基于隐马尔可夫模型的 3D 手写识别方法*

王琦进[1]，齐晓霞[1]，江义渊[2]，徐旺兴[2]
(1. 安徽新华学院 信息工程学院，合肥 230088；2. 台湾万能科技大学 资讯工程系，台湾 中壢 32061)

摘　要：为实现便捷高效的人机交互，提高交互能力，在基于隐马尔可夫模型(HMM)的基础上，提出了一种新的 3D 手写识别方法。该方法使用带有三轴加速度传感器的手持设备去采集各种手写数据；使用插值及快速傅里叶变换(FFT)滤波等方法对采集的数据进行预处理；使用隐马尔可夫模型对每个手写动作进行模型训练；使用训练过的手写模型对采集的数据进行手写识别。数据测试结果表明，该方法在手持移动设备上数据分类的准确性可达到 84.5%。

关键词：手写识别；加速度传感器；隐马尔可夫模型；快速傅里叶变换
中图分类号：TP391.4　　**文献标志码**：A　　**文章编号**：1001-3695(2012)09-3565-03
doi:10.3969/j.issn.1001-3695.2012.09.097

3D handwriting recognition method based on hidden Markov models

WANG Qi-jin[1], QI Xiao-xia[1], CHIANG Yi-yuan[2], HSU Wang-hsin[2]
(1. *Collage of Information Engineering*, *Anhui Xinhua University*, *Hefei* 230088, *China*; 2. *Dept. of Computer Science & Information Engineering*, *Vanung University*, *Chungli Taiwan* 32061, *China*)

Abstract: In order to achieve efficient and convenient human-computer interaction, improve interaction ability, this paper proposed a new 3D handwriting recognition method based on HMM. In this method, using of the tri-axis accelerometer mounted on a handheld device to collect various handwriting data; using of the interpolation and FFT filtering to preprocess the data; using of the hidden Markov models to train every handwriting movement ; using of the trained HMM to carry out the handwriting recognition. Tested on the handwriting mobile device, the results show that the classification accuracy can reach 84.5%.

Key words: handwriting recognition; accelerometers; HMM(hidden Markov models); FFT(fast Fourier transtorm)

手势识别是多通道人机交互领域的一个重要研究方向，其本质是一种模式识别，属于多维模式识别和智能计算机接口的范畴。近年来，随着移动传感设备的发展，手持移动设备也变得越来越普及。人们希望借用其便携性，随时随地地做出一些手势动作，实现人机交互。但由于手持移动设备在进行人机交互过程中出现了交互效率低、自然性不好、用户认知负担过重等问题，因此，如何研究出一种新兴的交互方法已成为人机交互领域的研究热点之一。

目前针对手势识别系统的研究主要有两类：a)基于视频(图形处理)的，b)基于传感器(加速度信号)的[1]。由于传感器技术的快速发展，一种能够识别用户手势动作的交互技术已经出现。其实，早在 20 世纪 90 年代初，有关科研人员就开始了基于传感器的手势交互研究。IBM 公司在 1996 年推出了一种具有自动感知运动状态的运动感知设备，用于玩具和日常工具中[2]；1998 年 Harrison 等人对使用传感器作为用户操作接口的可行性进行了探讨；2000 年 Bartlett 等人在 PDA 上使用旋转、倾斜等手写实现滚屏、选择等功能[3]；2001 年 Rekimoto[4] 将传感器移植到可穿戴的输入设备上，完成各种交互；2004 年 Ikjin 和 Wonbae 研究了手持设备中的加速度信号的处理方法[5]；2005 年 Kela 等人[6] 定义了手写接口的基本概念，设计了手写识别系统 Smart Design Studio；2006 年 Bake 等人[7] 探讨了如何识别用户连续的动作和手势；2007 年 Ferscha 等人[8] 将手势的概念泛化，根据手势的语法规则分为原子手势和复杂手势两种，并建立了手势库 Glib，可供基于加速度传感器的系统使用。

对于如何使用加速度信号进行手势识别，有些文献也做了介绍，但它们的大多数工作集中在识别简单或者相同的手势，如简单的线性移动和变向以及阿拉伯数字等。Hsu 等人[9] 尝试了 LCS-SVM 的方法对连续字母即英文单词进行识别，取得了较好的效果，但该方法对单个字母的识别率相对较差。本文在上述研究的基础上，在三维空间中针对 26 个英文字母提出了一种基于隐马尔可夫模型的 3D 手写识别方法。隐马尔可夫模型是语音识别领域中最成功的统计模型之一，但其在 3D 加速度传感领域中的手势识别还是一个崭新的研究方向。

1　隐马尔可夫模型及其学习算法

1.1　隐马尔可夫模型

隐马尔可夫模型(HMM)是一种具有学习能力的统计模型，该模型对时变、非平稳时间序列具有较好的分析效果，已被

收稿日期：2011-12-15；**修回日期**：2012-02-01　　**基金项目**：安徽省高等学校省级优秀青年人才基金资助项目(2012SQRL228)；安徽高校省级自然科学研究项目(KJ2012Z134)

作者简介：王琦进(1980-)，男，安徽人，讲师，硕士，主要研究方向为模式识别与应用、Ad hoc 无线网络(qjwang118@yahoo.com.cn)；齐晓霞(1980-)，女，安徽人，讲师，硕士，主要研究方向为人工智能与模式识别；江义渊(1969-)，男，台湾人，助理教授，博士(后)，主要研究方向为模式识别与智能系统；徐旺兴(1970-)，男，台湾人，助理教授，博士，主要研究方向为模式识别与智能系统、计算机网络与通信.

成功用于语音信号识别、故障检测等领域[10]。

$$\text{HMM 可以记为 } \lambda = (N, M, A, B, \pi)$$
$$\text{简写为 } \lambda = (A, B, \pi) \tag{1}$$

其中：N 为隐藏状态数；M 为每个状态下的观察序列长度；A 为状态转移概率矩阵，记为 $A = \{a_{ij}\}$，$a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$，表示从状态 s_i 转移到状态 s_j 的概率；B 为观察值概率矩阵，记为 $B = b_j(k) = P(v_k \ at \ t \mid q_t = s_j)$，表示在状态 s_i 时输出符号 v_k 的概率；π 为初始概率分布矢量，记为 $\pi = \{\pi_i\} = P(i_1 = i)$，表示状态 i 在初始时间 $t = 1$ 时的概率。

从上述隐马尔可夫模型的定义以及给定的模型参数可以看出，隐马尔可夫模型是一个数学上的双重随机过程，可以应用于时变数据的建模工作。

1.2 Baum-Welch 算法

HMM 模型的训练采用 Baum-Welch 算法，它是一种基于期望最大化的算法，根据最大似然准则（maximum likelihood，ML）对模型参数进行迭代的重新估算，直至得到最大似然意义下的最优模型参数值[11]。它通过估计模型参数 A、B 和 π，使模型生成观察序列的概率 $P(O \mid \lambda)$ 最大化，以增强模型的预测性能。

算法描述为：定义变量 $\xi_t(i,j)$，$\xi_t(i,j)$ 表示在给定模型 λ 和观察序列 O 的情况下，模型 t 时刻处于状态 i，$t+1$ 时刻处于状态 j 的概率，即 $\xi_t(i,j) = P(q_t = i, q_{t+1} = j \mid O, \lambda)$。经过推导 $\xi_t(i,j)$ 可以表示为 $\alpha_t(i)$ 和 $\beta_t(j)$ 的函数，即

$$\xi_t(i,j) = P(q_t = i, q_{t+1} = j \mid O, \lambda) = \frac{P(q_t = i, q_{t+1} = j, O \mid \lambda)}{P(O \mid \lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \tag{2}$$

其中：$\alpha_t(i)$ 和 $\beta_{t+1}(j)$ 分别为前向变量和后向变量。

若 t 时刻状态为 S_i 的概率为

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{3}$$

则 Baum-Welch 算法可表示为

$$\tilde{\pi}_i = \frac{\alpha_1(i)\beta_1(i)}{\sum_{j=1}^{N} \alpha_1(j)\beta_1(j)}$$
$$\tilde{\alpha}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$
$$\tilde{\beta}_{jk} = \frac{\sum_{\substack{t=1 \\ O_K = V_K}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} = \frac{\sum_{t=1}^{T} \alpha_t(j)\beta_t(j)\delta(O_t, V_K)}{\sum_{t=1}^{T} \alpha_t(j)\beta_t(j)} \tag{4}$$

1.3 前向算法

前向算法可看成求一个隐马尔可夫模型与一个观察值序列的匹配程度[12]。该算法从若干模型中选择最优的，进而得到最匹配的序列模型。

算法描述为：定义前向变量 $\alpha_t(i) = P(O_1, O_2, \cdots, O_t, q_t = s_i \mid \lambda)$，表示模型 λ 下、在 t 时刻，观察事件为 O_t、状态为 s_i 的概率，则在 $t+1$ 时刻：

$$\alpha_{t+1}(j) = P(O_1, O_2, \cdots, O_{t+1}, q_{t+1} = s_j \mid \lambda) = \left[\sum_{i=1}^{N} a_i(i) a_{ij}\right] b_j(O_{t+1}) \tag{5}$$

则前向算法为

初始：
$$a_1(i) = \pi_i b_i(O_1) \qquad 1 \leqslant i \leqslant N \tag{6}$$

递归：
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}) \qquad 1 \leqslant t \leqslant N-1, 1 \leqslant j \leqslant N \tag{7}$$

终止：
$$P(O \mid \lambda) = \sum_{i=1}^{n} \alpha_T(i) \tag{8}$$

2 拟建的 3D 手写加速度识别系统

拟建的 3D 手写加速度识别系统的流程如图 1 所示，它分为数据训练（建模）流程和数据测试（识别）流程，主要由数据采集、数据预处理、数据训练以及模式识别四部分构成。

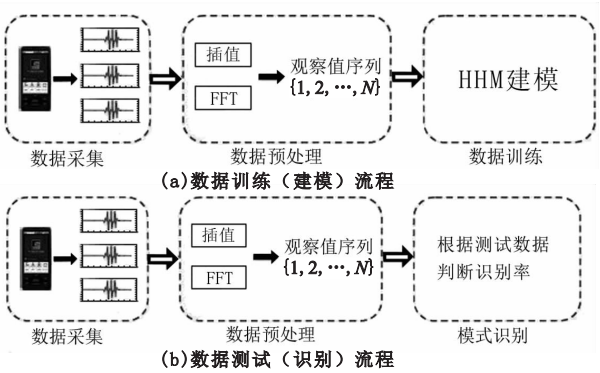


图1 3D手写加速度识别系统的流程

2.1 数据采集

实验使用 HTC phone、iphone 等带有内置加速度传感器的手持设备进行数据采集，每隔约 18 ms 采集一组数据，每个手写字母的数据采集组数在 100 ~ 300 不等。传感单元以一定的频率获取运动过程中 x、y、z 三轴加速度值 a_x、a_y、a_z，并构成每一时刻的运动加速度 O_i，记为 $O_i = (a_x, a_y, a_z)$。数据采集过程中，由于个体差异，不同人执行同一个手势动作，其动作幅度和执行速度有一定的差别，即使是同一个人多次执行同一动作也会有差异，这样就造成了采样值的不确定性。本文对手势运动的速度和幅度作了规定，同时要求同一人对相同采样对象进行数据采集，从而保证手势动作执行的规范性。实验对 26 个大写字母“A ~ Z”的手写数据进行了采集，每个字母采样 50 次。

2.2 数据预处理

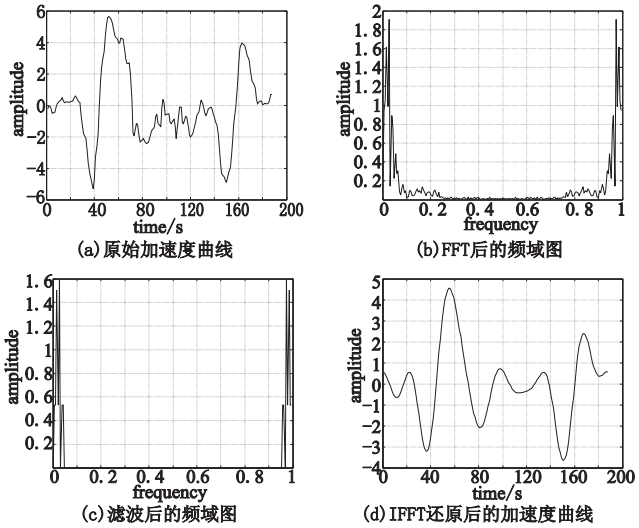
2.2.1 插值法保持采样数据长度一致

在手写运动过程中，加速度值在 x、y、z 三轴上具有沿时间连续变化的属性。如果某次采样速度偏离正常值，便会造成采样时间的快慢不同，导致数据长度的不同，从而使识别成功率降低。同时由于 MATLAB 的 HMM toolbox 要求实验数据在长度上保持一致，因此使用插值法来解决。为了不丢失采样数据所携带的信息，实验没有选用求平均数据长度的方法，而是选择一组采集数据中长度最长的一个，其余数据通过插值算法与最长数据保持一致。

2.2.2 FFT 滤波去杂讯

快速傅里叶变换（FFT）是离散傅里叶变换的快速算法，它在数字信号处理、计算大整数乘法、求解偏微分方程等方面有着广泛地应用。在利用 FFT 进行滤波时，通常使用截止频，通过设置一个高低通频率界限，从而起到滤波的效果。本实验使用手持加速度设备进行数据采集，采集的数据即使在外界无动作的情况下依然会产生很小的杂讯噪声，并且手写运动过程也会不可避免地产生轻微抖动，产生杂讯。实验需要去掉的是不

足以表征手写信息的杂讯噪声,通过将频率域中的某些频率成分幅值进行设置,然后运用 IFFT 变换到时间域从而达到滤波的效果。图 2 所示为字母“F”在 x 轴上使用 FFT 及 IFFT 做滤波的过程。



(a) 原始加速度曲线　(b) FFT后的频域图　(c) 滤波后的频域图　(d) IFFT还原后的加速度曲线

图2　FFT及IFFT滤波过程

如图 2(a)所示,由于采样过程中的轻微抖动,使得原始的未经处理的数据的频率带有杂讯波,曲线呈锯齿状;如图 2(b)所示,在经 FFT 转换到频域图后,仍可以看到存在低能量的杂讯波。实验经过多次测试,当频率成分幅值为 50 时进行杂讯噪声的去除,效果最好(如图 2(c)所示)。图 2(d)与(a)相比,经过 IFFT 转换回时间域的频率曲线变的平滑,达到了滤波去杂讯的效果。实验表明,进行 FFT 后的手写识别率比未经 FFT 时要高 9.3%。

2.2.3　状态数与观察值序列

根据隐马尔可夫模型的定义,在解决模式识别的问题中,应首先确定模型的隐藏状态数 N。理论上讲,状态数越多越好,这是因为随着状态数的增加,识别的错误率会降低到一个很稳定的程度。然而由于训练样本是有限的,所以状态数不能太大。如果实验中状态数 N 选取过大,则训练后很多状态在参数模型的对应项中将变为 0 或非常接近于 0,成为冗余项,通常实验中状态数按其复杂程度固定为 2 ~ 8 不等的数目。

每个状态下不同的观察值个数 O,观察值与要构建系统的输出息息相关,这里的观察值就是手势运动过程中的加速度向量,标记为 $O = \{a_1, a_2, \cdots, a_n\}$。图 3、4 为一组字母“A ~ Z”在手写过程中加速度观察值的量化过程。从图 3 的加速度分布可以看出,图形的波峰与波谷处是手写运动过程中特征信息点的位置,这些点的变化转移也是手势运动方向的变化转移过程,通常伴随着大加速度值的产生。在图 4 的加速度直方图中,加速度值“0”附近,堆集了大量的手写信息点,而这些点通常不带有过多的特征信息,它只是手写过程中一个相对平稳的过程,较少出现运动方向的变化转移。

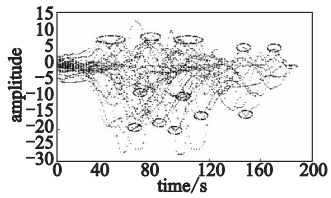


图3　观察值分布图

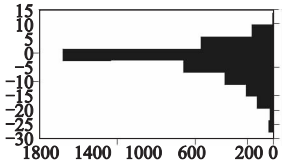


图4　观察值直方图

2.3　数据训练

使用观察值序列调整 HMM 参数称为一个训练过程。训练问题是一个非常重要的问题,因为它可以使模型参数最为理想地适应所观察到的训练数据。数据在经过预处理并得到观察值序列后,由 Baum-Welch 算法训练出离散的 HMM 数据模型。

实验在每个手写字母的 50 笔采样数据中随机抽取 40 笔进行训练。由于训练对 x、y、z 三轴加速度数据单独进行,因此每个手写字母的数据模型均由 x、y、z 三轴单独模型所组成。模型表示为

$$\lambda_{s,t}, s \in \{A, B, C, \cdots, Z\}, t \in \{x, y, z\} \tag{9}$$

其中:s 表示 A ~ Z 的 26 个大写字母,t 表示 x、y、z 三轴。

2.4　模式识别

在完成数据训练并得到相应的 HMM 模型后,利用每个字母的剩余 10 组数据进行测试。本文把测试数据分别丢入每个模型,通过得到的对数似然函数(log-likelihood, LL)值来判别测试数据与统计模型的成功识别率。由于 x、y、z 三轴加速度模型为单独训练,因此每个字母的对数似然函数值定义为

$$LL(s) \overset{\text{def}}{=} LL(s,x) \times LL(s,y) \times LL(s,z), s \in \{A, B, C, \cdots, Z\} \tag{10}$$

实验依据 max(LL)对采样数据进行分类,如 $LL(s =$ ‘M’),而 max($LL(s)$) = LL(M),那么分类结果与数据出处相吻合,识别成功。

3　实验结果

为了评价本文的手写识别模型,对 26 个大写字母,共 260 笔数据进行了测试。实验选用的状态数为 4,观察值个数为 8。表 1 为字母识别成功率结果(由于篇幅限制,选择部分字母)。

表 1　字母测试的成功识别率(部分)　/%

字母	A	B	C	D	E	F	G	H	I
识别率	50	80	100	100	40	80	100	90	80
字母	J	K	L	M	N	…	X	Y	Z
识别率	80	90	20	100	90	…	100	90	90

从表 1 的字母识别率可以看出,除个别字母的识别率较低外,其他字母识别率均达 80% 以上。其原因为:“L”(识别率为 20%)“U”(识别率为 30%)等个别字母识别率过低,从而影响到总体识别率。个别字母由于在手写过程中动作幅度变化不明显,没能产生足够的特征信息,因此识别率过低。实验对所有测试数据进行逐一比对,总体识别成功率达 84.5%。

4　结束语

本文提出了一种基于隐马尔可夫模型的 3D 手写加速度识别系统。它以手势作为人机交互对象,根据手和胳膊的运动轨迹(三维)找出其中的含义,从而实现了一种高效、自然的人机交互手段。系统在实现过程中,首先使用带有加速度传感器的手持设备进行数据采集;在获取了加速度值之后,再对数据进行预处理;然后从同一动作的诸多加速度序列中选择若干序列进行模型的构建;最后使用剩余序列对该模型进行测试,从而达到手写识别的目的。实验结果表明,本文方法是有效的,其分类精度可达 84.5%。

在下一步工作中,将考虑使用精度更高的加速度传感设备进行数据采集,以获得更加准确的 3D 手势运动信息,从而进一步改善手写识别的效果。此外,使用连续型隐马尔可夫模型以及隐马尔可夫模型与高斯混合模型相结合 (下转第 3570 页)

(上接第 3567 页)的方法也是笔者进一步的研究方向。

参考文献:

[1] 徐光佑,陶霖密,史元春,等.普适计算模式下的人机交互[J].计算机学报,2007,30(7):1041-1053.

[2] BARTLETT J F. Rock 'n' scroll is here to stay[J]. IEEE Computer Graphics and Applications,2000,20(3):40-45.

[3] HINCKLEY K,PIERCE J,SINCLAIR M, *et al*. Sensing techniques for moblie interaction[C]//Proc of the 13th Annual ACM Symposium on User Interface Software and Technology. New York: ACM,2000: 91-100.

[4] REKIMOTO J. GestureWrist and GesturePad: unobtrusive wearable interaction devices[C]//Proc of the 5th IEEE International Symposium on Wearable Computers. Washington DC: IEEE Computer Society,2001:21-27.

[5] JANG I J,PARK W B. Signal processing of the accelerometer for gesture awareness on handheld devices[C]//Proc of the 12th IEEE International Workshop on Robot and Human Interactive Communication. 2003:139-144.

[6] KELA J,KORPIPAA P,MANTYJARVI J, *et al*. Accelerometer-based gesture control for a design environment[J]. Personal and Ubiquitous Computing,2005,10(5):285-299.

[7] BAEK J,YUN B J. A sequence-action recognition applying state machine for user interface[J]. IEEE Trans on Consumer Electronics,2008,54(2):715-726.

[8] FERSCHA A,RESMRITA S. Gestural interaction in the pervasive computing landscape[J]. Elektrotechnik & Informationstechnik, 2007,124(1):17-25.

[9] HSU Wang-hsin,CHIANG Yi-yuan,WU Jung-shyr. Integrating LCS and SVM for 3D handwriting recognition on handheld devices using accelerometers[J]. WSEAS Trans on Computers,2010,9(2):235-251.

[10] RABINER L R,JUANG B R,JUANG B H. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceeding of the IEEE,1989,77(2):257-286.

[11] 王西颖,戴国忠,张习文,等.基于 HMM-FNN 模型的复杂动态手势识别[J].软件学报,2008,19(9):2302-2312.

[12] 王悦,岳玮宁,董士海,等.手持移动计算中的多通道交互[J].软件学报,2005,16(1):29-35.