基于手势识别的人机交互发展研究

任雅祥
(绍兴县水务集团有限公司，浙江 绍兴 312030)

摘 要:近年来手势识别技术的快速发展,基于手势识别技术的人机交互应用系统的建立使得人机交互的发展前景广阔。从手形、手势和手形手势的建模出发,介绍了模板匹配、特征提取、神经网络和隐马尔可夫模型4种手势识别的方法,并且综述了基于手势识别技术人机交互的发展,详细介绍了3类人机交互系统:漫游型系统、编辑型系统和操作型系统。

关键词:手形; 手势; 手势识别; 人机交互
中图法分类号:TP391 **文献标识码**:A **文章编号**:1000-7024 (2006) 07-1201-04

Suvery of human-computer interaction development based on hand posture and gesture recognition

REN Ya-xiang
(Shaoxing County Water Group Company Limited, Shaoxing 312030, China)

Abstract: The rapid progress of techniques in hand posture and gesture recognition provides a promising approach for human-computer interaction (HCI) and based on this progress many applications are established. Basic concepts of hand posture and hand gesture, two modeling methods of hand posture, two key problems of hand gesture and four recognition techniques are reviewed. And three types of HCI systems influenced by hand gesture recognition are introduced in detail.

Key words: hand posture; hand gesture; recognition; human-computer interaction

0 引 言

整个社会的计算机化为我们带来一种新的交互方式,那就是人机交互(human-computer interaction)。相对于计算机软硬件的日新月异,人机交互技术发展缓慢,文本时代的象征—键盘,在人机交互设备上依然占据主导地位。虽然图形界面带来了鼠标、手柄和触摸屏等,但是它们仍然不够有效和自然。当虚拟现实越来越为人熟知的时候,人们自然而然的想到用手势在虚拟环境中表达某种意图。比如,人们用手指的指向来表示前进的方向,用挥手来表示告别。在过去的几年中,越来越多的应用系统以手势识别作为人机交互的接口。因此,有必要对于手势识别以及在此基础上的人机交互发展作一个综述。

本文的第1章概述了手形(hand posture)、手形的建模以及手形的识别技术。在第2章,介绍了手势(hand gesture)及其识别技术。第3章介绍了3类基于手势识别技术的人机交互应用系统:漫游型系统(navigation systems)、编辑型系统(editing systems)和操作型系统(manipulating systems)。在第4章,讨论了基于手势识别技术的人机交互系统今后的发展方向,并且对本文作了一个总结。

1 手 形

手形是一种静态的手势,可以由手的位置、方向等来描述。用来表示手形的模型可以分成两种:基于三维的模型和基于图像的模型。基于三维的模型由若干从输入图像中提取出来的三维属性组成,这些三维属性包括:指尖坐标、手掌中心、手掌方向和各个关节的角度等。基于图像的模型由一系列图像属性构成,如边界线、轮廓线、点的分布、颜色和纹理等。

1.1 基于三维的模型

在J.Lee和T. L.Kunii的系统中,指尖信息被用来构造三维模型[1,2];在J.M.Rehg和T.Kanade的"Digiteyes"系统中[3],手指用圆柱体来建模,指尖用半球体来建模,如图1所示,这些都是基于三维的模型。

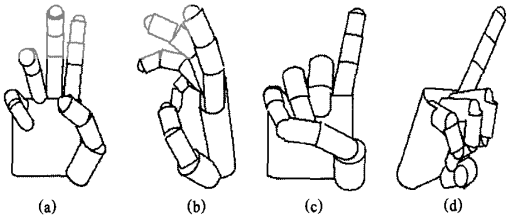

图1 "Digiteyes"系统中的4个手形

为了增强特性检测(如指尖的检测)的效果,在点特征之外,许多系统利用了线特征[2,4,5]。一个典型的例子就是用样条曲线来拟合手形[2,6],另一个例子是结合对边界线和轮廓线的

收稿日期:2005-01-01。
作者简介:任雅祥 (1969—),男,浙江绍兴人,工程师,研究方向为人机交互与供排水系统自动化控制。

观察来实现人体跟踪[2,7]。

无论手形在三维上如何建模，来自图像的二维属性和三维模型的比较不可避免。这种比较可以通过两种途径：把从图像中提取出来的三维信息与三维模型比较[8]；或把从三维模型提取二维信息与图像中提取出来的二维信息比较。但无论用何种方法，三维模型和二维图像之间的比较缺乏效率。

1.2 基于图像的模型

在基于图像的模型中，边界和轮廓是两个可从输入图像中提取的最直接的属性。J.Segen 使用基于边界线的技术，使其系统可以实时识别十种手形[9]。而 Ulrich Brockl-fox 使用了基于轮廓线的技术，使其系统可以实时识别 16 个自由度(degree of freedom)的三维交互[10]。T.E. Starner 使用了一组属性来建立手形模型，它们包括 4 个元素：手的 x 和 y 坐标、最小惯性轴的角度和轮廓椭圆的离心率。

另一种基于图像的模型是主成份因素分析(principal component analysis)。对于一个图像集合的主成份因素分析确定了一个图像的正交集，而此正交集可以用来描述原图像集合。相似的图像在一个特征空间上有相似的投影，因此在特征空间上的距离可以用来衡量图像的相似性。所以，这种方法可以用于比较输入手形和手形模板的相似性。Jerome Martin 和 James L.Crowley 在他们的系统中采用了这种技术[11]。

综上所述，基于三维的模型表达能力强，几乎能够表达一切手形，但是缺乏效率、计算繁琐。而基于图像的模型简单、高效，但是缺乏通用性。所以两种方法各有所长。

1.3 手形识别

手形识别技术包括：模板匹配、特征提取和神经网络等，本节介绍前两种方法，神经网的方法络将在 2.3 节中介绍。

模板匹配：总的来说，模板匹配是把输入数据分类到各个模板集的过程[12]。手形集合中的每一个手形都根据其三维或图像属性建立模板。然后把输入数据与模板进行匹配，找出最相匹配的模板。一个典型的模板匹配的例子就是计算输入手形数据和模板手形数据的距离，以找出距离最小的。所有距离差值的绝对值之和或平方和可以作为距离量度[12]。模板匹配是一种最简单的手形识别方法。当模板集相对较小，这种方法非常准确，但是当模板集较大的时候，在整个模板集中搜索将非常费时。

特征提取：在特征提取过程中，输入数据中的低层特征被提取出来，经过分析而转化成包含手形语义的高层特征，然后用高层特征进行手形识别[12]。低层特征可以是轮廓线、指尖坐标、关节角度、颜色、纹理等。而具体的手形语义就可以从这些特征中分析得到。在 Ulrich Bröckl-Fox 的系统中[10]，二维低层特征包括：由 B 样条曲线表示的轮廓线，轮廓线长度、重心位置；而手形语义集合包括模式手形和操作手形。3 个模式手形是“照相机在手中”、“场景在手中”和“飞过场景”。而操作手形包括平移、旋转、后退、滑行、重置和取消上一步动作。

2 手 势

手势是用来表达或强调一种意念、感情或态度一个手的动作。因此，从起初的意图到最终的动作，手势由一个时间段内所做的一系列手形组成。所以在手势识别的过程中，一个基础的工作就是进行手势分解，即把手势按时间顺序分解成若干手形。

2.1 手势分解

Kendon 把一个手势分为 3 个阶段：准备、动作和收回。Quek 则定义了一系列规则来规范手势分解。①整个手势包含 3 个阶段：缓慢的初始动作、加速的中间过程和返回初始位置；②在中间过程中，手做了一个包含具体语义的手势；③手在静止位置附近的小扰动不算手势；④手的动作不应超出某个空间范围；⑤静态手势应该在一个有限的时间段内被识别；⑥重复的动作可以作为手势。

根据以上规则，手势的组成类似于语言的组成。手形的识别相当于词法分析，而手势的识别则相当于语言的解释。因此一系列在语言分析中成功应用的技术可以被引入手势识别领域。其中比较成功的是隐马尔可夫模型(hidden markov model)。我们将在 2.3 中介绍它在手势识别中的应用。

尽管几乎所有手势都可以被分成 3 个阶段，但是只有第 2 阶段才有具体的语义。所以一个手势如何解释决定于第 2 阶段如何分解和解释。分解第 2 阶段的困难在于人们的习惯不同，即不同的人可以以不同的方式做同一个的动作。对于这个问题，引入神经网络将是一个合适而有效的方法。我们将在 2.3 中讨论神经网络在手势识别中的应用。

2.2 手势分类

在手势识别中另一个重要的问题是如何按照语义对手势分类。也就是说，手势可以表示什么？手势表示的语义可以分为哪几类？

Wexelblat 认为手势可以分为 6 类[13]：符号型手势、模式型手势、模仿型手势、图标型手势、指示型手势和强调型手势。符号型手势最简单，像食指和中指形成“V”状表示胜利(victory)等。模式型手势经常表示一些情感，如耸耸肩、摊开手表示不知道。模仿型手势经常用来模仿人们使用、操纵某个物体，像模仿操纵汽车的方向盘等。图标型手势经常在某些特定的场景中表示某些事件的发生等。指示型手势就是用手指指一个物体，或指示物体的位置或运动方向等。强调型手势主要是在人们说话时起到辅助作用。

另一种分类方法如图 2 所示[14]。手的动作可以分为两类：手势和无意识动作。手势又有两种：操作型手势和通信型手势。操作型手势表示在某环境中对某些物体的操作，如推箱子。通信型手势就是用手势作为交互的工具，如手语。通信型手势又可以分为动作和符号。符号与上一种分类方法中的

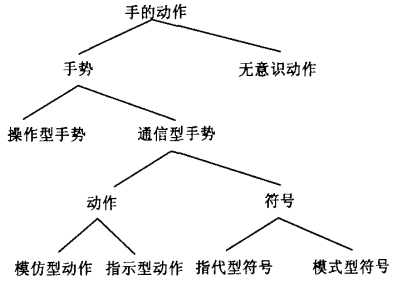

图 2 手势分类

符号型手势一样表示一种事先约定的意思，分为指代型符号和模式型符号。动作则包含模仿型动作和指示型动作，这与上一种分类中的模仿型手势和指示型手势一致[14]。

2.3 手势识别

神经网络：考虑到输入图像流的噪音以及不同的人的不同习惯，许多系统同采用神经网络的方法来识别手形和手势。传统的神经网络都用来识别手形，因为所有的连接和神经元都设定为瞬时或固定时延的。要使用神经网络识别手势，可以使用循环神经网络(recurrent networks)或时延神经网络(time delay neural networks)。循环神经网络应用反馈连接，使得神经网络可以记录当前场景和手势按时序分解而成的手形。而时延神经网络则把输入数据按时间分隔，因此，一个动态输入数据流被分隔成了一个静态输入数据集，而静态数据集可以由传统的神经网络来处理。Michael R. Berthold 提出了一个时延神经网络的变体—时延放射状基函数网络 (time delay radial basis function network)。该网络既保持了时延神经网络的分时特性，又具有 RBF 网络的训练快速的特性。

Murakami 在其系统中使用神经网络进行手势识别。他的系统是一个包含了 403 个节点的循环神经网络，能够识别 42 个日本手语中的手形。系统对于训练过神经网络的人的手形的识别率高达 98 %，而对于没有训练过神经网络的人的手形的识别率为 77 %。Anders Sandberg 提出了一种基于复合型人工神经网络的识别模型，该模型融合了放射状基函数和贝叶斯分类网络。此系统在场景信息的支持下可以识别手势，甚至一些比较复杂的手势。

隐马尔可夫模型：隐马尔可夫模型已经在口语识别中得到了成功的应用，同时由于手势与口语之间的相似性，隐马尔可夫模型被引入到了手势识别领域。隐马尔可夫模型是一个有限自动机，可以由一个四元组$M=(K,\Sigma,\Delta,s)$来表示。K是一个有限状态集，Σ是一个输出符号集，Δ是一个状态转换的集合，s是开始状态。在Δ中的每个状态转换都被赋予概率值，表示从一个状态转换到另一个状态的概率。

Starner和Pentland在他们的系统中用隐马尔可夫模型来识别一个美国手语的子集。该系统对有语法和无语法的输入进行了测试，其中在包含语法的输入中，经过训练的数据和未经训练的数据的识别率分别是 99.5 %和 99.2 %。在不包含语法的输入中，经过训练和未经训练的数据的识别率分别是 92 %和 91.3 %。

3 人机交互界面

任何人机交互界面的成功的关键都在于其对人的意图的支持程度。这种支持包括跟踪人原有的意图、刺激人应该有的意图、帮助人执行其意图[15]。相比于传统的交互方式，手势识别为人机交互界面提供了一种更加自然的方式。现在，基于手势识别的人机交互系统大概可以分为 3 类：漫游型系统、编辑型系统和操作型系统。

3.1 漫游型系统

这里的漫游是指飞过或走过一个虚拟环境，就像人们平常漫游风景区一样。虚拟环境中的漫游有一些基本要求：允许用户在任何位置以任何视点进行观察；允许用户拾起或放置物体；允许用户看到自己的动作和其他用户的动作；实现一系列约束，使得不可能的状态不会发生，如行进时穿透墙体等。在虚拟环境中漫游可通过几种方法：①在虚拟环境中直接用手势表示漫游方向。Sato Y.的系统定义了两类手势，一类是手掌全部张开，以手掌的位置和朝向的变化来改变漫游者的视点或使漫游者移动。另一类手势只有一个手指张开，用来指示物体，如手指指向某一个建筑物，然后系统返回此建筑物的名字。②在虚拟环境中用手操纵特定的工具进行漫游。在Stoakley的系统中，用户首先在一个手持的虚拟环境的缩略模型上确定视点，然后通过这个缩略模型的图形平移功能飞往指定点。

在以上两种方法中，手直接或间接地确定了前进的方向和身体的朝向。但是在真实世界中，手并不作此用。因此，研究人员正在努力寻找其它方法来表示前进的方向和身体朝向，他们的目的是解放手的功能。一些系统已经开始用头和脚来表示前进方向和身体朝向。Fuhrmann 在其系统中用头的方向表示前进方向和漫游速度。这种方法的优点是手解放了，但是其缺点也很明显，随意的头部运动将被系统误解而造成误操作。而且当人朝一个方向运动时，他不能看其它方向。而Joseph J在其漫游系统中用脚来表示是否前进或后退，用身体的倾斜表示左转或右转，在一定程度上解决了此问题。

3.2 编辑型系统

在这里，编辑的领域包括文本编辑、图形编辑等。Ho-Sub Yoon 在其图形编辑系统中用手势作为主要编辑工具。在系统中，手势分为两类，一类是图元手势和编辑命令，另一类包括数字 0~9 和字母。图元手势包括：圆、三角形、矩形、弧、水平线和垂直线。编辑命令包括移动、拷贝、取消、交换、除去、关闭。如图 3 所示。

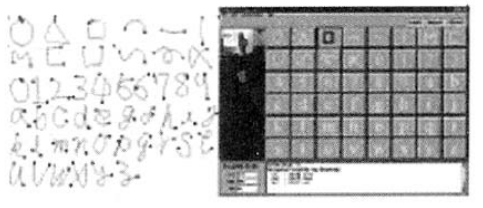

图 3 左图是手势数据，右图是系统接口

在 C. v. Hardenberg 的 Brainstorm 系统中，用户可以先把自己想要输入的物体通过无限键盘输入到墙上，然后可以用手指重新排列这些物体。如图 4 所示。

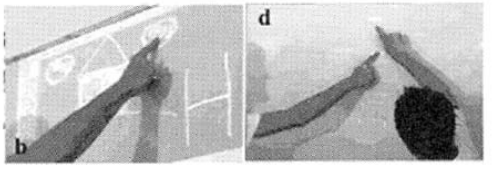

图 4 左图是用手势画图，右图是多个用户用手势重新调整物体位置

3.3 操作型系统

物体操作和系统操作是操作型系统中的两个非常适合手

势的操作类型。物体操作就是对物体本身进行操作，如对物体的挤压、拖拉等。系统操作就是通过手势远程控制一个系统，如用手势来控制计算机操作系统，用手势辅助演讲。

C.V.Hardenberg 的 FingerMouse 系统用手指来代替鼠标，在其另一个系统 FreeHandPresentation 中，用户可以通过手势翻动幻灯片来辅助演讲。在此系统中，伸出两个手指表示“下一页”，3 个手指表示“前一页”，5 个手指表示“幻灯片菜单”。如图 5 所示。

图 5　左图是用手势控制浏览器，右图是用手势翻动幻灯片

用手势进行物体操作十分自然、高效，特别是对于实体和曲面的建模和变形。在 J.Dorman 的计算机雕刻系统中，用户可以用手抓住一个虚拟曲面，并可以对此曲面延其法向进行推拉。手的张开程度决定了被推拉区域的大小。如图 6 所示。

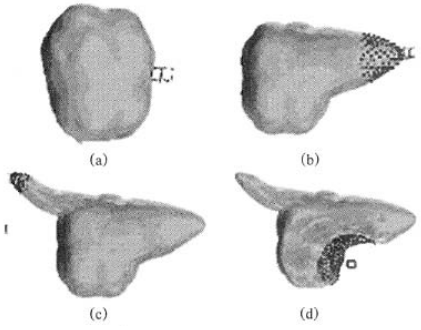

图 6　(a)物体的初始形状。(b)物体被拉伸。
(c)较小面积的物体被拉伸。(d)物体被推挤。

4　结束语

提供一个自然而且有效的人机交互界面始终是人机交互研究的目的。口语、手势、形体语言和人脸这四者作为人机交互的手段在虚拟环境中得到了初步的应用，它们的应用前景十分广阔。

在本文中，我们已经概述了快速发展的手势识别技术以及在此基础上人机交互接口的进步。手形有两种建模方式：基于三维的建模和基于图像的建模。前者有强大的表达能力，几乎能够表达所有的手形，但是缺乏效率。而后者简单高效，但是缺乏通用性。手势由一段时间内的一系列手形组成。手势的分解和分类是手势识别的两个重要问题。在此基础上，我们讨论了几种识别技术：模板匹配、特征提取、神经网络和隐马尔可夫模型。在手势识别技术的影响下，新的人机交互应用系统可以分为 3 类：漫游型系统、编辑型系统、操作型系统。

手势识别在人机交互中的应用还处于起步阶段[5]。手势作为人机交互的手段虽然自然简单，但是不准确。所以，有时候口语和手势应该结合起来以达到既准确又自然方便的效果。在多用户虚拟环境中，人脸又可以作为区分用户的属性。所以，口语、手势和人脸在虚拟环境中可以起到互补的作用，它们的综合利用在人机交互系统中将产生深远的影响。

参考文献:

[1] Lee J,Kunii T L.Model-based analysis of hand postur[J].*Computer Graphics and Applications*, 1995,15(5):41-43.

[2] Ying Wu,John Lin,Thomas S Huang. Capturing natural hand articulation[A].Proc IEEE Int'l Conf on Computer Vision (ICCV' 01)[C]. 2001. 2:426-432.

[3] Rehag,James M,Takeo Kanade.DigitEyes:vision-based human hand tracking[C].*Technical Report CMU-CS*-93-220,*School of Computer Science,Carnegie Mellon University*, 1993.

[4] Rehg J,T Kanade.Model-based tracking of self-occluding articulated objects[A].Proc of IEEE international Conference of Computer Vision[C]. 1995.612-617.

[5] Jakub Segen,Senthil Kumar. Shadow gesture: 3d hand pose estimation using a single camera[A].Proc IEEE Conf on Computer Vision and Pattern Recognition[C]. 1999. 479-485.

[6] Kuch James J,Huang Thomas S. Vision-based hand modeling and tracking for virtual teleconferencing and telecollaboration [A].Proc of IEEE Int'l Conf on Computer Vision[C]. *Cambridge, MA*, 1995.666-671.

[7] Jon Deutscher, Andrew Blake, Ian Reid. Articulated body motion capture by annealed particle filtering[A].Proc IEEE Conf on Computer Vision and Pattern Recognition, volume II[C]. *Hilton Head Island, South Carolina*, 2000.126-133.

[8] Schlenzig J,Hunter E, Jain R. Recursive identification of gesture inputs using hidden markov model[A].Proc Second Annual Conference on Applications of Computer Vision[C].1994.187-194.

[9] Segen J. Controlling computers with gloveless gestures[A]. Proc Virtual Reality Systems Conf[C]. 1993.2-6.

[10] Bröckl-Fox U. Real-time 3-d interaction with up to 16 degrees of freedom from monocular image flows[A].Proc of IWAFGR' 95 (Zurich)[C]. 1995.172-178,

[11] Martin J, Crowley J.An appearance-based approach to gesture recognition[A].Proceedings of Ninth International Conference on Image Analysis and Processing[C]. 1997. 340-347.

[12] Joseph J, LaViola Jr. A Survey of hand posture and gesture recognition techniques and technology [R]. *Brown University: Department of Computer Science*, 1999.

[13] Wexelblat.Afeature-based approach to continuous-gesture anal-ysis [D]. *SM Thesis, MIT Program in Media Arts and Sciences*, 1994.

[14] Pavlovic I, Sharma R, Huang T S. Visual interpretation of hand gestures for human-computer interaction: A review [C]. *IEEE Trans on PAMI*, 1997.

[15] Haasbroek L J. Advanced human-computer interfaces and intent support: A survey and perspective[C]. *Systems, Man and Cybernetics*, 1993.350 -355.