

# 基于深度图像技术的手势识别方法

曹维清, 李瑞峰, 赵立军

(哈尔滨工业大学机器人技术与系统国家重点实验室, 哈尔滨 150001)

**摘 要:** 针对复杂环境下的手势识别问题, 提出一种基于深度图像技术的手势识别方法。利用深度图像信息从复杂环境中提取手势区域, 综合手势的表观特征, 建立决策树实现手势的识别。对常见的 9 种手势在复杂背景条件下进行测试, 实验结果表明, 手势的平均识别率可达到 98.4%, 速度达到每秒 25 帧。

**关键词:** 手势识别; 深度图像; 表观特征; 复杂背景; 决策树

## Hand Posture Recognition Method Based on Depth Image Technology

CAO Chu-qing, LI Rui-feng, ZHAO Li-jun

(State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China)

**【Abstract】** Aiming at the problem of hand posture recognition from complex backgrounds, this paper proposes a hand posture recognition method based on depth image technology. The hand posture region is extracted from complex background via depth image. Appearance features are integrated to build the decision tree for hand posture recognition. Nine common postures with complex background are tested. Experimental results demonstrate that recognition rate is 98.4% and speed rate achieves 25 frames per second.

**【Key words】** hand posture recognition; depth image; appearance feature; complex background; decision tree

DOI: 10.3969/j.issn.1000-3428.2012.08.006

### 1 概述

随着机器人技术的发展, 智能服务机器人开始逐步融入人类的生活, 人机交互技术的发展是智能服务机器人研究的重要组成部分。人机交互的研究方向主要有表情、声音和手势识别技术。其中, 手势是一种自然直观的人机交流模式, 在信息表达上比表情更明确、信息量更丰富。

在人机交互中, 基于视觉的手势识别是实现新一代人机交互的关键技术。基于视觉的手势识别通过图像信息, 让机器人获取人的手势姿态信息, 对不同的手势信息进行分类。近年来, 学者们对基于视觉的手势识别进行研究, 文献[1]提出特征空间匹配法, 文献[2]采用提取同类手势的 SIFT 特征方法进行手势识别, 文献[3]提出手势图解析分类法, 文献[4]针对复杂背景采用空间分布特征对手势进行识别。

复杂背景下手势识别的研究<sup>[4-5]</sup>一般使用肤色模型实现手势区域分割, 虽然可以区分肤色和非肤色区域, 但是实际应用图像都包含有肢体和面部肤色区域, 它们对手势区域提取有着较大的影响, 同时光线也直接影响肤色模型。现有的文献中并没有考虑强光和弱光环境下的手势识别情况, 在实际应用中往往伴随光线的变化问题, 这些问题都直接影响了人机交互的实际效果。本文提出一种基于深度图像技术的手势识别方法, 从深度图像信息中提取手势区域, 运用几何方法提取手势表观特征并分类, 从而实现复杂环境下基于手势的人机交互。

### 2 基于深度图像技术的手势识别方法

#### 2.1 手势识别系统

基于深度图像信息的手势识别系统框架如图 1 所示。深度图像的采集采用非接触测量方式, 获取的场景中深度信息不受物体自身的颜色、纹理特征以及背景环境光线强弱的影响。本文选用微软公司的 Kinect 进行深度图像的采集, 获取

深度信息值, 转换到灰度值图像, 经过手势区域分割、特征提取、手势识别, 转换为手势信息, 提供人机交互使用。

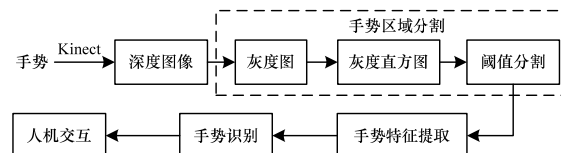


图 1 面向人机交互的手势识别框架

#### 2.2 手势区域分割

##### 2.2.1 深度图像成像原理

深度图像技术在近年来逐步得到广泛的关注, Kinect 采集 640×480 像素的深度图像速度可以达到 30 f/s, 深度的分辨率在 5 mm 左右。本文把 Kinect 采集到的场景深度值转换到灰度值空间, 实际场景的 RGB 图像如图 2 所示。在深度图像中像素点灰度值的大小对应于场景中点的不同深度值, 如图 3 所示, 图像中的灰度值越大, 表明该区域距离深度摄像头的距离越近。



图 2 场景 RGB 图像



图 3 深度图像

**基金项目:** 国家自然科学基金资助项目(61075081); 机器人技术与系统国家重点实验室课题基金资助项目(SKLRS200802A02)

**作者简介:** 曹维清(1982—), 男, 博士研究生, 主研方向: 模式识别, 人机交互; 李瑞峰, 教授、博士生导师; 赵立军, 讲师

**收稿日期:** 2011-07-11 **E-mail:** caochuqing@gmail.com

### 2.2.2 手势区域的分割

在人机交互的过程中, 手势动作往往都置于身体位姿之前, 本文利用手势区域的与背景区域不同的深度值分割出手势区域。同一深度的像素点在深度图像中灰度值相同, 但是每次人和深度摄像头的之间的距离都不完全相同, 无法用固定深度阈值实现区域的分割。本文使用基于灰度直方图的方法寻找手势区域与背景的分割阈值。

灰度直方图表示图像中具有每种灰度级的像素的个数, 反映图像中每种灰度出现的频率<sup>[6]</sup>。本文针对深度值所对应的灰度图像, 计算出灰度直方图。手势部位往往是离深度摄像头较近的区域, 且相对于背景区域面积较小, 因此从灰度值由大到小开始寻找像数点变化, 比较大的灰度值处作为区域分割的灰度阈值。图4为图3对应的灰度直方图, 可见当灰度值为235时有明显的数值变化区域, 根据该阈值对手势区域与背景区域二值化, 分割效果如图5所示。

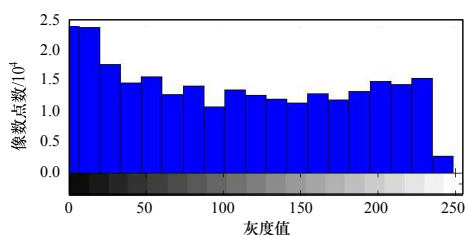


图4 灰度直方图



图5 手势区域分割效果

## 2.3 特征提取

### 2.3.1 手势的表现特征

本文通过提取手势的表现特征, 根据表征手势的指头数以及指间的夹角的不同对手势进行分类, 实现手势在旋转缩放条件下的快速识别。相比于其他的手势特征提取方法, 手势的表现特征具有更加直观、无需训练样本、适应性强、运算速度快实时性好等优点。

手势表现特征的提取步骤如下:

(1)通过数学形态学中的腐蚀操作<sup>[6]</sup>获取手势区域中心位置点。由于手掌作为手势表现的主要组成部分, 在手势区域中它占有最大的面积而且点较为集中。通过连续的腐蚀操作, 可以消除手势区域的边界点, 使手势区域逐步缩小, 最终得到手势区域的中心位置点  $C_0$ , 实际效果如图6所示。

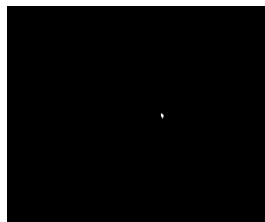


图6 手势区域中心点

(2)计算出中心点与手势区域边缘的最大距离值  $l$ , 对距离进行10等分, 以  $d=l/10$ 。以手势区域中心点为圆心做出圆轨迹, 圆半径从  $d$  开始到  $l$ , 每次递加值为  $d$ , 得到10个圆轨迹线, 如图7所示。

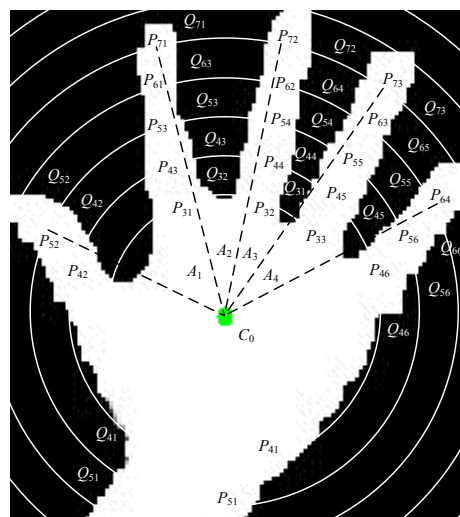


图7 手势表现特征提取

(3)按顺时针方向记录每条圆轨迹线上像素值变化点  $P_{ij}$  (0~1, 即从黑色区域到白色区域) 和  $Q_{ij}$  (1~0, 即从白色区域到黑色区域) 的位置坐标值,  $i$  表示轨迹圆的编号,  $j$  表示同一轨迹圆上  $P$  或者  $Q$  点编号, 同时删除单独存在的  $P_{ij}$  点和  $Q_{ij}$  点。

(4)通过  $P_{ij}$  和  $Q_{ij}$  的位置坐标, 计算每对  $P_{ij}$  和  $Q_{ij}$  之间的距离  $D_{ij}$ , 当轨迹圆与指尖部分区域相截时获得较小的  $D_{ij}$ , 不能表现手指的实际宽度值, 所以当  $D_{ij}$  小于阈值  $\delta$  时, 删除对应的  $P_{ij}$  和  $Q_{ij}$  点, 由图2可以看出一些手指的指间处  $P_{ij}$  和  $Q_{ij}$  点被去除, 根据经验设置阈值为  $\delta=d/4$ 。

(5)每个轨迹圆上获得的  $j$  最大值, 为与手掌相连分支数量总和  $N=\max(j)$ 。由于分支中包含手指和手腕分支, 则手指数量是  $N_f=N-1$ 。

(6)由每个分支的平均值计算得到分支的平均宽度  $W_j=D_{ij}$ 。在手形上可知手腕宽度大于手指, 在分支中手腕对应于最大宽度的分支。由除手腕分支外的其他分支在各自截取的最大轨迹圆上取  $P_{ij}$  和  $Q_{ij}$  的中点与中心位置点  $C_0$  连线, 得到指间夹角  $A_{j-1}$ 。

### 2.3.2 旋转和缩放不变性

人机交互中需要自然的手势动作, 而不局限于某些特定手势位姿或是特定手势区域大小。距离的远近以及手势的位姿会造成图像中手势大小以及旋转的问题。本文的特征提取方法, 利用等分圆轨迹消除手势区域大小对特征提取的影响, 同时对轨迹上的变化点进行特征计算, 提取的手指数量  $N_f$  和指间夹角  $A_{j-1}$  特征都具有旋转和缩放的不变性, 不受手势距离远近和旋转的影响。

## 2.4 手势识别

9种常见的手势如图8所示。

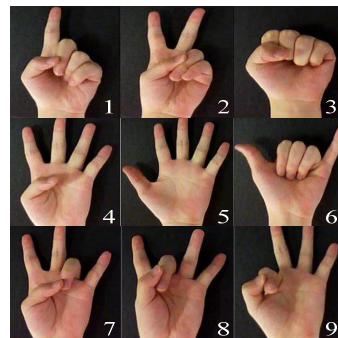


图8 9种常见的手势

本文通过提取手势的表观特征，建立决策树模型，对手势进行识别分类。决策树是通过对训练样本进行归纳学习，生成决策树或决策规则，然后使用决策树或决策规则对新数据进行分类的一种数学方法<sup>[7]</sup>。决策树通过把实例从根节点排列到某个叶子节点来分类，叶子节点即为实例所属的分类。构造好的决策树的关键在于如何选择恰当的逻辑判断或属性。

手势主要的提取特征是手指数量  $N_f$  和指间夹角  $A_{j-1}$  特征，作为决策树的分类节点。不同手势之间手指数量特征区分比较明显，建立的决策树首先以手指数量  $N_f$  作为决策树的根节点，对于手指数相同的手势，再建立子节点对各个指间夹角  $A_{j-1}$  差异情况进行区分。子节点上对于指间夹角  $A_{j-1}$ ，通过小样本的测试训练获取分类条件，建立的决策树如图 9 所示。

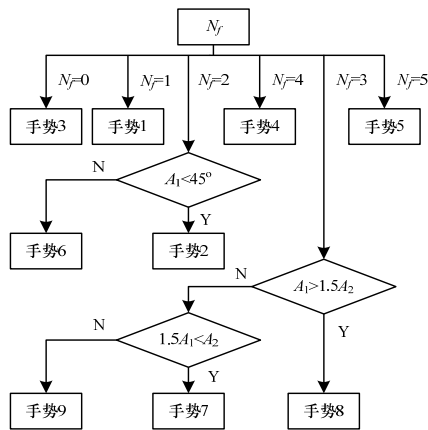


图 9 手势分类决策树

手势 1、手势 3、手势 4、手势 5 的手指数量特征具有唯一性，可直接通过根节点分类；手势 2 和手势 6 的  $N_f=2$ ，下一层子节点通过判断指间夹角  $A_1$  大小加以区分；手势 7、手势 8、手势 9 的  $N_f=3$ ，需要通过 2 层子节点判断指间夹角  $A_1$  和  $A_1$  大小进行区分。

### 3 实验结果及分析

#### 3.1 手势识别实验

为验证本文采用的实验方法，基于机器人平台，Kinect 通过 USB 口与机器人主机进行数据通信，机器人主机的配置是 2.0 GHz 的 CPU 和 2 GB 的 RAM，当采样图片大小为 640×480 像素时，识别速度可以达到 25 f/s，达到实时手势识别的要求。

实验针对 9 种常见的手势进行识别，由 5 个人在强光、弱光和普通光线的背景条件下对每种手势测试 200 次，共计 1 800 次测试。手势识别结果如表 1 所示。

表 1 手势识别结果

测试 手势	识别次数									识别率 (%)
	手势 1	手势 2	手势 3	手势 4	手势 5	手势 6	手势 7	手势 8	手势 9	
手势 1	200	0	0	0	0	0	0	0	0	100.0
手势 2	0	200	0	0	0	0	0	0	0	100.0
手势 3	0	0	200	0	0	0	0	0	0	100.0
手势 4	0	0	0	200	0	0	0	0	0	100.0
手势 5	0	0	0	0	200	0	0	0	0	100.0
手势 6	1	5	0	0	0	194	0	0	0	97.0
手势 7	0	0	0	0	0	0	190	10	0	95.0
手势 8	0	0	0	0	0	0	8	192	0	96.0
手势 9	0	0	0	0	0	0	2	3	195	97.5

从表 1 中可以看出，手势 1~手势 5 的识别率都为 100%。由于非特定人物对手势 6~手势 9 表达差异，当手势的表达差异较大时会造成测试过程中的识别错误，但从整体上看，平均手势识别成功率达到 98.4%，说明了本文手势识别方法的有效性。

#### 3.2 复杂的背景条件

##### 3.2.1 光照的强弱

环境中光线的变化是手势区域提取常见的难题之一，常见的手势区域分割方法多依赖于手势区域颜色，光线的强弱变化造成手势区域颜色较大的变化，当光线较强或者较弱时，手势区域部分甚至完全与背景色相近，无法分割出手势区域，严重的影响了手势识别效果。

由于深度信息的获取主要通过于深度摄像头自身对非可见光信号的发射和采集，环境外部可见光的变化对深度图像的采集没有任何的影响，室内场景中光线的变化对手势区域分割效果如图 10 所示，即使在室外强光甚至室内几乎没有光线的情况下，仍可以准确提取手势区域。

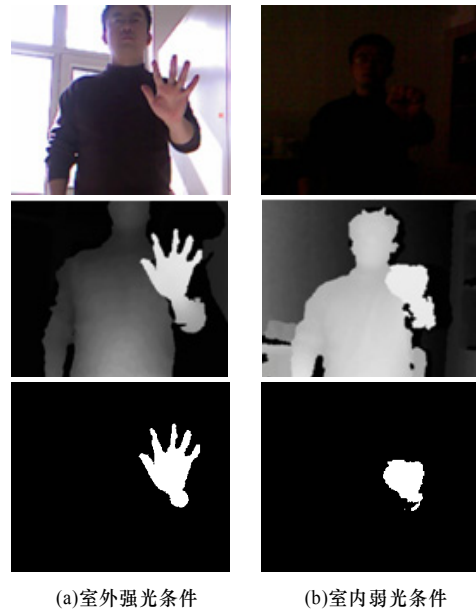


图 10 复杂背景下手势提取效果

##### 3.2.2 肤色区域重合

肤色重合时的手势区域分割是比较常见的难题<sup>[4]</sup>，如手势和脸部区域重合情况。在彩色图像中手势和脸部等肤色区域的重合影响了手势区域的分割效果。由于面部和手势区域颜色相近，无法通过肤色模型直接分割出手势区域，导致无法完整的提取出手势区域。在该情况下手势识别失败率很高。然而在深度图像中，手势和脸部等区域通常处于不同的深度区域下，通过本文的深度图像分割方法可以实现非常好的分割效果，复杂的背景环境不会影响手势的识别效果。

### 4 结束语

本文针对复杂背景下手势识别存在的难点，综合利用深度图像信息以及手势的表观特征，解决了在光照及肤色区域重合时的手势区域分割问题，实时的手势识别不受手势旋转和缩放的影响。该手势识别系统已应用于机器人平台，实验结果证明了本文方法的实用性和稳定性。该方法对背景环境有很强的鲁棒性，能为多方式的人机交互提供快速有效的信息。

(下转第 21 页)

Step4 calculate  $allp_i$  of  $Q_i$  using formula (7);

Step5 if  $allp_i \leq budget_i$

assign  $Q_i$  to  $C_j$  when  $ask_j^i$  is minimal ;

else refuse  $Q_i$  ;

Step6 output  $i$  and  $j$ ;  $p_i$  and  $allp_i$

End

本文提出的定价算法使资源提供商代理的要价与工作负载呈正比, 从而可避免用户任务超过截止时间仍无法完成的情况, 同时在一定程度上可使用户花费最小。

## 4 实验结果与分析

### 4.1 实验环境设计

采用 CloudSim 模拟器对本文提出的 DSP 定价算法进行仿真, 模拟的云计算环境包括 100 个用户和 20 个云资源提供商, 每个用户向云市场提交一个或多个独立的任务请求, 共有 200 个。 $\alpha$  表示工作负载对要价的影响程度,  $\alpha = \{0.25, 0.5, 0.75, 1\}$ ;  $\beta$  表示响应时间对出价的影响程度,  $\beta = \{0.25, 0.5, 0.75, 1\}$ 。实验进行 16 次, 结果取 16 次实验的平均结果。

将 DSP 算法与采用固定比例的定价算法(Fixed Pricing, FP)从下面的指标进行比较:

(1)用户利益: 用户的出价与单位交易价格的差值。

(2)资源提供商的利益: 单位交易价格与资源提供商的要价的差值。

### 4.2 性能分析

在不同工作负载情况下, DSP 算法与 FP 算法得到的资源提供商利益的比较情况如图 1 所示。可以看出, 资源提供商的利益随着工作负载的变化而逐渐增大, 但 FP 的资源提供商利益始终低于 DSP, 这是由于 FP 在调度中没有考虑工作负载的动态性。

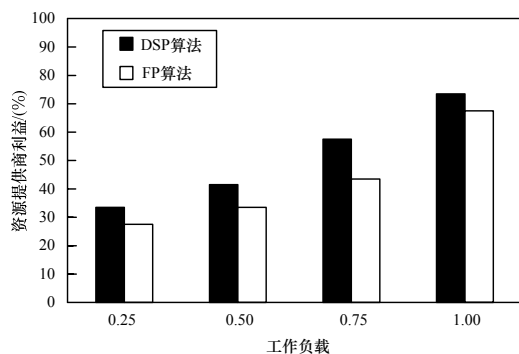


图1 2种算法的资源提供商利益比较

DSP 算法与 FP 算法的用户利益比较如图 2 所示。可以看出, 用户的利益随着响应时间与最大响应时间比值的增大

而逐渐增大, 但 DSP 算法明显优于 FP 算法, 这是因为 DSP 考虑了用户的 QoS 需求, 在资源负载较重时, 用户的某些服务由于受到截止时间的约束而不会推迟完成, 从而可以提高用户的利益。

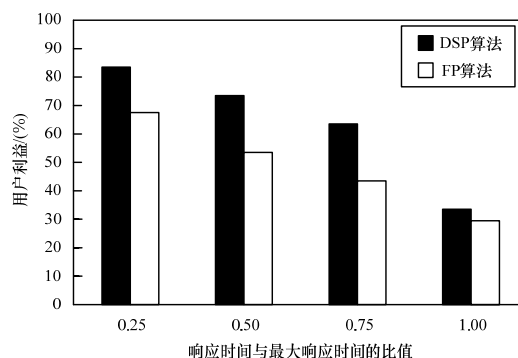


图2 2种算法的用户利益比较

## 5 结束语

本文提出一种云环境下基于组合双向拍卖机制的资源定价模型, 并给出相应算法。仿真实验结果表明, 本文算法在资源利用率、用户和资源提供商利益方面的性能明显优于固定比率定价算法。下一步将在模型中引入信任度以反映服务质量, 进一步完善定价算法。

## 参考文献

- [1] 程仕伟, 潘 郁. 云计算环境下基于可信性的动态资源分配策略[J]. 计算机工程, 2011, 37(11): 45-48.
- [2] Sui Xin, Leung Ho-Fung. An Adaptive Bidding Strategy in Multi-round Combinatorial Auctions for Resource Allocation[C]//Proc. of International Conf. on Tools with Artificial Intelligence. Singapore: IEEE Computer Society, 2008.
- [3] Teng Fei, Magoules F. Resource Pricing and Equilibrium Allocation Policy in Cloud Computing[C]//Proc. of International Conf. on Computer and Information Technology. [S. l.]: IEEE Computer Society, 2010.
- [4] Mihailescu M, Teo Y M. Dynamic Resource Pricing on Federated Clouds[C]//Proc. of International Conf. on Cluster, Cloud and Grid Computing. [S. l.]: IEEE Computer Society, 2010.
- [5] Shang Shifeng, Jiang Jinlei, Wu Yongwei, et al. DABGPM: A Double Auction Bayesian Game-based Pricing Model in Cloud Market[C]//Proc. of International Federation for Information Processing. Brisbane, Australia: [s. n.], 2010.

编辑 顾姣健

(上接第18页)

## 参考文献

- [1] Kelly D, McDonald J, Markham C. A Person Independent System for Recognition of Hand Postures Used in Sign Language[J]. Pattern Recognition Letters, 2010, 31(11): 1359-1368.
- [2] Wang Chieh-Chih, Wang Ko-Chih. Hand Posture Recognition Using Adaboost with Sift for Human Robot Interaction[C]//Proc. of International Conference on Advanced Robotics. Jeju Island, Korea: [s. n.], 2007.
- [3] Flasiński M, Mysliński S. On the Use of Graph Parsing for Recognition of Isolated Hand Postures of Polish Sign Language[J]. Pattern Recognition, 2010, 43(6): 2249-2264.
- [4] 杨 波, 宋晓娜, 冯志全. 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学报, 2010, 22(10): 1841-1848.
- [5] 陈一民, 张云华. 基于手势识别的机器人人机交互技术研究[J]. 机器人, 2009, 31(4): 351-356.
- [6] 蓝章礼, 李益才. 数字图像处理与图像通信[M]. 北京: 清华大学出版社, 2009.
- [7] Witten I H, Frank E. Data Mining: Practical Machine Learning Tools and Techniques[M]. Burlington, USA: Morgan Kaufmann Publishers, 2005.

编辑 顾姣健