

一种基于加速计的手势识别算法 及其在虚拟环境中的应用*

刘剑锋¹, 张大兴²

(1. 浙江大学 CAD & CG 国家重点实验室, 浙江 杭州 310058; 2. 杭州电子科技大学 计算机学院, 浙江 杭州 310018)

摘要:针对虚拟环境下的三维交互问题,提出了基于加速计的手势识别算法。识别之前先对加速计输出做量化处理,再用离散型隐马尔科夫模型(HMM)进行建模、训练与识别。基于此识别算法,将手势作为一种交互方式应用在三维交互系统中,建立了手势-语义映射表,以训练标准手势,最后对输入手势进行了识别,根据识别结果完成了交互语义。将算法应用于虚拟奥运博物馆实例中,实验结果表明,该算法可有效快速识别出用户输入手势,且平均识别率达到 90% 以上,用户凭借此方法能准确无误地完成交互。

关键词:三轴加速计;隐马尔科夫模型;手势识别;三维交互;虚拟现实

中图分类号:TH164;TP301

文献标识码:A

文章编号:1001-4551(2010)08-0091-06

An accelerometer-based gesture recognition algorithm and its application in virtual environment

LIU Jian-feng¹, ZHANG Da-xing²

(1. State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310058, China;

2. College of Computer Science, Hangzhou Dianzi University, Hangzhou 310058, China)

Abstract: Aiming at the 3D interaction in virtual environment, an accelerometer-based gesture recognition algorithm was proposed. As a pre-process procedure, raw data output by accelerometer was quantized, and then the discrete hidden Markov model(HMM) was used to train and recognize them. Based upon this recognition algorithm, gestures was treated as a method of human-computer interaction and used in 3D interaction subsystem in VR system. The gesture-semantic map was established, the standard gestures was trained, finally the recognition was done. Experimental results show that the system can recognize input gestures quickly with a reliable recogniton rate. The users are able to perform most of the typical interaction tasks in virtual environment by this accelerometer-based device.

Key words: tri-axes accelerometer; hidden Markov model(HMM); gesture recognition; 3-D interaction; virtual reality

0 引 言

虚拟现实系统在 CAD、E-learning、体育仿真、数字娱乐等领域有着广泛的应用。例如浙江大学 VRMM 研制的 VNM(Virtual Network Marathon),虚拟生物实验室、国家体育总局体育科学研究所研制的虚拟自行车训练系统。虚拟现实系统的主要任务包括两方面:

①有效地组织与管理虚拟场景,并将虚拟物体尽可能清晰准确地绘制出来;②准确无误地完成用户与虚拟

物体之间的交互任务^[1]。

得益于真实感图形学在近几十年的迅速发展,目前的三维虚拟现实系统已经有能力绘制出逼真的虚拟环境,并让用户获得了较强的沉浸感。对于虚拟现实系统的另一主要任务即三维人机交互(3-d interaction),D. Bowman^[2]等人把这个概念定义为:在空间环境(spatial context)下完成的用户交互行为称为三维交互。虚拟环境中的三维交互方式较之传统二维基于 WIMP 隐喻式(metaphor)交互方式存在着交互内容的

复杂性、交互通道(modal)的多维性、交互类型的多样性的特点。这也是传统二维交互方式无法满足于虚拟现实系统的原因,因此需要提供一系列丰富有效的三维交互手段来满足虚拟环境中的交互需求。

基于加速计的手势识别技术运用于人机交互已经在许多文献中被提出。Juha Kela^[3]等人提出将手势识别技术作为一种辅助性的交互手段,运用于指定的实际环境中,并开发出系统 Smart Design Studio 用于支持这一交互方式。其系统测试情况表明,对不同用户开发不同手势命令有助于提高系统的识别率。Portillo-Rodriguez^[4]等人提出一种基于 FSM 的手势匹配方法,将手势根据其在特定构型空间中的位置分解成一组状态序列,以这组序列在 FSM 中能否达到终点状态来判断是否可识别。Hofmann, F.^[5]提出了一种基于手势运动加速度向量的手势识别方法。在此基础之上,Thomas Schlomer^[6]等人借助 Wiimote(一种集成了加速计的游戏控制器)实现了基于离散型 HMM 的识别算法。

本研究借鉴基于加速计的手势识别算法,并将其应用在三维交互系统中。

1 基于加速计的手势识别算法

集成了 3 轴加速计的传感器作为手握式(hand held)三维交互输入设备来获取人体手部运动时产生的加速度信息。人体每完成一个手势,可获得一串输入信号,将这组信号流称为一个输入模式(input pattern)。

定义 1 模式 $P = \{V_t | 0 \leq t \leq T\}$, V_t 为 t 时刻获得的三轴加速度数向量, T 为一个信号流终止时刻。

根据日常经验可知,人体在执行(perform)相同手势时,手部运动每次产生的模式在某种程度上满足一定的统计规律。标准模式(standard pattern)是预先定义好的一类模式,每个标准模式对应一个专门的输入语义(input semantic),当交互处理系统接收到输入语义信息时做出必要的反馈行为,此过程即达到了基于手势交互的目的。基于加速计的三维交互过程,即可视为“从获得输入模式,通过模式识别得到与其最接近的标准模式,系统再对输入的标准模式进行反馈”的过程。

1.1 模式建立与预处理

本研究采用的 BOSCH SMB380 加速计芯片的采样频率和灵敏度偏高,产生数值特点是近似连续的且噪声与冗余较大。在使用数据之前需作去噪与量化处理。

去噪过程即对输入模式 P 的每个加速度分量 V_t 做平滑(smoothing)处理。平滑函数如下:

$$S(V_t) = \begin{cases} V_t & t=0 \\ S(V_{t-1}) + \alpha(V_t - S(V_{t-1})) & 0 < t \leq T \end{cases} \quad (1)$$

其中,平滑因子 α 取值为 $0 \sim 1$, α 值越大平滑效果越明显。

量化处理要求采用一种能有效减少数据复杂性且编码本(code book)固定的方法。量化对象 V 为三维空间向量,且对象之间的相关性以欧几里得距离来衡量。因此空间划分(partitioning)聚类算法符合本研究要求。常用的空间划分聚类算法有 k-means 聚类和 k-medoids 聚类。k-medoids 聚类算法与 k-means 聚类算法相比,对噪声和异常数据有较强的鲁棒性^[7],因此本研究选择后者作为聚类算法。

初始聚类中心(initial medoid)预设为均匀分布在半径为 R_{cluster} 的且中心在原点的球体表面上的 M 个点。

聚类算法如下:

- (1) 设定 R_{cluster} 与 M 的初始值,计算得到初始聚类中心点集合 $M_{\text{ini}} = \{m_1, m_2, \dots, m_M\}$;
- (2) 将所有对象 V 分到各个类中(根据与 medoid 最相近的原则);
- (3) 对于每个类(C_i),顺序选取其中的一个对象 V_x ,计算用 V_x 代替 m_i 后的消耗 $-E(V_x)$ 。选择 E 最小的 V_x 来代替 m_i 。
- (4) 重复步骤(2)~(3),直到 medoid 确定。

理想的聚类结果是任意一个对象 V 与其所在同一聚类中的其他对象之间的距离尽可能小,与聚类之外的任意对象之间的距离尽可能大。

根据聚类结果,对数据进行量化。量化之前需确定编码本,编码本的大小即上一步骤得到的聚类数 M 。编码内容置为与之对应的聚类中心的 ID。量化对象 V_t 的量化结果即是 V_t 所在聚类的聚类 ID。

综上所述,输入模式 P 的量化结果为 $P' = \{ID_{V_t} | V_t \in P\}$ 。量化结果有如下特点:①数据离散化;②数据复杂性小,每个分量的值只可能为 $1 \sim M$ 之间的整数。可知,上述平滑和量化两个预处理步骤有效解决了初始数据的噪声和复杂度大的问题。

1.2 模式分类

常用的模式分类方法有支持向量机(Support Vector Machine)方法、神经网络(Neural Networks)方法、隐马尔可夫模型(Hidden Markov Models)方法。本研究

采用离散型一维隐马尔可夫模型。

1.2.1 训练

定义2 手势(Gesture)为一个二元组 $G = \langle P, Att \rangle$ 。 P 为一个输入模式经量化的结果, Att 为设备的一组附加输入信息,如本研究所采用手握式传感器的其他按钮信息。

本研究用一组手势来代表一组特定的交互语义。

对于交互系统中的每一个标准手势,本研究通过人体重复执行这一手势来获得一组样本 $T = \langle G_1, G_2, \dots, G_n \rangle$,用这组样本来训练一个 HMM。

训练步骤如下:

(1) 初始化一 HMM,使用多维观测值集合 $\{O_1, O_2, \dots, O_L\}$ 作为训练数据;

(2) 考虑到多维观测值,且各个观测值序列为独立事件,可记 $P(O|\lambda) = \prod_{i=1}^L P(O_i|\lambda)$,其中 L 为该组训练样本的数据的个数。通过 Baum-Welch 算法^[8] 对该 HMM 的参数 $\{A, B, \pi\}$ 进行重新估计,得到新的 HMM λ' ,并计算 $P(O|\lambda')$ 。

(3) 若 $P(O|\lambda') - P(O|\lambda) > \varepsilon, \lambda = \lambda'$; 重复步骤(2);若 $P(O|\lambda') - P(O|\lambda) \leq \varepsilon$,则视为算法收敛,所得 λ 即为局部最优解。

训练的结果是,得到一组与表1中 Gesture a-h 对应的 HMM。

1.2.2 识别

识别过程即获取一个随机输入手势并从标准手势库中找到最优匹配手势的过程。识别过程的形式化描述如下:

(1) 随机输入手势 $G_x \langle P, Att \rangle$,取 P 分量作为观测序列 O 。

(2) 求解 $\sigma_n = \operatorname{argmax} P(O|\lambda_i), \lambda_i$ 为经训练的代表每一标准手势的 HMM。

(3) 所得最大值 σ_n 对应的 HMM 所代表的标准动作 G ,即为 O 的识别结果。

1.2.3 阈值型 HMM 方法

上述识别算法对任意一个输入手势皆有解。该方法在用户误输入时将产生严重错误(系统产生错误的识别结果,且不可挽回)。考虑采用阈值 HMM(threshold HMM)方法^[9],在步骤(2)中,考虑 σ_n 小于某一特定阈值 ε 时,可认为该输入为无效输入,以此降低错误风险。

2 虚拟环境中三维交互系统的设计

虚拟环境下三维人机交互通常有3个主要任务:

(1) 虚拟物体操控(Virtual object manipulation):指对交互对象—三维虚拟物体的选择、缩放、旋转、移动、创建、删除、编辑等操作。

(2) 视点控制(Viewpoint manipulation):视点(point-of-view)也称虚拟眼或摄像机,包含 Zoom Factor 和 Field of View 等参数。视点控制即通过移动视点在三维世界坐标系下的位置来完成虚拟环境中的漫游(roam)或导航(navigation)。

(3) 应用程序控制(Application control):与三维虚拟环境无关的对应用程序的一般控制任务皆可纳入 Application control 范畴。

下文所描述的虚拟环境中三维交互模型的目标即是准确有效地完成上述3个任务。

在描述交互模型之前先定义如下概念:

定义3 交互原语(interaction atom)记为 A ,指完成某种特定功能的且不可再分割的交互操作。通常是机器指令级的实现。设计交互系统时需预定义交互原语集 A^* 。

定义4 执行例程 $E = \{A_0, A_1, \dots, A_i\}$,其中 A_i 表示某一交互原语,执行例程 E 为一具有时序性的交互原语序列。执行例程用于完成某一项特定的交互任务。

定义5 交互语义 $s = \langle I, C, E, con \rangle$,交互语义 S 为一个四元组,元素 I 表示输入数据流, C 表示交互行为触发时刻的上下文快照, E 表示该语义对应的执行例程, con 为例程 E 激活的条件。

2.1 交互模型体系结构

下面提出一种适于虚拟现实系统的三维交互模型:

交互模型构成组件如下:

- (1) 输入/输出流(I/O stream);
- (2) 输入通道(input channel);
- (3) 解析器(interpreter);
- (4) 执行器(executer);
- (5) 反馈通道(feedback channel)。

另外,交互系统还包括3个外部对象:

- (1) 空间交互上下文(spatial context);
- (2) 媒体库(media);
- (3) 交互语义库(Semantic database)。

交互模型按功能不同可分为3层:

(1) 交互解释层:从输入流中提取交互任务信息,结合上下文,分析得到交互语义。

(2) 交互执行层:根据交互语义,对交互对象进行某种原语(atom)级操作,交互对象一般存在于媒体库中。

(3) 交互反馈层: 将执行层的返回结果, 翻译成用户能够理解的信息, 并通过输出通道反馈给用户。

该分层交互模型的体系结构如图 1 所示。

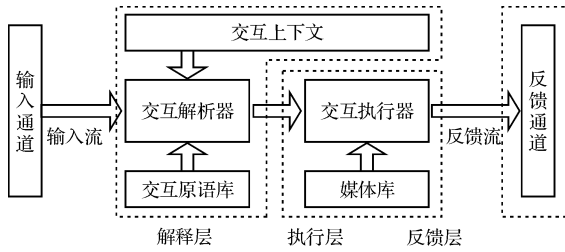


图 1 三维交互模型体系结构图

2.2 建立交互语义库

交互语义指一组输入数据流在特定交互上下文下所表示的“含义”, 在上述交互系统中“含义”可解释为: 一条或多条交互命令及其命令参数(通常包括交互对象)组成的集合。交互语义由解析器通过分析输入流, 并参考上下文快照得出。执行器接收解析器输出的交互语义序列, 逐条译成交互原语, 最终执行所有交互原语。

在设计交互系统时, 需枚举系统提供的交互功能, 并且以交互语义的形式存储在交互语义库中。交互语义库必须提供添加/删除、检索的功能。

2.3 交互结果反馈

交互结果反馈是交互过程中的重要环节, 良好的交互系统要求反馈信息具备明确性与完整性。明确性是指反馈信息必须包含对交互结果的明确指示; 完整性要求反馈信息包括交互任务的状态、交互任务引起的上下文变化情况、交互任务失败时的原因分析以及替代方案的建议。

结果反馈过程分以下步骤:

- (1) 执行层提交执行例程的返回值至反馈层;
- (2) 反馈层分析返回值与执行例程的操作代码, 组织反馈信息;

(3) 反馈信息包含多通道内容(如显示设备通道、音频设备通道、机械型设备通道), 反馈层创建各个通道的反馈信息, 最终通过各个反馈通道输出反馈结果。

3 应用实例与分析

此部分描述利用上述基于加速计的手势识别技术的三维交互系统, 在应用级系统 - 虚拟奥运博物馆(VDOM)^[10] 漫游系统中的实现。

3.1 建立手势 - 语义映射表

VDOM 中的三维交互功能归纳如下:

(1) Agent 控制: turn left/right, move forward/back;

(2) Object 控制: select, rotate, scaling。

在本研究中, 对上述交互功能, 预定义交互语义库。在本系统中, 以 XML 文件的形式存储交互语义, 文件名为 *.sdb.xml。

在为交互语义定义标准手势时, 需从手势的形象性与简洁性出发, 目的要使用户能够快速、自然地接受手势命令集。VDOM 系统定制的手势 - 语义映射表(部分)如表 1 所示。手势 a ~ h 分别对应如图 2 所示的手势运动 a ~ h。

表 1 手势 - 语义映射表

Gesture	Interaction Semantic
a	move left/turn left
b	move right/turn right
c	rotate left
d	rotate right
e	move up
f	move down
g	confirm/select/zoom in
h	cancel/delete/zoom out

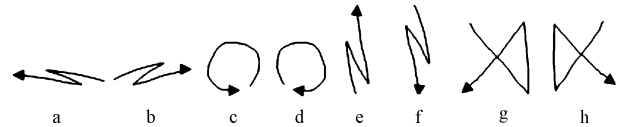


图 2 标准手势图

3.2 训练标准手势

各组训练样本的情况如表 2 所示。考虑到手势 g (confirm)、h (cancel) 较为复杂, 可适当增加样本集数目以提高训练效果。利用 1.2.1 中提出的训练算法, 经训练得到标准动作 a ~ h 对应的 8 个 HMM。

表 2 各手势训练数据及识别结果

Interaction Commands	Training	Testing	Recognition	
	Data	Data	Correct	Ratio
move/turn left	150	50	47	0.94
move/turn right	150	50	46	0.92
rotate left	150	50	49	0.98
rotate right	150	50	47	0.94
move up	150	50	48	0.96
move down	150	50	49	0.99
confirm	200	50	44	0.88
cancel	200	50	45	0.90

在训练过程中, 分别考虑 HMM 状态数的选择对训练效果的影响情况(如图 3 所示)。训练效果可以通过以下方式衡量: $P(G|\lambda)$ 表示 HMM λ 中产生观测序列 G 的概率。 $P(G|\lambda)$ 越高意味着 HMM λ 与目标手势 G 更贴切。

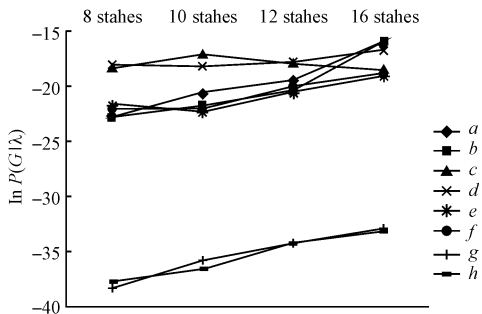


图 3 HMM 状态数对训练效果的影响

将 HMM 状态数由 8 增加到 12 的过程中各个手势的训练效果均有提高。将 HMM 状态数由 12 增加至 16 时,手势 *a, b, d* 的训练效果增加明显,但手势 *c, f, e* 的训练效果呈下降趋势,结合训练以及识别速度等方面的考虑,本系统使用 12 个状态的 HMM。

3.3 设计基于手势的三维交互子系统

三维交互子系统从加速计传感器获取加速度原始数据,经滤波和量化两道预处理程序后,得到平滑且离散化的数据流,分析器从数据流中分析而得手势数据。手势数据可以通过训练器 (Trainer) 来对现有 HMM 进行训练,也可提交给识别器 (Recognizer) 进行识别。手势识别模块的输出结果将提交给交互任务执行器,并由其来完成 VDOM 中某项具体的三维交互任务。

VDOM 中基于手势的三维交互子系统的处理流程如图 4 所示。

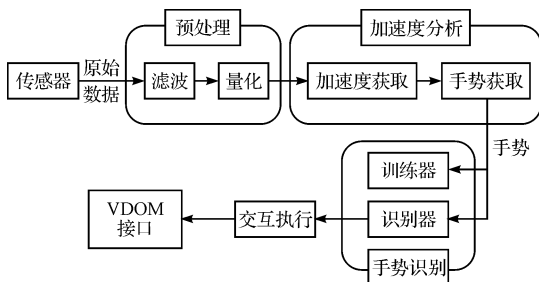


图 4 VDOM 三维交互子系统的处理流程

3.4 实验结果与性能分析

在本研究中,完成训练工作之后,对系统进行测试。每个手势分别使用 50 个输入样本来测试。分析测试结果可得,系统对手势 *g* 与 *h* 的识别率平均在 89% 左右,略显不足,其余手势均在 95% 及其以上,效果较理想。

分析图 4 中 HMM 状态数对手势 *g, h* 的影响,可作出如下推断:①增加 HMM 状态数可以在一定程度上增加 *g, h* 对应 HMM 的训练效果,从而增加其 HMM 对 *g, h* 手势的识别率;②适当增加训练样本数,也可提

高 HMM 对 *g, h* 手势的识别率。结合上述两个方法能够改善三维交互系统对手势 *g, h* 的识别率。

总体上,本研究提出的基于加速计的三维交互技术在虚拟现实系统实例 - VDOM 中成功地实现,且取得了较为可靠的测试结果。用户在绝大多数情况能够借助加速计准确无误地完成交互。系统运行时截图如图 5 所示。图 5 中左列子图分别是用户输入的手势,根据表 1 可知其代表含义分别为虚拟人向右转、虚拟人向左转、观察物体向左旋转、观察物体向右旋转。各个手势的交互结果分别体现在图 5 的右列子图中。



图 5 运行结果

4 结束语

在基于手势的三维交互系统中,造成手势识别率不高的原因有:①训练度的不足导致 HMM 与标准手势之间关联性不强;②对于复杂的手势,在量化过程中损失信息量较大;③样本集不够完整,仅仅来源于几位用户。在实际应用中,可分别通过如下方法对识别率做改进:①增加样本集容量,涵盖不同习惯用户的手势样本;②适当增加量化阶段编码本的大小,即增加聚类算法中聚类中心数目,以此减少原始数据在量化过程中的损失;③对于复杂的手势,可适当增加其对应的 HMM 状态数。如何最大化利用人类手势这一最自然的交互方式,建立一套形象且便于记忆的手势集,扩展现有的基于手势的交互功能库,丰富现有三维交互系统将是下一步的主要工作。

参考文献 (References) :

[1] 汪成为,高文,王行仁. 灵境(虚拟现实)技术的理论、实现及应用[M]. 北京:清华大学出版社,1996.

[2] BOWMAN D, KRUIJFF E. 3D user interfaces; theory and practice[M]. USA Boston: Addison-Wesley,2004.

[3] KELA J, KORPIA P, MANTYJARVI J. Accelerometer based gesture control for a design environment[J]. **Personal and Ubiquitous Computing**,2006,10(5):285-299.

[4] RODRIGUEZ P, GONZALEZ O S. Development of a 3D Real Time Gesture Recognition Methodology for Virtual Environment Control[C]//The 17th IEEE International Symposium on Robot and Human Interactive Communication. Munich, Germany: [s. n.],2008:279-284.

[5] HOFMANN F, HEYER P. Velocity Profile Based Recognition of Dynamic Gestures with Discrete Hidden Markov Models[C]//Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction. Bielefeld, Germany: [s. n.],1997:81-95.

[6] SCHLOMER T, POPPINGA B. Gesture Recognition with a Wii Controller [C]//Proceedings of the 2nd international conference on Tangible and embedded interaction. Bonn, Germany: [s. n.],2008:11-14.

[7] NG R T, HAN J. Efficient and Effective Clustering Methods for Spatial Data Mining[C]//Proceedings of the 20th VLDB Conference. Santiago de Chile, Chile: [s. n.],1994:144-155.

[8] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. **Proceedings of the IEEE**,1989,77(2):257-285.

[9] LEE H, KIM J H. An HMM-based threshold model approach for gesture recognition[J]. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1999, 21(10):961-973.

[10] PAN Zhi-geng, CHEN Wen-zhi, ZHANG Ming-min, et al. Virtual reality in the digital olympic museum[J]. **IEEE Computer Graphics and Applications**,2009,29(5):91-95. [编辑:李辉]

(上接第 83 页)

以产生足够大的光功率。O/E 转换的光接收器采用 Agilent 公司的 HFBR2412,以实现数字光通信时 HFBR2412 后需加一接收电路。接收电路中,在 VCC 与电源之间接了 1 个 10 Ω 的限流电阻和 1 个 0.1 μF 的旁路电容以去除噪声^[7]。

本研究用光发送/接收单元传输模/数转换器的移位时钟来测试光纤数字传输系统,移位时钟是频率为 1 MHz 的方波。从测得的整个光纤传输系统的输入/输出波形图中可以得到该光纤数字传输系统的输出延迟时间约为 150 ns,符合电子式电流互感器系统所允许的误差范围之内。

3 实验验证

本研究用准确度为 0.01 级的常规电磁式标准互感器对上述设计的电子式电流互感器数据采集系统进行了测试,测量通道实验结果如表 1 所示。从表 1 中看出实验结果中数据的线性度很好,通过示波器的观察发现,采样时钟信号中参杂了很多的高频干扰信号,A/D 转换结束后信号振荡现象比较严重,但是通过一个低通滤波器,很好地解决了上述问题,得到了一个稳定的输出^[8]。

表 1 测量通道实验结果

一次电流/kA	A/D 输出折合值/V	比差/(%)	角差/(′)
0.25	0.510 3	0.192	10
0.4	0.809 7	0.169	7
0.5	1.100 3	0.16	7
1	1.820 5	0.117	6

4 结束语

基于 TMS320F2812 的数据采集系统已经设计并调试完成,实验结果表明,该数据采集系统抗干扰性好,测量精度高,符合 IEC60044-8 的 0.2 级精度标准,能很好地完成信号的采集,满足电子式电流互感器对信号采集的实时性、可靠性、准确性的要求,同时该设计降低了系统的功耗,提高了系统的性能。

参考文献 (References) :

[1] 贾春荣,邸志刚. 一种新型电子式电流互感器模拟信号处理系统设计[J]. 电气应用,2007,26(1):108-110.

[2] 龙忠琪. 模拟集成电路教程[M]. 北京:科学出版社,2004.

[3] 苏奎峰,吕强. TMS320F2812 原理与开发[M]. 北京:电子工业出版社,2005.

[4] Texas Instrument. TMS320F2812 Digital Signal Processors Data Manual[M]. Texas Instrument,2003.

[5] Texas Instrument. TMS320F28x Event Manager (EV) Peripheral Reference Guide[M]. Texas Instrument,2002.

[6] 刘田豹,方春恩,李伟,等. 混合式光电电流互感器数据采集系统的设计[J]. 智能电器及计算机应用,2009(1):25-28.

[7] 丁全响,周有庆,吴桂清. 混合式光电电流互感器数据采集系统的研究[J]. 电力科学与工程,2006(1):34-40.

[8] 季小林,高晓蓉. 基于 TMS320F2812 的数据采集系统[J]. DSP 开发与应用,2007,23(8):169-171. [编辑:张翔]