

手势识别技术综述

胡友树 安徽合肥工业大学工业培训中心 230601

摘 要: 本文简单介绍了手势识别的发展过程以及技术难点, 在分析了手势的定义、手势建模后, 具体阐述了目前广为研究的基于数据手套的手势识别和基于计算机视觉的手势识别, 最后介绍了手势识别技术在人机交互中的应用。

关键词: 手势 识别技术 数据手套 视觉

引言

人与计算机的交互活动越来越成为人们日常生活的一个重要组成部分, 特别是最近几年, 随着计算机技术的迅猛发展, 研究符合人际交流习惯的新颖人机交互技术变得异常活跃, 也取得了可喜的进步, 这些研究包括人脸识别、面部表情识别、唇读、头部运动跟踪、凝视跟踪、手势识别以及姿势识别等等。总的来说, 人机交互技术已经从以计算机为中心逐步转移到以人为中心, 是多媒体、多种模式的交互技术。

手势是一种自然、直观、易于学习的人机交互手段。以人手直接作为计算机的输入设备, 人机间的通讯将不再需要中间的媒体, 用户可以简单地定义一种适当的手势来对周围的机器进行控制^[1]。手势研究分为手势合成和手势识别, 前者属于计算机图形学的问题, 后者属于模式识别的问题。手势识别技术分为基于数据手套和基于计算机视觉两大类。

以人手直接作为输入手段与其它输入方法相比较, 具有自然性、简洁性和丰富性、直接性的特点。因此说用计算机来识别手势提供了一个更自然的人机接口。但是由于其难度较大, 目前的研究结果尚不能实用化。

1. 手势识别方法的发展^[1]

最初的研究主要集中在做一种专用硬件设备来进行输入。例如数据手套, 即人可以戴上一个类似于手套的传感器, 计算机通过它可以获取手的位置、手指的伸展状况等丰富信息。如 1993 年 B. Thomas 等人做的自由手遥控目标的系统是凭借数据手套作为输入的媒介, 但这需要实验者戴上一个专用设备。

之后人们又致力于标记手势的研究, 即通过在手上作标记, 例如在手腕和手指处贴上或画上特殊颜色的圆点, 用来识别手势。这虽然给识别带来了方便, 但同样给实验者带来麻烦。

最后人们终于把注意力集中到自然手上, 通过专用加速硬件和脱机训练, 一些研究者成功地研制了手势系统, 但其识别的手势仅限几种。例如 Freeman 和 Roth 等人提出的基于方向直方图的手势识别系统。1994 年, 作者高文等人提出了一种静态复杂背景中手势目标的捕获与识别, 1995 年又提出了动态复杂背景中手势目标的捕获与识别方法。

2. 手势的定义

由于手势 (gesture) 本身具有多样性和多义性, 具有在时间空间上的差异性加上不同文化背景的影响对手势的定义是不同的。这里把手势定义为: 手势是手或者手和臂结合所产生的各种姿势和动作, 它包括静态手势 (指姿态, 单个手形) 和动态手势 (指动作, 由一系列姿态组成)。静态手势对应空间里的一个点, 而动态手势对应着模型参数空间里的一条轨迹, 需要使用随时间变化的空间特征来表述。手势和姿势 (posture) 的主要区别在于姿势更为强调手和身体的形态和状态而手势更为强调手的

运动^[3]。

3. 手势识别的技术难点^[1]

3.1, 现在大多数的研究重点都在静态手势的识别, 其技术难点都以下几点:

3.1.1, 手势目标检测的困难;

3.1.2, 手势目标识别的困难。

目标的实时截取是指在人以复杂的背景条件下从图像流中截取目标来, 这是机器视觉主要研究的课题之一。目前已有许多针对专用自动视觉系统的较为成熟且易于实现的技术。例如, 利用目标窗与背景窗的脂肪图分割目标的方法, 基于多图像信息的目标分割方法以及二维熵的阈值分割方法等。

3.2, 手势识别则是根据人手的姿态以及变化过程来解释其高层次的含义, 提取出具有几何不变性的特征是其关键技术。

由于手势具有以下特点:

3.2.1, 手是弹性物体, 故同一种手势之间差别很大;

3.2.2, 手有大量冗余信息, 由于人识别手势关键是识别手指特征, 故手掌特征是冗余信息;

3.2.3, 手的位置是在三维空间, 因此难以定位, 并且计算机获取的图像是三维向二维的投影, 因此投影方向很关键;

3.2.4, 由于手的表面是非光滑的, 因此易产生阴影。

上面两个问题目前还远未解决, 具体实现时必须加一定的限制条件。

4. 计算机手势识别技术和方法^[3]

利用计算机识别和解释手势输入是将手势应用于人机交互的关键前提。

4.1, 目前人们采用了不同手段来识别手势:

4.1.1, 基于鼠标器和笔, 缺点是只能识别手的整体运动而不能识别手指的动作; 优点是仅利用软件算法来实现, 从而适合于一般桌面系统。需要说明, 仅当用鼠标光标或笔尖的运动或方向变化来传达信息时, 才可将鼠标器或笔看作手势表达工具。这类技术可用于文字校对等应用。

4.1.2, 基于数据手套 (Data Glove), 主要优点是可以测定手指的姿势和手势, 但是相对而言较为昂贵, 并且有时会给用户带来不便 (如出汗)。

4.1.3, 基于计算机视觉, 即利用摄像机输入手势, 其优点是不干扰用户, 这是一种很有前途的技术, 目前有许多研究者致力于此项工作。但在技术上存在很多困难, 目前的计算机视觉技术还难以胜任手势识别和理解的任务。

4.2, 所采用的手势识别方法目前主要有^[6]:

4.2.1, 模板匹配技术, 这是一种最简单的识别技术, 它将传感器输入的原始数据与预先存储的模板进行匹配, 通过测量两者之间的相似度来完成识别任务。

4.2.2, 神经网络技术, 这是一种较新的模式识别技术, 具有自组织和自学习能力, 具有分布性特点, 能有效抗噪声和处理不完整的模式以及

具有模式推广能力。

4.2.3, 统计分析技术, 通过统计样本特征向量来确定分类器的一种基于概率的分类方法。在模式识别中一般采用贝叶斯极大似然理论确定分类函数, 该技术的缺点是要求人们从原始数据中提取特定的特征向量, 而不能直接对原始数据进行识别。

目前较为实用的手势识别是基于数据手套的, 因为数据手套不仅可以输入包括一维空间运动在内的较为全面的手势信息, 而且比基于计算机视觉的手势在技术上要容易得多。但是基于视觉的手势识别是未来的趋势。

5. 基于数据手套的手势识别^[4]

数据手套是虚拟现实技术中广泛使用的交互设备。基于数据手套的手势识别严格来说其实不能算作一种真正的手势识别, 传统的交互设备, 如鼠标 (笔) 等其实也可以认为是一些手势输入设备。基于数据手套手势输入 (Glove-based Gesture Input) 的优点是输入数据量小, 速度快, 就直接获得手在空间的三维信息和手指的运动信息。可识别的手势种类多, 已能够进行实时地识别。

基于数据手套的手势识别目前较多采用神经网络等方法。由于神经网络可以用静态和动态的输入, 很适合用快速、交互的方式进行训练, 而不必用一种解析的方式定义传递特征。还可以根据用户个人情况调整网络的连接权值, 使手势识别程序能适应不同的用户。存在的不足是手势识别网络依赖于设备。当使用不同的手套设备时, 要改变网络的拓扑结构, 并重新训练网络得到新的连接权值。

6. 基于视觉的手势识别^[2]

自然手的研究是未来的趋势, 因此基于视觉的手势识别是顺应潮流的。因此本文将介绍基于视觉的手势识别技术研究的几个主要方面。

一个基于视觉的手势识别系统的总体构成如图 1 所示。首先, 通过一个或多个摄像机获取视频数据流。接着, 系统根据手势输入的交互模型检测数据流里是否有手势出现。如果有, 则把该手势从视频信号中切分出来然后, 选择手势模型进行手势分析, 分析过程包括特征检测和模型参数估计识别阶段, 根据模型参数对手势进行分类并根据需要生成手势描述。最后, 系统根据生成的描述去驱动具体应用。

6.1, 手势建模

手势模型对于手势识别系统至关重要, 特

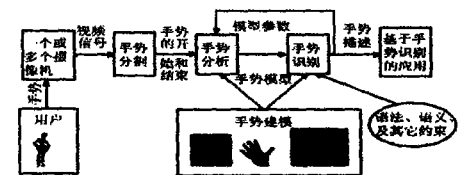


图1 连续动态手势识别系统构成图

(下转第 41 页)

一种更实用、更安全的用户管理方式。所以我们在学生宿舍区采用基于 802.1X 的认证方式。

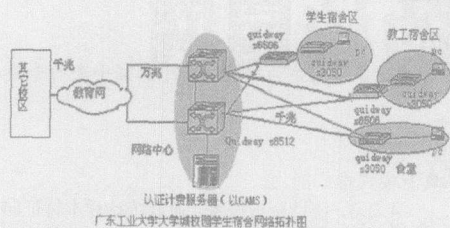


图2

图2是我校学生宿舍区网络接入的拓扑图。本方案的认证过程如3所述。这种方案在性价比、安全性和可管理方面较有优势。用户发出认证报文，是使用特定的组播MAC地址，设备发送用户的报文使用单播MAC地址，解决了认证报文的广播问题，其他用户不能侦听到认证过程，从而无法知道用户的密码、账号和MAC地址。

址。认证通过后的用户账号与接入设备IP地址、VLAN、MAC地址与端口进行绑定。通过这种绑定，用户只能从指定的地点，使用特定的计算机认证上网，可以增强用户认证的安全性，防止账号盗用和非法接入。认证服务器系统在加强身份认证安全性的同时，配合IP地址分配策略控制，还可以防止IP地址盗用和冲突，有效地阻止了用户篡改MAC地址和IP地址的行为。认证计费形式灵活多样，操作方便。可防止学生设置代理服务器，比如单网卡代理、双网卡代理、终端服务代理或HTTP等其他方式的代理。还有网络性能高，对学生用户的认证和计费不会对网络性能造成瓶颈，而降低上网的速度。802.1X认证是分布式认证，没有单点故障。

5. 结束语

本文简要分析了802.1X认证工作原理及工作机制，对我校学生宿舍区的802.1X认证

计费解决方案特点的进行分析，该方案在充分发挥交换式以太网接入优点的前提下，可以有效地解决学生的网络认证和网络安全问题。考虑接入网络安全性、管理性和建网成本，作为宽带网接入的安全解决方案，802.1x必将是未来的发展主流。

参考文献：

- 1、《使用IEEE802.1X技术优化校园网计费管理》米鲁沈，孙剑颖等，武汉：华中科技大学学报（自然科学版），2003.10:224-226
- 2、《802.1x协议及其在宽带接入中的应用》高祥，周林，http://www.net130.com/2004/6-6/222941-2.html
- 3、《IP技术与综合宽带网》[M]巴继东，北京：人民邮电出版社，2000
- 4、《802.1x在MA5200E/F上的实现和应用》宋海波，华为技术报第164期（维护专刊之三十），2002.12.31

(上接第42页)

别是对确定识别范围起关键性作用，模型的选取根本上取决于具体应用。对于某个给定的应用，一个非常简单并且粗糙的模型使用图像梯度方向直方图去跟踪人手以及识别静态手势可能就是充分的。然而，如果要实现自然的人机交互，那么必须建立一个精细有效的手势模型如3D人手模型，使得识别系统能够对用户所做的绝大多数（如果不是所有的）手势作出正确的反应（识别或拒识）。从目前的文献来看，几乎所有的手势建模方法都可以归结为两大类：基于表现的手势建模和基于3D模型的手势建模。基于表现的手势模型是建立在手（臂）图像的表现之上，它通过分析手势在图像（序列）里的表现特征去给手势建模。基于3D模型的手势建模方法考虑了手势产生的中间媒体（手和臂），一般遵循两步建模过程：首先给手和臂的运动以及姿态建模，然后从运动和姿态模型参数估计手势模型参数。

基于3D手（臂）模型的手势模型又可以分为体模型、网格模型、几何模型以及骨架模型。人体3D体模型主要用于跟踪和识别身体姿态。最常使用的3D模型是3D骨架模型，其参数是经过简化的关节角度参数和指节长度。人手的物理特性可以为3D骨架模型提供两组约束：静态约束（关节角度范围）和动态约束（运动依赖关系）。使用手（臂）3D模型时存在两个主要问题：其一，参数空间的维数高；其二，通过视觉技术获取这些模型的参数困难重重并且非常复杂。

第一类基于表现的手势模型使用2D灰度图像本身建立手势模型。第二类基于表现的手势模型建立在手（臂）的可变形2D模版的基础上。第三类基于表现的手势模型建立在图像属性的基础上。第四类基于表现的手势模型通过计算图像运动参数，抽取手势模型参数。这类表现模型主要用在动态手势识别里。

6.2. 手势分析

手势分析阶段的任务就是估计选定的手势模型的参数。分析阶段一般由特征检测和参数估计两个串行任务组成。在特征检测过程中，首先必须定位做手势的主体（人手）根据所用的线索不同，可以把定位技术分为基于颜色定位、基

于运动定位以及多模式定位等三种。绝大多数颜色定位技术依赖于直方图匹配或者利用皮肤的训练数据建立查找表的方法。基于颜色定位技术的主要缺点是在不同的光照条件下皮肤颜色变化较大，这经常导致未被发现的皮肤或误检测出非皮肤区域。利用限制性背景或者颜色手套，使得高效地、甚至实时地定位人手成为可能，然而对用户以及对接口设备施加了明显限制。

尽管不同手势模型的参数各不相同，但是用于计算模型参数的图像特征基元通常是非常相似的。常用的图像特征基元包括灰度图像、二值影像、区域、边界及轮廓或者指尖等。

模型参数的估计要根据模型涉及的参数不同而采用不同的参数估计的方法。对于3D手模型通常包括初始参数估计和参数随时间更新两个环节。对于2D的四类基于表现的手势模型要根据分类来采取不同的方法。

6.3. 手势识别

手势识别就是把模型参数空间里的轨迹（或点）分类到该空间里某个子集的过程。静态手势对应着模型参数空间里一个点，而动态手势则对应着模型参数空间里的一条轨迹，因此它们的识别方法有所不同。静态手势识别算法包括基于经典参数聚类技术的识别和基于非线性聚类技术的识别。

与静态手势不同，动态手势涉及时间及空间上下文。绝大多数动态手势被建模为参数空间里的一条轨迹。不同用户做手势时存在的速率差异、熟练程度会在轨迹的时间轴上引起非线性波动。如何消除这些非线性波动是动态手势识别技术必须克服的一个重要问题。考虑到对时间轴的不同处理，现有的动态手势识别技术可以分归一类：基于隐马尔可夫模型（Hidden Markov Models, HMM）的识别，基于动态时间规整（Dynamic Time Warping, DTW）的识别，基于压缩时间轴的识别。

7. 手势识别在人机交互中的应用^[4]

作为一种三维交互设备。

7.1. 用于虚拟环境的交互。如虚拟制造和虚拟装配、产品设计等。虚拟装配通过手的运动直接进行零件的装配，同时通过手势与语音的合成来灵活的定义零件之间的装配关系。还可

以将手势识别用于复杂设计信息的输入。

7.2. 用于手语识别。手语识别的研究目标是让机器看懂聋人的语言。手语识别和手语合成相结合，构成一个“人一机手语翻译系统”，便于聋人与周围环境的交流。手语识别同样分为基于数据手套的和基于视觉的手语识别两种。

7.3. 用于多通道、多媒体用户界面。正如鼠标没有取代键盘，手势输入也不能取代键盘、鼠标等传统交互设备，这一方面由于手势识别的设备和技术问题，一方面也由于手势固有的多义性、多样性、差异性、不精确性等特点。手势输入在人机交互中应用的精髓不在于用未独立地用作空间指点，而是为语言、视线、唇语等交互手段通道提供空间的或其他的约束信息，以消除在单通道输入时存在的歧义。这种做法是试图以充分性取代精确性。

7.4. 用于机器人机械手的抓取。机器人机械手的自然抓取一直是机器人研究领域的难点。手势识别，尤其是基于数据手套的手势识别的研究对克服这个问题有重要的意义，是手势识别的重要应用领域之一。

8. 总结

本文介绍了手势识别涉及的各个方面，包括手势的定义、建模以及识别的关键技术，并且简单阐述了手势识别的应用。

参考文献

- [1]《灵境（虚拟现实）技术的理论、实现及应用》汪成为，高温，王行仁，北京：清华大学出版社，1996
- [2]《基于视觉手势识别的研究综述》任海兵，祝远新等，电子学报，2000.2
- [3]《计算机手势输入及其在人机交互技术中的应用》方志刚，小型微型计算机系统，1999.6
- [4]《手势识别技术及其在人机交互中的应用》李清水，方志刚，人类工效学，2002.3
- [5]《VR环境中产品设计手势的定义与合成》邹晨，张数有等，工程图学学报，2000.2
- [6]《基于数据手套的手语识别方法》杨吉广等，智能计算机接口与应用进展，北京：电子工业出版社，1997