手势识别的神经网络方法

焦圣品① 徐大海 白英彩

摘 要 手势识别正成为人机交互技术研究中的一种重要模式。本文介绍了手势的基本概念及手势输入的两种方式，分析了人机交互过程中用户产生手势和计算机系统"感知"手势的基本过程，提出了手势识别的两种途径。在此基础上，使用数据手套 5th Glove'95 输入手势，运用 BP 神经网络实现了静态手势的识别。

主题词 手势 数据手套 神经网络 人机交互

一、引言

在当前广泛使用的 WIMP 图形用户接口(Graphical User Interfaces based on Windows, Icon, Menus and Pointing devices)中，用户通过键盘、鼠标向计算机输入信息的交互方式已成为人机通信的瓶颈[1]，因为，从用户到计算机的通信是串行的，而且用户也不能以习惯的方式(如手势、语音)与计算机进行交互。

随着多媒体技术的发展，计算机已经具备了处理语音、图形、图象和文字等多种通信媒体的能力，从计算机到用户的通信带宽得到了进一步的提高。为了提高计算机的使用效率，克服现有交互技术的种种不足，必须研究新的输入技术[2]如语音输入技术、在线手写体输入技术、手势输入技术，以扩大从用户到计算机的通信带宽，使用户更加方便自然地和计算机系统交互。

特别地，新的人机交互技术——虚拟现实(VR-Virtual Reality Technology)技术的发展更进一步促进了手势识别技术的研究[3]。VR 技术是一种以人为中心的新型人机接口技术。它利用计算机技术生成由三维逼真的图象、声音等构成的虚拟环境(VE—Virtual Environment)，刺激用户的感官，同时向用户提供三维输入技术，使用户能以日常生活的经验、技能与虚拟环境进行交互。VR 技术不同于传统的 WIMP 人机接口技术的主要方面是，它特别注重用户能否在与 VE 交互时产生沉浸感(Immersion)。制约用户产生沉浸感的主要因素是，用户能否自然、直观地运用三维交互手段与 VE 进行交互[3]。要达到这一目标，一个重要的方面就是研究手势识别技术，使用户能使用手势以自然、直观的方式与 VE 进行交互。

二、手势的基本概念

在日常生活中，不论我们从事何种工作，几乎都离不开手。我们通过手接触周围的环境，了解外部世界；使用手操作工具或物体完成工作任务；除了使用语言之外，我们还使用手势作为一种通信手段与周围的人进行交流。

广义地讲，手势是指人有意识地作出的手的运动(包括手指的弯曲、伸展，手腕的转动和手在空间的移动)。不论是操纵工具或物体执行某种任务，还是进行交流，手势都表达出手势者的某种意图。由于长期的学习和人与人之间的相互影响，人们作出的手势表达的含义基本上是确定的，即在确定的场合下表达确定的含义。如我们握拳伸出大拇指朝上，表示对某人或某事的赞赏。因此，手势包括两方面的含义：手的运动及其表达出的手势者的意图或者说手势所表达的概念。在使用"手势"一词时，一般不作区别，可根据上下文判别。手势的识别就是根据用户的手势识别手势的含义。

手势的执行是一个动态的过程，其特征表现在手指弯曲引起的手的形状的变化、手在空间的位置和方位的变化，需要从时间和空间两方面来描述。根据手势的时变特征，可将手势分为静态手势(Static Gestures)和动态手势(Dynamic Gestures)。静态手势是指只需用手的形状特征来表示的手势。如在美国手势语(ASL)中，用手指的弯曲引起的手的形状变化来表示英语中的字母。从测量的角度看，静态手势可用某一时刻手的空间特征的测量值来表示。动态手势是指需要使用随时间变化的空间特征来描述的手势。如起重工人在工作中伸出手掌朝上并摆动表示将货物上吊。从测量的角度看，动态手势需用一段时间内手的空间特征的一组测量序列值来表示。

由于人的手有 20 多个自由度，运动十分灵活、复杂；而且不同的人，手的大小不同。因此，同样的手势，不同的人作出时手的运动会存在差别；同一个人在不同的时间、地点作出手势也不一样。可见，手势又具有随机性。

本文主要研究静态手势的识别。

① 焦圣品 上海交通大学金桥网络工程中心 上海 200030

三、手势的输入

手势的输入有两种实现方式[4]，一种方式是基于视觉的手势输入(Vision-based Gesture Input)[4,5,6]，即使用摄象机运用计算机视觉技术来捕获手势，实现手势的输入。基于视觉输入方式的优点是对用户的运动限制较少。但由于计算机视觉技术仍然不成熟，这种输入方式存在很多不足，例如，为了便于从背景图象中提取出手势，需要采用特殊的背景或要求用户戴上标有特殊颜色的手套；需要处理的数据量庞大，处理方法复杂，难以实现地识别手势[4,5]。

手势输入的另一种方式是使用数据手套(Data Glove)及跟踪器来跟踪手势的变化，实现手势的输入。这种方式称为基于手套的手势输入(Glove-based Gesture Input)。基于手套的输入方式的优点是，输入数据量小、速度高，能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，且能对手势进行实时地识别[4]。

数据手套是虚拟现实技术中广泛使用的交互设备。数据手套有多种，性能不同[6]。比较简单的数据手套只用几个传感器来测量手指总的弯曲度，而复杂的数据手套(如 CyberGlove)使用 18 个或 22 个传感器测量手指的弯曲，更高级的数据手套还具有力反馈。本文使用 Fifth Dimension Technologies(5DT)公司生产的数据手套(5th Glove'95)实现手势的输入。5th Glove'95 手套的每个手指上有一个传感器，分别测量每个手指的总弯曲度，分辨率为 8 位，即可以测量每个手指的 256 种弯曲状态。另外，该手套上还有 2 轴倾斜传感器，分别测量手的俯仰角和倾斜角。总的采样速率可达 200Hz。5th glove'95 与 PC 机之间通过标准的 RS-232 串行口相连。

四、手势识别方法的分析

从用户产生手势到系统"感知"手势的过程可用下图表示。

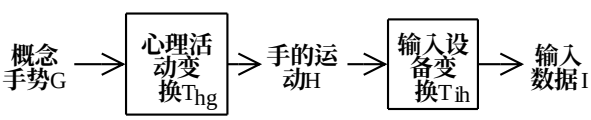

图1　用户产生手势和系统感知手势的原理

由图 1 可知，在交互过程中，用户根据系统的输出确定交互任务，明确操作要求，形成概念手势 G。经过一定的心理活动，用户将概念手势以手的运动 H 表达出来。这个心理活动过程以变换 T_{hg} 表示，该变换十分复杂。系统输入设备监测用户手的运动，产生输入数据 I。T_{ih} 表示从用户手的运动 H 到输入数据 I 的变换，即输入设备的变换。当使用数据手套及跟踪设备实现手势输入时，I 就是手套等的输出数据序列；当使用摄象机采用计算机视觉方式来捕获手的运动时，I 就是系统通过摄象机得到的手图象(图象序列)。上述手势的产生和"感知"过程可用下式表示

$$H = T_{hg}G \tag{1}$$

$$I = T_{ih}H \tag{2}$$

由(1)(2)两式有

$$I = T_{ih}T_{hg}G = T_{ig}G \tag{3}$$

从模式识别的角度可以认为手势 H 是需要识别的模式，概念手势 G 是模式 H 所属的类别。为了识别手势，必须以某种参数化方式来描述手的运动 H。静态手势对应于描述参数空间的随机点；动态手势对应于参数空间中手势执行时间内的一条轨迹。

显然，手势的识别就是从输入数据 I 求得手势 G。由式(1)(2)和(3)知，手势的识别的途径有两种。一种是直接从输入数据 I(或从其中提取得到的特征)求得手势 G，即

$$G = T_{ig}^{-1}I \tag{4}$$

其中 T_{ig}^{-1} 是 T_{ig} 的逆变换

另一种是先按式(5)从输入数据 I 求得手的运动 H

$$H = T_{ih}^{-1}I \tag{5}$$

再按式(6)从手的运动 H 求得手势 G。

$$G = T_{hg}^{-1}H \tag{6}$$

其中 T_{ih}^{-1}、T_{hg}^{-1} 分别是 T_{hg}、T_{ih} 的逆变换。

换言之，第二种方式就是先根据变换从输入数据 I 求得手的运动 H，再根据变换确定用户作出的手势 G。

由于数据手套 5th Glove'95 只能测量每个手指的总弯曲度，且手册没有给出如何根据手套的读数计算每个手指上的每个关节弯曲角的方法。因此，为了简化处理工作，我们采用第一种途径，运用 BP 神经网络实现从 I 到 G 的变换，从而实现手势识别。

初步规定系统必须识别日常生活中常用的 9 种手势，即 {G}= {"1","2","3","4","5","6","7","8","9"}。在识别时这些手势，不考虑手的方位，对应的手形状如下表 1。

表 1　9 个手势的手形描述及编号、期望输出

手势	手的形状	序号	期望输出
"1"	食指前伸，其它手指成完全弯曲状	0	{1, 0, 0, 0, 0, 0, 0, 0, 0}
"2"	食指和中指前伸，其它手指成完全弯曲状	1	{0, 1, 0, 0, 0, 0, 0, 0, 0}
"3"	食指、中指和无名指前伸，另两指成弯曲状	2	{0, 0, 1, 0, 0, 0, 0, 0, 0}
"4"	大拇指弯曲，其它四指伸展朝前	3	{0, 0, 0, 1, 0, 0, 0, 0, 0}
"5"	所有手指伸展朝前	4	{0, 0, 0, 0, 1, 0, 0, 0, 0}
"6"	大拇指和小手指伸展，其它 3 指成弯曲状	5	{0, 0, 0, 0, 0, 1, 0, 0, 0}
"7"	大拇指和食指、中指指尖接触，另 2 指弯曲	6	{0, 0, 0, 0, 0, 0, 1, 0, 0}

“8”	大拇指和食指伸展, 其它 3 指弯曲	7	{0, 0, 0, 0, 0, 0, 0, 1, 0}
“9”	食指先朝前伸展再将其前端关节弯曲, 其它手指成弯曲状	9	{0, 0, 0, 0, 0, 0, 0, 0, 1}

五、用 BP 神经网络识别静态手势

神经网络是一种大规模并行处理网络, 由许多具有非线性映射能力的神经元组成。神经元之间通过权相连。神经网络能实现复杂的非线性映射, 映射关系是通过学习(或训练)得到的。神经网络具有很高的计算速度、很强的容错性和鲁棒性, 特别适合用于模式识别。

BP 神经网络是目前使用最广泛、方便的一种单向传播的多层前向神经网络, 除输入输出节点外还有一层或多层隐层节点。同层节点之间没有耦合。输入信号从输入层节点, 依次传过各层节点, 最后到达输出层节点。每一层节点的输出只影响下一层节点的输出。神经元的激活函数为 S 型函数, 即

$$f(u) = 1.0/(1.0 + \exp(-u)) \tag{7}$$

本文使用的BP 神经网络为三层前馈的, 即输入层、中间隐含层和输出分别有 5、50、9 个神经元, 如下图 2。

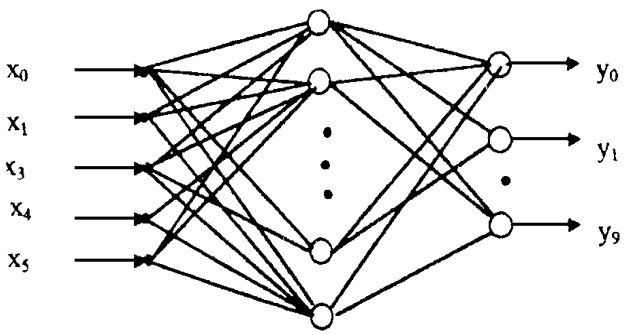

图 2　BP 网络拓扑结构

网络第 l(l= 1, 2 .)层有 n_l 个神经元, $y_k^{(l)}$ 为该层第 k 个神经元的输出, 有:

$$net_k^{(l)} = \sum_{j=0}^{n_{l-1}} w_{kj}^{(l)} y_j^{(l-1)} \tag{8}$$

$$(k = 1, 2, \cdots n_1) \tag{9}$$

$$y_k^{(l)} = f(net_{k(l)})$$

其中 $w_{kj}^{(l)}$ 为第 l- 1 层神经元节点到该节点的连接权值。

为了使用BP 神经网络识别手势, 将手势{G}编号, 确定神经网络的期望输出, 如下表(1)。另外, 将数据手套的读数进行归一化处理, 将表示每个手指弯曲的读数折算到[0, 1]区间内。实验表明, 虽然每个手指传感器的分辨率为 8 位, 但实际使用时, 由于受手指弯曲/伸展范围的制约, 每个手指传感器读数的最大(握拳时手指完全弯曲产生最大值)和最小值(所有手指伸展时产生最小值)并都不是 255 和 0。考虑到每个手指读数的最大、最小值不同, 按式(10)对每个手指的读数进行归一化处理,

$$x_i = \frac{r - min_i}{max_i - min_i} \quad (i = 0, 1, 2, 3, 4) \tag{10}$$

其中, r 为手套第 i 个手指的读数

max_i, min_i 分别为第 i 个手指读数的最大、最小值

x_i 为归一化处理后手指弯曲值, 即神经网络的输入

离线样本采集程序、网络训练程序和手势在线识别程序都用C++ 实现。离线样本采集程序的基本工作过程是, 用户右手戴上数据手套并按要求作出一个手势, 保持稳定后(持续时间 0.2s 以上)左手按下键盘上相应的数字键, 则经过归一化处理后的手指弯曲值和对应的期望输出值(组成一个样本)就被记录并保存在手势库文件中。每个手势重复 10 次, 共记录了 90 个样本, 再用样本对BP 网络进行离线训练, 将训练后网络的连接权值存储在网络描述文件中, 供手势在线识别程序使用。

手势的在线识别程序在执行时, 首先从网络描述文件中读入连接权值, 设置网络, 之后等待用户作出规定的手势。其基本工作过程为, 捕获用户作出的手势, 对捕获的手势数据进行归一化处理, 再输送给识别模块处理。识别模块由BP 网络和查表部分组成, 它根据网络的输出值确定手势序号, 从而得到识别结果。实验表明, 对于每种手势, 用户作两次, 系统都能正确地识别(训练数据取自同一个用户)。

六、讨论

Fels 等研究了将手势转换为语音的神经网络接口, 从其实现看, 该接口使用神经网络直接将手势映射为语音参数, 如基波频率和峰值。而本文研究的手势识别则是运用神经网络实现了手势到抽象符号的映射。运用神经网络的优点是, 可以根据用户个人情况调整网络的连接权值, 使手势认别程序能适应不同的用户。使用其他模式识别方法则无法做到这点。

本文是直接根据手套的输入来识别手势。这种方式存在的不足是, 手势识别网络依赖于设备。当使用不同的手套设备时, 可能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。如果按照第二种途径来实现手势识别, 从输入数据中得到通用的手势表示 H, 比如手指各个关节的弯曲角、手的方位和位置, 神经网络根据手势表示 H 来识别手势, 则可以使手势识别网络仅与手势的表示有关而与设备无关, 即设备无关性。

此外, 由于没有对手势执行过程种手的位置和方位进行跟踪的设备, 本文识别的手势仅限于静态手势, 不包括动态手势。为了识别任意种类的手势, 今后需要进一步研究动态手势的识别技术。

参考文献

[1]Andries van Dam, Post-WIMP user interfaces, Commun-

(下转第 64 页)

条不紊。从中不难看出, Excel 的这种功能在试验数据处理中是非常实用和有效的。

如果准备处理的下一张表格与上一张格式相同, 那么最有效的处理方法是: 将上一张表格文挡复制成为赋予新文件名的文挡, 然后对这个文挡进行相应的数据"覆盖录入", 当数据输入完毕, 相应的数据也处理完毕。

4. 数据图表的形成

对于包含一大堆数据的表格, 数据间的关系一时难以体现出来, 采用座标纸来逐点描绘出数据曲线, 即可体现出数据间的关系。Excel 提供了一种高效的数据图表方式, 只需选取好目标数据块, 然后单击"图表向导", 在多种图表类型中(面积图、条形图、柱形图、折线图、饼图、组合图、圆环图等)选取所需的样式, 然后单击"完成"即可自动将这些数据转换形成最能体现其关系的图表。有了这种工具, 那种逐点描绘座标纸的时代一去不复返。图 1 所示的是供试品种从第二天到第十四天发芽数分布曲线图, 其所用数据是从 C5 到 P5, C8 到 P8, C704 到 P704 的数据块。

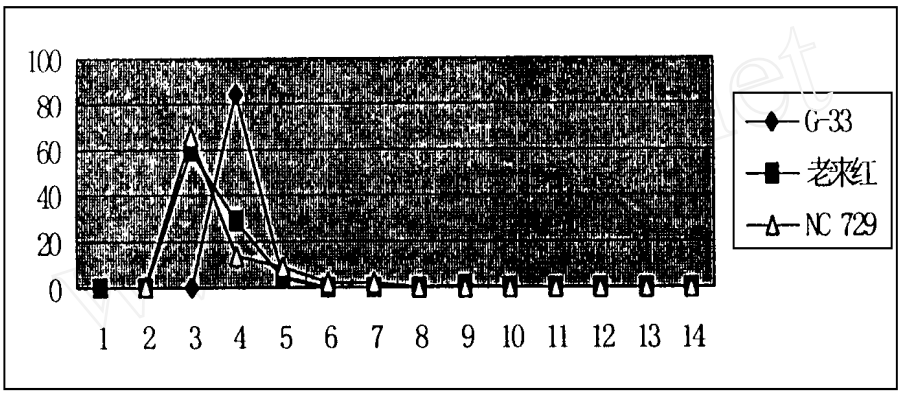

图 1　供试品种从第二天到第十四天发芽数分布曲线图

5. 表格和图表嵌入 Word 文挡

撰写科研报告、科研总结或科研论文是我们从事科研活动的一项重要工作, 用 Excel 处理所得到的数据、图表等是有价值的结果, 可喜的是这些表格或图表完全可以很方便地嵌入 Word 文挡, 然后一并打印输出。

四、结束语

自 Microsoft Office 4 X 及 Microsoft Office 97 推出以来, 现代办公自动化方式和手段发生了翻天覆地的变革, 它将功能全面的应用软件有机地集合在一起, 使它们发挥出更为强大的作用。本文所谈及的 Excel 便是其中的一个重要组件, 在速度上, 该软件充分发挥了 Windows95 之所长, 支持 32 位应用程序、长文件名、桌面块捷方式及抢先式多任务处理, 使用户可在应用程序之间快速切换; 在使用上, 它能充分利用 Windows 95 的图形界面设计, 使用户在操作时更加轻松自如, 而且在学习、操作等方面也不是一件难事。作为农业科技人员, 花点时间和精力来学习这种数据处理方法是非常值得的, 特别是看到自己利用 Excel 来有效解决以往颇为费时费力的数据处理工作时, 更能体会到这种数据处理的轻松愉快。

参考文献

[1]木林森. 高峰霞. 中文版 Excel 97 使用手册, 北京: 清华大学出版社, 1997

[2]马育华. 田间试验和统计方法. 北京: 农业出版社, 1983

[3]许美玲, 卢秀萍. 烟草种子超干燥保存技术研究. 贵阳: 种子, 1996(3)

(收稿日期: 1998 年 9 月)

(上接第 29 页)

cations of the ACM, V 40 N 2, Feb 1997, p63- 67

[2]Kenneth Tolman, Multimodal communication, Proceedings of SPIE-The International Society for Optical Engineering, Soc for Optical Engineering, WA, USA, v1785 1993, p298- 306

[3]Mauro Figueiredo, Klaus Bohm, Jose Teixeira, Advanced interaction techniques in virtual environments, Computers & Graphics v17 n6 Nov-Dec 1993, p655- 661

[4]David J. Sturman, David Zeltzer, A survey of glove-based input, IEEE Computer Graphics and Applications, Vol14, No. 1, 1994, p30- 39

[5]J. Davis, M. Shah, Visual gesture recognition, IEE Proceeding: Vision, Image and Signal Processing, v141 n2 Apr 1994, p101- 106

[6]Lee Jintae, Kunii Tosiyasu L. Model-based analysis of hand posture IEEE Computer Graphics and Applications, v15 n5 Sep 1995, p77- 86

(收稿日期: 1998 年 10 月)