

基于隐性马尔可夫模型的手势识别设计和优化

刘珠峰, 周 良, 丁秋林

(南京航空航天大学 信息科学与技术学院 计算机系, 南京 210016)

摘 要: 新型的笔式交互技术要求能够高效地识别用户手势, 适应用户的手绘风格。建立了基于隐性马尔可夫的手势识别模型, 在此基础上提出了在重采样阶段的中点补偿和编码阶段的方向编码优化方法。实验结果表明该识别模型能以更精简的采样点数量表示手势并给出良好的识别结果, 减少了模型训练的运算量。

关键词: 手势识别; 隐性马尔可夫模型; 重采样; 方向编码

中图分类号: TP391.3

文献标志码: A

文章编号: 1001-3695(2011)06-2386-03

doi:10.3969/j.issn.1001-3695.2011.06.107

HMM-based sketch recognition system: design and optimization

LIU Zhu-feng, ZHOU Liang, DING Qiu-lin

(College of Information Science & Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China)

Abstract: Newly pen-based interaction technology requires efficient user gesture recognition and user style adaption. This paper built an HMM-based sketch recognition system and used a new method called midpoint-compensation in resampling phase and an improved direction coding method for optimization. The experiment shows that the sketch recognition model can recognizes user gesture efficiently meanwhile with smaller data to represent the gesture and less computation for training the model.

Key words: gesture recognition; HMM; resampling; direction coding

近年来, 基于多媒体的人机交互技术在国内外受到了高度的重视。随着便携式终端的增多和基于电阻/电容触摸设备的日渐普及, 众多研究者投入了基于语音与手势的人机交互界面的研究^[1]。作为对人类在自然界交流方式的模拟, 手势具有表达自然、易于学习的优点, 在不具备标准全尺寸键盘的设备上, 更可以作为输入指令, 是快速实现用户意图的理想工具。

手势相关方面的研究至今已经取得了很多研究成果。Rubine^[2]提出了 GRANDMA 系统, 该系统通过提取手势中笔画的几何特征和动态特征对手势进行描述, 并可以快速添加新的手势指令。Mark^[3]提出了一种利用九宫格矩阵对手势的笔画路径进行编码, 但不能扩展至丰富的手势指令库。在借鉴了神经网络在其他认知领域的成功后, 也有学者将其应用到手势识别领域的研究中。Sun^[4]、Golubitsky 等人^[5]提出了一种基于 SVM 向量机的手绘图形识别, 可以较好地适应不同用户的使用习惯并具有较高的识别率。Willems^[6]将手势看做图形, 从全局计算特征量。

由于用户绘制手势可以被视为依次绘制的一组连续笔画的时序过程, 这与语音的输入具有一定的相似性, 而随着隐性马尔可夫模型(HMM)在语音识别领域的成功, 不少研究者尝试将该模型应用到手势识别领域, 尤其是以像素驱动的手势识别中。本文以手势的笔画方向为观测序列建立了 HMM 模型, 并优化了特征点的提取和观测值的计算。

1 手势识别系统框架

手势识别系统框架如图 1 所示。手势识别模块的总体目标是构建一个健壮的分类器, 以对手绘手势进行分类和识别。

从图中可以看出, 主要的处理工作包括获取用户的输入数据、对数据的预处理、对特征的提取和鉴别、利用分类器对手势分类、后期处理并对结果进行评估。



图 1 手势识别匹配框架

虽然各种手势识别的模型大体上相同, 但由于 HMM 是基于统计的概率评价模型, 每一步所得出的结果都会影响后续的处理。预处理所得数据的多少直接影响到特征提取和训练的代价; 特征的选择影响着模型状态的数目及模型的结构。本文将从预处理和特征提取着手, 优化手势识别系统, 减少训练代价并提高手势识别的效率。

1.1 手势重采样

手势可视做用户在系统交互界面绘制的一系列点坐标的集合, 可以通过系统的 MouseMove 事件来获取。由于用户在绘制笔画过程中快慢不同, 在同一时间内计算机得到的笔画采样点的数量不尽相同。因此, 手势相较于图符, 除了二维的构成特征外, 还应包括先后和速率信息。可以定义一个手势为

$$\text{gesture} = \{ \text{point}(x, y), \text{index}, \text{time} \}$$

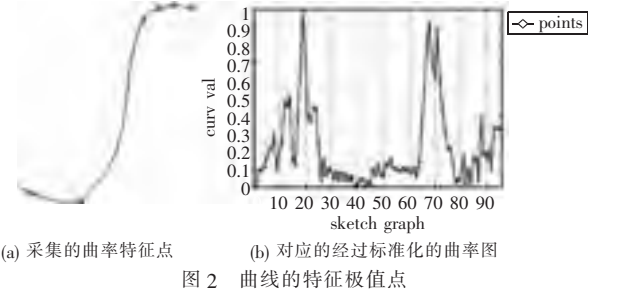
其中: $\text{point}(x, y)$ 代表了当前点在屏幕上的位置, index 表明了点在原始采样中的顺序, time 表示了用户自提笔开始绘制 $\text{point}(x, y)$ 时所经过的时间。用户的速率信息则可以通过点的空间和时间关系来体现。这些信息都绑定在一起, 在初始采样中记录, 并不随着后续的处理而变化。

预处理中最重要的部分为点集的重采样, 在以往的研究中主要基于平均距离和特征角点两种方法。平均距离以在相同

收稿日期: 2010-10-28; 修回日期: 2010-12-10

作者简介: 刘珠峰(1986-), 男, 江苏泗洪人, 硕士研究生, 主要研究方向为人机交互技术(lzf8888@gmail.com); 周良(1966-), 男, 副教授, 主要研究方向为人机交互技术、信息系统与信息安全、知识工程; 丁秋林, 男, 教授, 主要研究方向为 CIM、DSS、MIS。

的距离 d 内固定最终采样点个数为 K ;另一种方法是计算各点的曲率极值点,并取阈值过滤掉曲率较低的点。第一种方案容易将真正的特征点删除,这对于本文中模型特征值的计算是致命的;第二种方案在草图识别中经常使用,但曲率极值点往往在局部过密,单纯设置阈值难以反映缓和的方向变化。如图 2 所示,图中用空心方框标注的点为曲率极值点,图中第二点和第三点之间具有缓和的方向变化,这让单纯地采集曲率极值点无法反映特征需求。

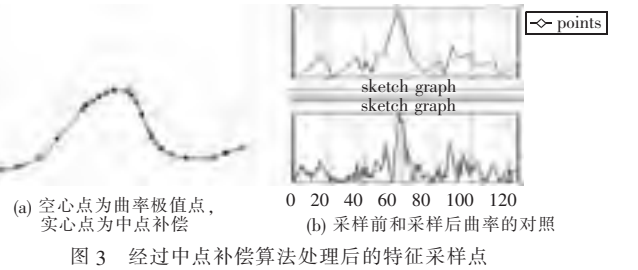


本文提出了一种基于曲率特征的中点补偿采样算法,既能较好地反映曲线的曲率特征变化,又能够有效控制点集的规模。定义 $distance_1$ 、 $distance_2$ 为阈值距离, $p(i)$ 为笔画中的原始采样点, $curv(i)$ 为 i 点的曲率值, $curv_thresh$ 为曲率点选取的阈值。该算法的描述如下:

```
for each point p(i) in the original points
if curv(i) > curv_thresh
compute distance (p(i), p(i-1)) = d1
if (d1 < distance1)
将 p(i) 与 p(i-1) 中曲率最大的点保留, 不进行中点补偿
if (d1 > distance2)
寻找原始序列中 p(i) 与 p(i-1) 之间的近似中点 k, 将 p(i) 与 k (如果存在) 加入到保留的点集中
```

$Curv_thresh$ 的选取方法在众多手势研究论文中有所描述,本系统中 $curv_thresh$ 为所有原始采样点曲率平均值的 1.12 倍;平均距离阈值的 $distance_1$ 和 $distance_2$ 需要靠经验获取,且 $distance_2$ 一般需要取到 $distance_1$ 值的 2.5 ~ 3 倍。补偿中点 k 有时并不存在,由于在笔速极快的情况下,往往系统的原始采样点不足,会产生近似中点距离 $p(i)$ 或 $p(i-1)$ 小于 $distance_1$ 的情况。对于 k 的要求是满足 $distance_1(p(i-1), k)$ 与 $distance_2(p(i-1), k)$ 均在 $0.4d_1 \sim 0.6d_1$ 。

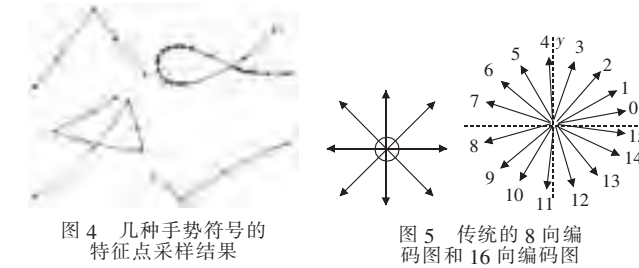
采用中点补偿算法对原始采样点处理后,采样点能够较好地反映曲率特征的变化,解决了特征点过于密集的问题,较好地描绘了整个手势的曲线走向。图 3 和 4 阐述了使用中点补偿法后的采样点采集。



1.2 特征提取

在之前的研究中,有很多特征用于表达手势的笔画特点,如 Rubine 特征^[2]、曲率和笔画速度^[4]、笔画数^[4]等。要建立一个高效的 HMM,必须能够保证观测序列具有一定的规律性,而基于方向编码可以较好地满足 HMM 建模的需求。方向编

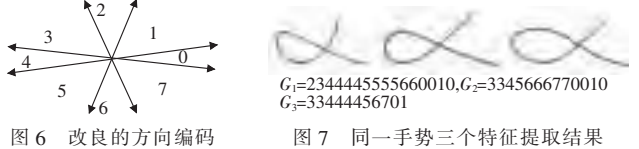
码早已应用在手势识别中^[4,7,8],手势可以描述为一系列具有方向向量构成的序列,而相近的向量可以共用一个符号来描述,而用于描述的不同符号的个数称之为观测值数目。常用的 HMM 观测值数目为 8 或者 16,而观测值则用整数来表示,如图 5 所示。一般认为 16 向由于更准确地描述了方向性,获得了更理想的观测序列。



经过实验可以发现,16 向编码并不总是能够产生良好的训练结果,有时甚至降低了模型的识别率,究其原因是因为 16 向太“精确”了。HMM 观测模型中,各个观测值是离散的,虽然观测值 6 和 7 在几何空间上很接近,但对于 HMM 来说,它们对应着不同的发射概率函数(emission probability)。因此,用户手势笔画微小的角度变化,都会对概率的估计带来巨大的影响。传统的 8 向编码同样有问题,因为人们很难在手绘时做到严格的垂直。而垂直/水平与斜线有着不同的自然语义,对于模型来说显然应该纳入不同的观测值。Derk 等人^[8]提出了一种改进,类似于图 5 中 16 向编码图的旋转,顺时针或者逆时针旋转一个角度 T ,但容易将一些倾斜不大的斜线纳入直线/垂直线的范畴。经过分析,本文提出了改良的基于方向的笔画编码方法。该方法对竖直、水平具有良好的容错性。编码的基本算法如下所示:

- a) 对于传统的方向编码图,顺时针旋转 $\pi/8$ 。
- b) 将 0、2、4、6 区间围绕它们所包含的 x 、 y 轴将所包含区间由 $\pi/4$ 缩减为 $\pi/5$ 。
- c) 根据向量所处的区间判定其观测值。

改良后的编码如图 6 所示。改良的基于方向的笔画编码更精确地判定了直线和水平线,同时,将相近的斜线归为了同一类。图 7 为实验给出的手势笔画特征编码示例。



经过改良之后,提取的特征值更容易相近,因此在模型训练完成后可以给出较高的概率评价。

2 建立隐性马尔可夫模型

隐性马尔可夫模型假设学习处理的过程是由一系列有限的离散和隐藏的状态序列组成。这些状态被假设为是隐藏的,但可以通过发出的特征值来观测。HMM 可以表示为: $HMM = \{\pi, A, B\}$ 。其中: π 为状态初始概率; A 为状态之间的转移概率; B 为发射概率,即在给定观测值下每个状态的概率值矩阵。

在无监督条件下,状态之间可以随意跳转,即使在大量的训练样本下所得出的模型仍不能给出很好的识别率。结合手势输入的实际情况,系统采用了单向跳转的 HMM,如图 8 所示。除了起始和结束状态,每一个状态只能指向自己或者下一

个状态。而状态初始概率中 $\pi(s) = 1$, 其余初始状态概率为 0。

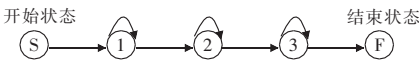


图 8 系统采用的隐性马尔可夫模型拓扑结构

本文中对模型的训练采用了传统的 Baum-Welch 算法, 通过一系列离散的观测序列循环训练 HMM 的参数。模型验证和概率计算通过传统的前/后向算法进行。

建立模型的另一主要问题在于状态数量选择。在理想情况下, 一个左向箭头的表示可能只需要两个隐藏状态 S1、S2。S1 表示直线的特征, S2 表示箭头特征。考虑由于用户的手绘风格以及特征提取阶段可能出现的误差, 两个状态的 HMM 模型并不能提供良好的分类结果, 系统通过细化状态数量的方式来建立可以良好分辨手势的模型。经过实验, 在 4~6 个状态下, 模型具有较好的识别能力。图 9 给出了矩形在不同状态数目下的识别率, 可以看出在 5 个隐藏状态下, 模型具有最好的识别效果。

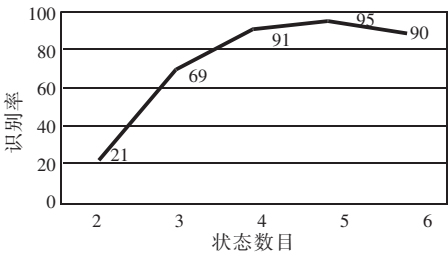


图 9 不同状态数下矩形的识别率

3 实验结果

利用本文所述方法, 系统分别用鼠标模拟笔画输入和手写板输入测试了手势识别率。在训练阶段, 系统采集了 12 个用户书写的 25 种手势, 每种手势同一用户书写 3 次, 使用这 36 个训练数据对识别器进行循环训练。在识别效果测试中, 测试用户数量为 30 人, 且均没有提前训练测试用户的绘制风格, 每人随机分配 6 种手势, 每种手势绘制 4 次。表 1 给出了部分手势的最终测试结果以及相应的识别率。

表 1 手势符号的识别率/%

手势符号					
鼠标输入	94	95	91	94	92
手写板输入	95	96	94	97	95
Derk ^[8] HMM	90	91	91	92	88

从表 1 中可以看到, 由于鼠标在输入笔画时容易产生抖动、漂移, 不如手写板输入手势流畅, 造成了在个别手势特别是

涉及曲线的手势上识别率稍低; 但是由于在手势的预处理部分去除了冗余点并优化了特征的提取, 平均识别率仍然能在 93% 以上。采用手写板输入的成绩要更好, 识别率在现阶段平均达到了 95%, 部分手势达到了 97%。与基于传统的特征点采集算法和特征提取算法的 HMM 建模相比较 (Derk 的 HMM), 本系统识别率有了明显的提高, 尤其是在手势符号较为复杂的情况下。鉴于基于 SVM、ANN 的机器学习方法版本众多、实现方法各异以及测试时所采用的手势模型不同, 系统无法与这些机器学习方法进行直观比较。与其他机器学习方法相比, 虽然其他方法也能够提供 90% 以上的识别率, 基本满足手势识别的应用需求, 但 HMM 自身特性决定了当有新手势加入时不需要对整个系统模型的现有手势进行重新学习, 从而减小了系统的重构负担, 这是 HMM 学习方法的优势。

4 结束语

本文建立了一个基于隐性马尔可夫模型的手势识别框架, 并在重采样和特征提取阶段提出了优化方法。实验结果表明, 本系统在用户手势识别的准确性方面相较于未优化前有了一定提高。下一步的工作目标是将手势识别结合到基于手绘的三维建模的原型系统上, 为建模系统提供手势功能支持, 并在实际操作中体现出手势交互相较于 WIMP 交互方式的优越性。系统在手势设计的合理性以及对用户语义的理解上仍有较大的提升空间, 该部分工作会在实践中继续完善。

参考文献:

[1] 孙正兴, 冯桂焕, 周若鸿. 基于草图的人机交互技术研究进展[J]. 计算机辅助设计与图形学学报, 2005, 17(9): 1889-1899.

[2] RUBINE D. Specifying gestures by example[J]. Computer Graphics, 1991, 21(4): 329-337.

[3] MARK W. Design and implementation of a stroke interface library[C]//IEEE Region 4 Student Paper Contest, 1997.

[4] SUN Zheng-xing, LIU Wen-yin, PENG Bin-bin, et al. User adaptation for online sketchy shape recognition[C]//Lecture Notes in Computer Science. Berlin: Springer-Verlag, 2004: 303-314.

[5] GOLUBITSKY O, WATT S. Online recognition of multi-stroke symbols with orthogonal series[C]//Proc of the 10th International Conference on Document Analysis and Recognition. 2009: 1265-1269.

[6] WILLEMS D. Iconic and multi-stroke gesture recognition[J]. Pattern Recognition, 2009, 42(12): 3303-3312.

[7] LI X, YEUNG D. On-line handwritten alphanumeric character recognition using dominant points in strokes[J]. Pattern Recognition, 1997, 30(1): 31-44.

[8] DERK A, CRAIG B. Hidden Markov model symbol recognition for sketch-based interfaces[C]//Proc of AAAI Fall Symposium. 2004: 15-21.

(上接第 2385 页)

[7] 李蔚清, 洪云轩, 吴惠中. 一种基于特征的实时 LOD 模型生成算法[J]. 系统仿真学报, 2005, 17(2): 429-431.

[8] 刘晓东, 熊海桥, 蒋立华, 等. 利用 Billboard 实现虚拟植物集群生长显示[J]. 计算机工程, 2003, 29(13): 52-54.

[9] 黄炳, 陈俊丽, 万旺根. 飞行仿真中三维云场景的渲染[J]. 上海大学学报: 自然科学版, 2009, 15(4): 342-345.

[10] 梁伟, 刘群, 吴渝. 飞机尾翼空中失火的场景模拟[J]. 计算机工程与应用, 2010, 46(9): 163-165.

[11] 肖永辉, 徐青, 周杨. 三维城市景观阴影绘制算法的研究与实现

[J]. 测绘科学技术学报, 2006, 23(4): 308-309.

[12] 张楠, 王建立, 王鸣浩. 基于图形处理器的边缘检测算法[J]. 计算机科学, 2010, 37(1): 265-267.

[13] 柳有权, 刘学慧, 吴恩华. 基于 GPU 带有复杂边界的三维实时流体模拟[J]. 软件学报, 2006, 17(3): 568-576.

[14] HOPF M, ERTL T. Hardware based wavelet transformations[C]//Proc of Vision, Modelling and Visualization Conference. 1999: 317-328.

[15] 周国亮, 冯海军, 何国明, 等. 图形处理器在数据管理领域的应用研究综述[J]. 计算机科学与探索, 2010, 4(4): 289-303.