胀腐蚀等技术获得最终的分割区域,肤色分割如图 1(a) 所示。

按照文献[8]的方法:排除小面积区域,计算肤色的高度(height)及宽度(width)的比例  $K$ ,若满足  $0.7 \leq K \leq 3.0$ ,即为手势区域。至此,可得到最终的手势区域。对于得到的二值图像,由以下几个条件去除干扰的肤色区域,进而得到最终的手势分割结果:

(1) 去除肤色区域面积  $< 400$  像素的区域。

(2) 为排除某些满足条件(1)的狭长区域,设定肤色区域的高和宽必须大于 20 像素。

(3) 肤色区域的重心在肤色区域内部,既重心在手掌上。

(4) 计算肤色的高度(height)及宽度(width)的比例  $K$ ,若满足  $0.7 \leq K \leq 3.0$ ,即为有意义的手势区域。

本文识别出手势区域之后,将手势区域分离出来,并画出手势轮廓,手势分割如图 1(b) 所示。



(a) 肤色分割

(b) 手势分割

图 1 肤色分割图和手势分割图

### 1.2 Graham scan 凸包算法

在做手势时,手指经常远离手掌中心,且指尖离得最远。根据这一情况建立数学模型。因为凸包包含指尖的点,基于 Graham scan 凸包算法找到手型外部的凸包<sup>[9]</sup>。具体过程如下:

(1) 在所有点中选取  $y$  坐标最小的一点  $H$ ,当作基点。如果存在多个点的  $y$  坐标都为最小值,则选取  $x$  坐标最小的一点。坐标相同的点应排除。然后按照其他各点  $p$  和基点构成的向量  $\langle H, p \rangle$  与  $x$  轴的夹角进行排序,夹角由大至小进行顺时针扫描,反之则进行逆时针扫描。

(2) 线段  $\langle H, K \rangle$  一定在凸包上,接着加入  $C$ 。假设线段  $\langle K, C \rangle$  也在凸包上,因为就  $H, K, C$  这 3 点而言,其的凸包就是由此 3 点所组成。但接下来加入  $D$  时会发现,线段  $\langle K, D \rangle$  才会在凸包上,所以将线段  $\langle K, C \rangle$  排除,  $C$  点不可能是凸包。

(3) 即当加入一点时,必须考虑到前面的线段是否会出现于凸包上。从基点开始,凸包上每条相邻的线段的旋转方向应该一致,并与扫描的方向相反。如果发现新加的点使得新线段与上线段的旋转方向发生变化,则可判定上一点必然不在凸包上。

在图 2 中,加入  $K$  点时,由于线段  $\langle H, K \rangle$  相对于  $\langle H, C \rangle$  为顺时针旋转,所以  $C$  点不在凸包上,应该删除,保留  $K$  点。接着加入  $D$  点,由于线段  $\langle K, D \rangle$  相对于  $\langle H, K \rangle$  为逆时针旋转,故  $D$  点保留。按照上述步骤进行扫描,直到点集中所有的点都遍历完成,即得到凸包。

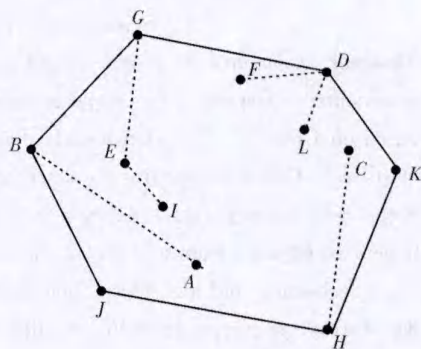


图 2 凸包示意图

### 1.3 指尖检测

提取到手的轮廓后,计算轮廓内接圆的中心坐标得到掌心。通过计算那些既在凸包上又在轮廓上的点来找到指尖<sup>[10]</sup>。

(1) 初始化  $C$  序列既为凸包上的点又是轮廓上的点的集合。

(2) 遍历  $C$  序列上的每个点  $p$ ,取其左右  $m$  和  $n$  点,分别计算它们到  $p$  的矢量值  $pm$  和  $pn$ 。

(3) 计算  $pm$  和  $pn$  的夹角。如果  $\alpha_i < 45^\circ, i = 1, \dots, N$ ,则确定为指尖,否则返回步骤(2)重新检查下一个候选点。

## 2 手势表现特征定义

通常情况下,图像特征的选取往往对识别结果的准确性有较大的影响,单一的特征一般会受环境的影响和其他因素的干扰。考虑到计算的复杂性和识别的实时性等因素,本文定义了 4 种手势表现特征。图 3 列出了 12 种常见的手势。



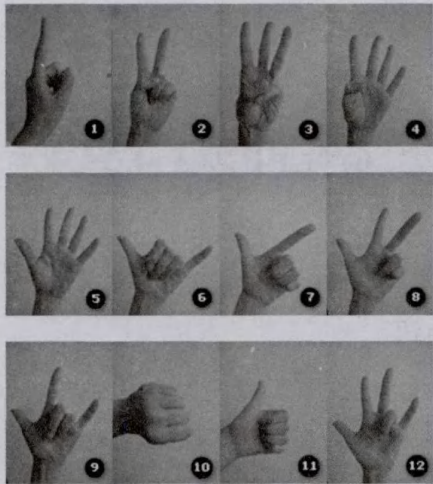


图3 12种常见的手势

根据上文的方法得到指尖,指尖依次记为  $P_1, P_2, \dots, P_N (N \leq 5)$ ;检测到的指尖个数记为  $N$ ;通过测量得知,正

常人5指分开后,手指间的角度在  $10^\circ \sim 180^\circ$  之间<sup>[11]</sup>。大拇指与食指间的角度为  $40^\circ \sim 90^\circ$ ,食指与中指、中指与无名指、无名指与小指的角度为  $10^\circ \sim 35^\circ$ ,大拇指与小指间的角度为  $95^\circ \sim 150^\circ$ ,这为本文的算法提供了依据。本文将指尖与掌心  $O$  以直线连接,计算手指间的相对角度。计算相邻两个指尖与掌心的角度  $\beta_{N-1}$ ,即  $\beta_1$  为  $P_1$  和  $P_2$  与掌心  $O$  的夹角; $\beta_2$  为  $P_2$  和  $P_3$  与掌心  $O$  的夹角; $\beta_3$  为  $P_3$  和  $P_4$  与掌心  $O$  的夹角; $\beta_4$  为  $P_4$  和  $P_5$  与掌心  $O$  的夹角。

本文定义4个手势表现特征:指尖个数  $N$ ;手势的最小外接矩形的长宽比  $K = L/W$ ;相邻两个指尖与掌心的角度  $\beta_{N-1}$  在  $40^\circ \sim 90^\circ$  之间的个数  $N_1$ 、在  $10^\circ \sim 35^\circ$  之间的个数  $N_2$ 。

3 手势分类与识别

根据以上4个特征,可通过建立决策树来分类识别上述12种常见的手势,其识别流程图如图4所示。

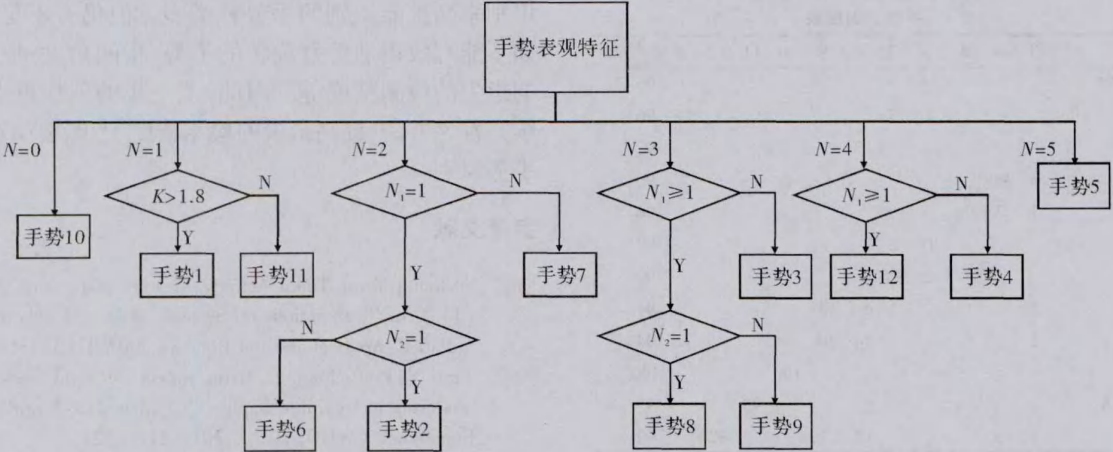


图4 手势决策树图

首先根据所检测到的指尖的个数  $N$  对12种手势加以区分,当  $N=5$  和  $N=0$  时可分别分离手势5和手势10;当  $N=1$  时,手势1和手势11由于指尖只有一个,无法用角度特征区分,只能用  $K$  加以区分。当  $K >$  时( $\sigma$ 取1.8),能分离手势1;当  $N=2$  时,手势7的  $N_1=1$ ,而手势2的  $N_1 \neq 1$  且  $N_2=1$ ,故此能分离手势2和6;当  $N=3$  时,食指与小指之间的角度范围在  $30^\circ \sim$

$105^\circ$  之间,而大拇指和食指之间的角度在  $40^\circ \sim 90^\circ$  所以根据  $N_1$  来区分手势8和手势9有较的误差,所以本文根据  $N_1 \geq 1$  来分离手势3,然后手势8的  $N_2=1$ ,手势9的  $N_2 \neq 1$ ;当  $N=4$  时,手势12和手势4根据  $N_1 \geq 1$  与否则来确定。至此,本文定义的12种常见的手势均可识别,部分手势识别结果如图5所示。



(a) 手势5



(b) 手势2

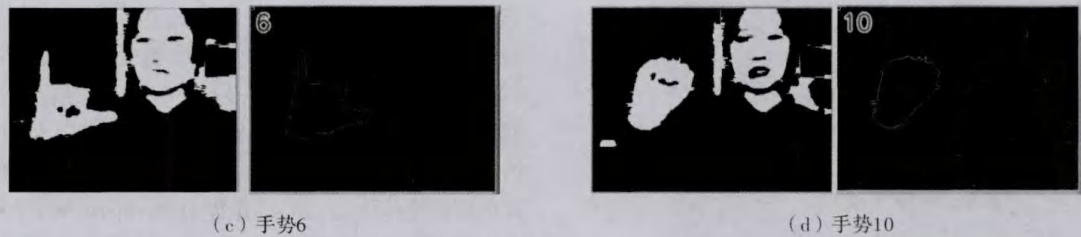


图5 部分手势识别结果

4 实验结果及分析

本文在 Windows 环境下,借助计算机视觉库 OpenCV 提供图像算法函数和微软 Visual Studio 2010 搭建开发平台,实现手势识别。实验在光照条件良好的环境下,由 10 人分别在不同背景条件下对每种手势做 10 次,并且每次都相对平面旋转一定得角度,手离摄像头的距离也在 30 ~ 50 cm 不等,每种手势共计 100 次,表 1 给出了 12 种手势的识别结果及准确率。

表1 手势识别结果

手势	1	2	3	4	5	6	7	8	9	10	11	12	准确率/%
1	95									5			95
2		96				1	3						96
3		3	92	2				3					92
4			4	90							6		90
5			1	3	96								96
6		1				97	2						97
7		1				2	97						97
8			3				3	91	3				91
9			2					4	94				94
10										100			100
11	5										95		95
12			1	3				2	2			92	92

由实验结果可知,识别手势 10 的准确率为 100%,手势 1,2,5,6,7,9,11 的识别准确率均在 94% 以上,而手势 3,手势 4,手势 8 和 12 识别率偏低,主要是由于不同的人在这些手势上表达有所不同,不同的人每次做出的手势手指间的角度不同,过小则可能导致指尖的识别产生误差。

文献[5]综合手势的空间相对密度特征和指节相对间距特征来识别手势,多数手势识别率低。本文分别采用这两种方法对两方法均能识别的部分手势做了对比实验,结果如表 2 所示,表明本文算法准确率较高。

表2 对比试验

手势	文献[7]方法识别率/%	本文方法识别率/%
1	90.0	95
2	71.0	96
5	68.8	96
6	56.0	97
12	56.9	92

5 结束语

本文实验在 Windows7 操作系统,VS2010 + OpenCV 环境下完成。本文提出的基于手势表现特征的静态手势识别方法,利用凸包和轮廓检测指尖,结合指尖个数,手指角度特征和轮廓长宽比来分类,利用本方法成功地对本文定义的 12 种手势进行了识别,通过多人在不同背景下做不同大小、角度的手势进行识别,结果表明该算法对平移、旋转以及缩放有鲁棒性,实时性好,识别率高且能识别的手势种类多。但仍有不足之处,如不能有效识别垂直旋转的手势,指间角度过小时影响指尖的检测精度等。因此,下一步的工作重点是继续完善该方法,提高指尖的检测精度,考虑增加双手的手势识别。

参考文献

[1] Sushmita Mitra, Tinku Acharya. Gesture recognition; a survey [J]. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 2007, 37(3): 311 - 324.

[2] Frati V, Prattichizzo D. Using kinect for hand tracking and rendering in wearable haptics [C]. Istanbul: World Haptics Conference (WHC), IEEE, 2011: 317 - 321.

[3] 张国良, 吴江琴, 高文, 等. 基于 Hausdorff 距离的手势识别[J]. 中国图象图形学报, 2002, 7(11): 1144 - 1150.

[4] 黄国范, 李英. 基于改进 Hu 矩的字母手势识别[J]. 赤峰学院学报, 2013, 29(3): 23 - 24.

[5] 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[J]. 计算机辅助设计与图形学报, 2010, 22(10): 1841 - 1848.

[6] 林水强, 吴亚东, 陈永辉. 基于几何特征的手势识别方法[J]. 计算机工程与设计, 2014, 35(2): 636 - 640.

[7] 成琳, 陈俊杰, 相洁. 图像颜色征提取技术的研究与应用[J]. 计算机工程与设计, 2009, 30(14): 3451 - 3454.

[8] 翁汉良, 战萌伟. 基于视觉的多特征手势识别[J]. 计算机工程与科学, 2012, 34(2): 123 - 127.

[9] Adams W, Belotti P, Shen R. Convex hull characterization of lexicographic orderings [J]. Writing, 2013, 23(6): 719 - 726.

[10] Zhang Binjue, Zhao Liaoyin, Wang Yixuan. Fingertip detection and gesture recognition based on kinectdepth data [J]. IEEE Transactions on Computer Science and Technology, 2014, 3(1): 9 - 14.

[11] Stenger B. Template - based hand pose recognition using multiple cues [C]. Haikou: Proceeding of ACCV'06, 2006: 551 - 560.