

张生军 著

VISION-BASED NON-LABELING HAND GESTURE RECOGNITION

# 基于视觉的无标记 手势识别

 吉林大学出版社  
JILIN UNIVERSITY PRESS

VISION-BASED NON-LABELING HAND GESTURE RECOGNITION

# 基于视觉的无标记 手势识别

张生军  著

 吉林大学出版社  
JILIN UNIVERSITY PRESS

图书在版编目 ( CIP ) 数据

基于视觉的无标记手势识别 / 张生军著. —长春:  
吉林大学出版社, 2016. 6

ISBN 978-7-5677-6761-4

I. ①基… II. ①张… III. ①计算机视觉 - 研究  
IV. ①TP302.7

中国版本图书馆CIP数据核字 (2016) 第146115号

书 名: 基于视觉的无标记手势识别

作 者: 张生军 著

责任编辑: 邵宇彤 责任校对: 丁 鑫

吉林大学出版社出版、发行

开本: 880 × 1230毫米 1/32

印张: 6 字数: 104千字

ISBN 978-7-5677-6761-4

封面设计: 中尚图

北京墨阁印刷有限公司 印刷

2016年6月 第1版

2016年6月 第1次印刷

定价: 32.00元

版权所有 翻印必究

社址: 长春市明德路501号邮编: 130021

发行部电话: 0431-89580028/29

网址: <http://www.jlup.com.cn>

E-mail: [jlup@mail.jlu.edu.cn](mailto:jlup@mail.jlu.edu.cn)

## 序 言

---

基于视觉的无标记手势识别是计算机交互领域内重要的研究课题,其主要是通过对摄像头捕获到的手势进行分析,然后结合各种算法实现无标记手势动作的自动识别。相对于其他人机交互方式,手势识别技术具有自然、方便、经济和良好的用户体验感等特点,同时也有很好的市场应用价值和科研价值。近年来,手势识别技术在计算机交互技术中受到越来越多的重视,但是如何实现一个通用、基于普通摄像头、对外界环境变化鲁棒性强、无标记以及实时的手势识别系统还没有很好的解决方案。针对该问题,本书分别从手部检测、手部跟踪和手势识别三个方面对无标记手势识别中相关算法进行了研究,在研究过程中还考虑了手势识别系统的整体性和实用性问题。本书的主要工作和贡献如下:

1. 手部检测方面。根据手部所具有的特点,本文采用多特征融合的方法对手部进行检测。根据人体肤色分布特性,为提高肤色判定准确性,本文采用了高斯肤色模型对肤色进行了建模。针对外界光照造成的肤色容易发生变化的问题,本文采用 HSV 颜色空间来表示不同的肤色。由于手势具有非刚体运动特性,在手部运动过程中会出现背景信息融入手部的情况,针对该问题本文在

使用 Haar-like 特征进行手部描述时采用了手部背景去除算法。为提高 Haar-like 特征分类的准确性和快速性,本文采用快速 Ada-Boost 分类器。为提高手部检测的准确性,并减少手势过程中由于双手相互遮挡造成的手部检测错误,本文针对双手存在相互遮挡、不容易区分的特点提出了基于 HOG 特征和 PCA 分析的左右手区分方法。

2. 手部跟踪方面。根据手部本身和手部运动过程中所具有的特点,本文提出了使用肤色、SIFT 和 HOG 多特征相融合,并采用 Mean-shift 作为路径寻优的连续手部跟踪算法 CASHS。首先,根据手势特点和运动中所具有的非刚体运动特性,算法首先提取手部的基于肤色的 SIFT 特征,通过该特征可以实现对手部的缩放、视角、亮度和部分缺失等变化的描述无关性,同时通过肤色特征可以辅助去除掉部分背景干扰并且缩小特征提取范围;其次,针对手部运动过程中具有的缩放、变形、模糊和旋转等特性,该算法采用了 HOG 特征对手势特征进行描述。通过这两种特征的结合可以使算法在一定程度上提高手部跟踪的鲁棒性。最后,本研究采用了具有较好时间特性和跟踪准确性的 Mean-shift 算法对手部运动进行跟踪,在跟踪过程中增加前述两个特征作为相似性度量依据,这有利于提高算法跟踪稳定性。

3. 手势识别方面。作为本研究的重点也是难点内容,该部分的研究内容分为用于系统控制指令的静态手势识别和用于信息交互的动态手势识别两个部分。针对静态手势识别本文研究了基于肤色的 SIFT 特征的手势识别方法,针对动态手势识别本文提出了

FCRFs(Fuzzy based Condition Random Fields)的手势标注方法和 FLDCRFs(Fuzzy based Latent-dynamic Condition Random Fields)的手势识别方法。首先,针对静态手势识别,本文在识别中采用基于肤色的 SIFT 特征。SIFT 特征是基于灰度图像的特征,由于在灰度空间内手部以外的其他信息尤其是背景区域对手部特征提取的干扰比较大,而通过增加肤色信息有助于区分背景和手部,可以在一定程度上解决背景干扰的问题,因此增加颜色信息的 SIFT 特征对于提高手部特征的描述具有较重要意义。同时由于手部运动属于非刚体的运动,即便是静态手势也不能保证每次手势都可以保持相对稳定的特征,因此我们增加了 HOG 特征以增强对手部缩放、旋转等情况的鲁棒性。其次,针对动态手势识别,本文从手势认知的角度对手势的定义和识别做了深入的分析,并在此基础上提出了对手势采用模糊集合描述的方法。通过这种方式定义的手势,不但更符合日常生活中人们对手势的定义,而且也便于机器识别。与精确集合表示方式相比,采用模糊集合方式进行的手势描述可以解决一些无法用精确集合区分的动作,这对于提高手势识别的准确性有较重要意义。本文在动态手势识别中采用了 CRF 辨识模型,与 HMMs 模型相比,该模型具有更好的精确性和更低的训练样本要求。在上述模糊特征和 CRFs 模型分析基础上,本文从两个层次对手势识别进行了研究:首先,提出了基于模糊特征的条件随机场 FCRFs 作为手势序列标注方法,通过标注可以很好地区分手势序列的开始和结束过程;其次,针对连续手势识别的需求,提出了基于模糊特征的隐动态条件随机场 FLDCRFs(fuzzy based Condition

Random Fields)模型。

本文对手势识别各模块中包括的算法流程进行了较详细的理论分析,并针对各种方法进行了相关实验验证。实验结果显示本文各模块算法具有较好的性能指标:手部检测算法对不同肤色具有一定的鲁棒性,对左右手遮挡具有自动识别的能力;手部跟踪算法对手型的变化、快速运动和复杂背景都有较好的跟踪准确度;静态手势识别算法对复杂背景下的简单手势具有较高的识别率;动态手势识别算法通过使用模糊手势特征对手势识别率的提示较大。同时,由于手势识别技术融合多个学科的知识,加之该领域的研究正处于发展当中,本研究中各部分还存在一些有待改善的地方。

张生军

2016年6月

# 目 录

|                             |    |
|-----------------------------|----|
| 第 1 章 引 言 .....             | 1  |
| 1.1 研究背景 .....              | 1  |
| 1.2 研究意义 .....              | 6  |
| 1.3 国内外研究现状 .....           | 7  |
| 1.3.1 手部检测 .....            | 9  |
| 1.3.2 手部跟踪 .....            | 12 |
| 1.3.3 静态手势识别 .....          | 17 |
| 1.3.4 动态手势识别 .....          | 19 |
| 1.4 本文工作和主要贡献 .....         | 22 |
| 1.5 本文结构安排 .....            | 29 |
| 第 2 章 手部检测 .....            | 30 |
| 2.1 基于混合特征的手部检测 .....       | 31 |
| 2.1.1 手部颜色特征 .....          | 31 |
| 2.1.2 Haar-like 特征 .....    | 33 |
| 2.1.3 Boost 分类器 .....       | 36 |
| 2.1.4 融合肤色信息的 Haar 特征 ..... | 42 |
| 2.1.5 实验数据分析 .....          | 43 |

|       |                                                    |    |
|-------|----------------------------------------------------|----|
| 2.2   | 基于 HOG 的左右手识别 .....                                | 47 |
| 2.2.1 | 梯度方向直方图 (HOG) .....                                | 49 |
| 2.2.2 | 主成分分析 (PCA) .....                                  | 52 |
| 2.2.3 | 特征分类 .....                                         | 56 |
| 2.2.4 | 实验与分析 .....                                        | 58 |
| 2.3   | 结论 .....                                           | 64 |
| 第3章   | 手部跟踪 .....                                         | 66 |
| 3.1   | Mean-shift 简介 .....                                | 67 |
| 3.1.1 | 核函数 .....                                          | 68 |
| 3.1.2 | Mean-shift 向量形式 .....                              | 69 |
| 3.1.3 | Mean-shift 概率密度梯度证明 .....                          | 70 |
| 3.1.4 | Mean-shift 算法优缺点 .....                             | 72 |
| 3.2   | Mean-shift 跟踪算法 .....                              | 72 |
| 3.2.1 | 目标特征提取 .....                                       | 74 |
| 3.2.2 | 相似度度量 .....                                        | 76 |
| 3.2.3 | 搜索策略 .....                                         | 78 |
| 3.3   | 混合特征手部跟踪 SHS (SIFT-Hist-Shift) .....               | 78 |
| 3.3.1 | 连续混合特征手部跟踪 CASHS (Continuously Adaptive SHS) ..... | 83 |
| 3.4   | 实验数据分析 .....                                       | 85 |
| 3.5   | 结论 .....                                           | 91 |
| 第4章   | 手势识别: 静态手势 .....                                   | 92 |
| 4.1   | SIFT (SIFT-Hist-Shift) 特征向量 .....                  | 93 |

|       |                                                                             |     |
|-------|-----------------------------------------------------------------------------|-----|
| 4.2   | 尺度空间极值检测 .....                                                              | 94  |
| 4.2.1 | 关键点定位 .....                                                                 | 98  |
| 4.2.2 | 关键点方向确定 .....                                                               | 100 |
| 4.2.3 | 特征向量生成 .....                                                                | 102 |
| 4.2.4 | 特征匹配 .....                                                                  | 104 |
| 4.3   | 基于多特征融合的手势识别 .....                                                          | 105 |
| 4.4   | 实验结果和分析 .....                                                               | 107 |
| 4.5   | 结论 .....                                                                    | 113 |
| 第5章   | 手势识别：动态手势 .....                                                             | 115 |
| 5.1   | 预备知识：模糊集合和手势 .....                                                          | 116 |
| 5.1.1 | 模糊集合 .....                                                                  | 116 |
| 5.1.2 | 手势的模糊性 .....                                                                | 117 |
| 5.1.3 | 模糊集描述的手势 .....                                                              | 118 |
| 5.2   | 基于 FCRFs (Fuzzy based Condition Random Fields) 的手势标注 .....                  | 122 |
| 5.2.1 | 条件随机场 CRFs (Condition Random Fields) .....                                  | 122 |
| 5.2.2 | 模糊条件随机场 FCRFs (Fuzzy based Condition Random Fields) .....                   | 123 |
| 5.2.3 | 实验和分析 .....                                                                 | 126 |
| 5.2.4 | 小结 .....                                                                    | 137 |
| 5.3   | 基于 FLDCRFs (Fuzzy based Latent-dynamic Condition Random Fields) 的手势识别 ..... | 137 |

|       |                                                                         |     |
|-------|-------------------------------------------------------------------------|-----|
| 5.3.1 | 隐动态条件随机场 LDCRFs (Latent-dynamic Condition Random Fields)                | 138 |
| 5.3.2 | 模糊隐动态条件随机场 FLDCRFs (Fuzzy based Latent-dynamic Condition Random Fields) | 139 |
| 5.3.3 | 实验和分析                                                                   | 144 |
| 5.3.4 | 小结                                                                      | 151 |
| 5.4   | 结论                                                                      | 152 |
| 第6章   | 结论与展望                                                                   | 153 |
| 6.1   | 研究内容总结                                                                  | 153 |
| 6.2   | 下一步研究和展望                                                                | 158 |
|       | 参考文献                                                                    | 162 |

## 1.1 研究背景

手势识别技术是一种基于计算机科学技术，以通过模式识别方法对人类手势动作进行解释为目标的新兴技术。基于视觉的无标记手势识别技术则是一种通过对摄像头捕获的无标记手势视频内容进行分析，进而实现手势识别的技术。使用这种方式的人机交互系统具有安装方便、成本低廉以及用户体验好等特点，它具有广泛的用途和巨大的商业价值。目前已有越来越多的研究机构和人员在该领域开展相关研究工作。以下我们对手势识别在国内外的研究背景进行介绍。

手势识别典型系统流程如图 1-1 所示：

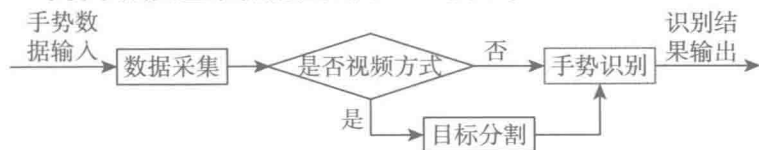


图 1-1 手势识别典型系统流程构成

首先,通过数据采集装置采集手势的相关数据,这些数据包括手部运动轨迹、手型和运动速度等,不同系统有不同的数据采集装置,常见的如摄像头、数据手套等;其次,通过特定的方法进行目标分割,本研究中目标分割的主要目的是找到“手”,从而找到要分析的目标;最后,借助各种模式识别方法对手势进行识别,如用于常见的静态手势的模板法和用于动态手势的辨识模型法等。由于在使用数据手套方式进行数据采集的手势识别系统中,手部的运动信息可以直接获取到,因此该类系统不需要进行手部目标分割。

手势识别作为一门新兴的技术,它在不同领域有不同的用途<sup>[1-4]</sup>。随着研究的深入,尤其是一些关键技术不断被突破,手势识别应用范围在不断扩展。以下就现有手势识别系统所涉及的应用领域作介绍。

- 健康助理。对于那些不方便使用语言或打字的人,可以通过手势识别设备,通过手势方式与外界进行交流。通过手语到文本的自动翻译,手势识别系统为远程手语交流和手语信息存储提供了可行性。
- 手指指向方向辨识。采用手指指向自动识别技术可以对机器人进行控制,并帮助其辨识所关注的对象或描述的内容<sup>[5]</sup>。
- 计算机交互。具有手势识别能力的计算机交互系统,允许用户使用手势完成常用的计算机操作,这种方法较传统的接触式键盘和鼠标具有较大优势<sup>[6-10]</sup>。

- 虚拟控制。对于一个需要寻找或者接近物理控制装置而花费大量时间的系统来说,手势可以被作为一种可行的替代机制。例如在车中控制车外的设备,或者电视机控制都是这种类型的应用<sup>[11]</sup>。通过使用手势识别技术可以控制远程多种设备,这种控制信号不仅可以指定所希望的反馈,而且可以指定要控制的设备<sup>[12-14]</sup>。
- 浸入式游戏技术。在视频游戏交互中,使用手势交互技术可以使游戏者经历更加身临其境的感觉和更强交互感的体验。

根据应用目标、环境及设备等的不同,在手势识别中的数据采集设备主要分为基于数据手套和基于图像两种大类,后者又分为深度感知摄像机、立体摄像机和普通摄像机等图像采集设备。

- 数据手套。通过电磁或者惯性跟踪设备,数据手套可以把手部的位置信息和旋转信息输入到计算机中。一些数据手套可以检测到手指的轻微弯曲,甚至可以给使用者提供触觉反馈。第一套商业化的手部跟踪数据手套是 DataGlove,它的数据采集设备可以检测手部位置、运动和手指弯曲<sup>[15]</sup>。
- 深度感知摄像机。这种类型的摄像机通常都是专门定制的专业摄像机,如时间相关摄像机在一定的量程内可以产生被摄场景的深度图谱,通过这种方式所获得的数据近似 3D 场景信息。对于运动范围较小的手势来说,使用这种摄像机可以达到很好的手部运动信息<sup>[16]</sup>。
- 立体摄像机。使用两个已知相对位置的摄像机,通过摄

像机标定技术和三维重建技术，可以构建目标的三维场景，其中摄像机的相对位置关系可以通过使用棋盘格标定或者红外发射装置获得。有研究者通过这种测量系统进行手部运动检测<sup>[17]</sup>。

- 普通摄像机。在资源或环境条件限制下不方便使用其他数据采集方式，但又要使用基于图像的模式识别情况下，普通的摄像机也可以被用于手势识别。虽然普通摄像机没有深度和立体摄像机那么有效，但是普通摄像机由于其设备安装简单和成本低廉等特点，其适用范围相较其他方式更广<sup>[18]</sup>。

对手势进行识别的方法众多，图 1-2 显示了在手势识别中常用的一些方法。不同算法对手势的识别率、时间复杂度和对设备的要求都不尽相同。例如，体模型包含了对手势进行详细分析的必要信息，但是实践证明这种方式非常耗时，而且为进行实时分析还需要其他辅助技术。另一方面，基于表观的模型通常来讲比较容易处理，但是由于其对复杂手势的处理存在缺陷，同样也不满足人机交互所需要的普遍适用性原则。

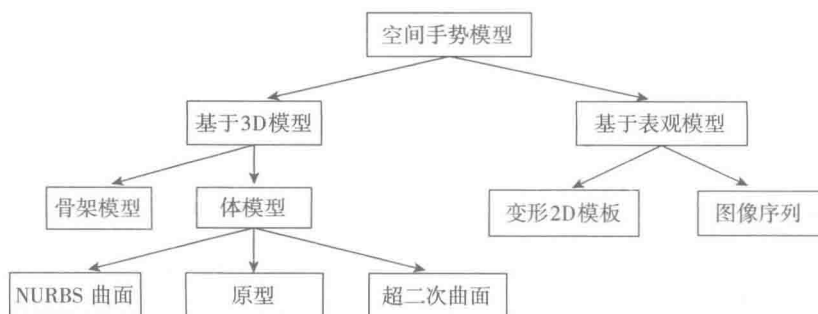


图 1-2 手势识别中常用方法

手势识别常用的方法主要分为2种不同的类型：基于3D模型和基于表观模型<sup>[19]</sup>。在3D模型法中，最重要的是使用手部运动中的手部关键帧信息获取手掌位置或者关节角度等重要参数。3D模型方法常使用体模型或者骨架模型，或者这两者的融合模型。而基于表观的系统则是直接使用视频或图像进行识别。

体模型方式在计算机动画产业和其他计算机视觉方面已经被大量地使用，这类模型通常由复杂的3D曲面构成，如NURBS或者POLYGON网格。3D建模方法的缺点是计算复杂度高，如何使其达到实时分析的程度还需要进一步研究。目前来说，更可行的方法是使用一些基本的模型组件与所要研究人体部分的相互映射（例如，用柱体表示胳膊和脖子，用球体表示头部），并且通过对这些基本模型的分析实现对手势运动的分析。此外，一些抽象的结构体，如超二次曲面和广义柱体在身体各部位拟合中表现出更好的性能，其中最引人注目的是这些结构体的参数相对简单。

针对3D模型耗时的处理过程和系统包含大量参数的问题，可以对某部分只使用一个含关节角度参数的简化模型来表示，这就是所谓的骨架表示法<sup>[20]</sup>。在骨架法中构建了虚拟的人体骨架，并且与人体特定部位相互映射。在构建虚拟骨架时主要使用被映射部分的位置、方向以及这些被映射部分的相互关系（例如关节间的角度以及相互位置和方向）。骨架模型的优势在于它只使用了一个关键参数，同时算法速度较快。由于参数较少，因此它可以使用基于模板库的模式匹配进行识别。同时，关键点的使用可以使检测程序将重点放在这些关键位置信息分析上，并有助于简化问题。

基于表观模型的方法不使用身体的空间信息，而是借助模板库直接从图像或视频中获取<sup>[21]</sup>。这种模型是基于人体的可变形2D模板，如手部。这种可变形模板通常是由一系列目标轮廓插值点构成。众多插值函数中最简单的是线性插值，该法从给定的一系列点的可变参数和外部形变计算一个平均形状。在手势识别中这些基本模板的模型通常主要用于手部跟踪，但是也可以用于简单的手势分类。另一种方法是使用图像序列作为手势模板。这种方法中的参数要么是图像本身，要么是从这些图像中得到的一些特征，在这种方法中，大多数时候仅仅使用一维（单目标）或三维（立体）。

## 1.2 研究意义

手势识别技术在应用和理论上都有比较重要的研究意义。从应用上讲，一方面使用手势识别技术的人机交互产品具有方便、灵活的特点，另一方面由于手势识别技术本身存在不少技术难题，目前手势识别技术应用还不广泛，产品还不成熟，因此有必要对相关产品进行进一步研发，随着这些技术问题的解决，其应用领域将不断扩大；从理论上讲，目前的手势识别研究还处于发展中，还有较多的待研究领域以及为提高识别率而对现有理论进行研究和改进的需求。

在手势识别方面有较多的问题亟待研究：首先，对于手势识