

# 基于上下文感知的多模式移动导游系统设计

张磊, 俞建新

(南京大学计算机科学与技术系国家重点实验室, 南京 210093)

**摘要:** 移动导游系统已经引起广泛的关注。作为移动计算的一个典型实例, 许多研究机构开发了各自的导游系统。借鉴这些系统的优秀设计思想, 设计了一个基于上下文感知的多模式交互导游系统, 以提高系统的人机交互性能。

**关键词:** 导游系统; 多模式交互; 上下文感知

## Design of Multi-model Mobile Guide System Based on Context Awareness

ZHANG Lei, YU Jianxin

(State Key Lab, Dept. of Computer Science, Nanjing University, Nanjing 210093)

**【Abstract】** Mobile guide system has drawn more attention as a typical case of mobile computing. Many research institutes have developed their own tour guide systems. Based on these models, this paper comes up with a multi-model interaction guide system based on context awareness, which can improve performance of the system.

**【Key words】** Guide system; Multimodal interaction; Context awareness

近几年来, 移动导游系统已经引起越来越多的关注。通过 PDA 或者高端手机, 导游系统不仅能够提供给游客多种服务, 更重要的是它提供给游人一种无时无刻获取资讯的途径。本文致力于开发一套位置感知的多模式交互移动导游系统。移动导游系统原型, 具有如下特性:

(1) 系统采用多模式交互方式, 以提高移动环境中多模式交互的效率。实验数据表明, 使用语音和笔输入的多模式交互方式, 在易用性和交互效率上具有明显的优势<sup>[1]</sup>。因此, 在我们的导游系统原型中, 采用语音、笔、键盘和上下文输入等交互模式与系统进行交互, 有效改善人机交互的自然性和提高人机交互的效率。

(2) 系统采用上下文感知技术, 以提高系统的交互智能性。上下文感知技术是提高系统交互智能性的重要途径。使用上下文感知技术, 可以有效减少用户输入, 提高系统的交互智能性。在导游系统原型中, 系统通过 GPS 自动感知用户位置, 并向位置服务器主动发送位置信息, 由位置服务器经上下文计算后改变系统行为。

(3) 系统采用客户端/服务器 3 层结构, 在客户端和各种应用服务器之间加入 Guide 服务器。Guide 服务器功能主要包括: 语音识别, 日记生成和 Web 服务代理。Guide 服务器增加了系统的性能和可扩展性, 它是移动终端功能的延伸。在系统中引入 Guide 服务器, 也是该系统与多数导游系统的不同之处。

### 1 相关研究

从 20 世纪 90 年代初至今, 以手机、PDA 和掌上电脑为代表的移动设备得到了日益广泛的应用, 手持移动设备的硬件性能逐步提高, 无线通信技术的飞速发展, 使得传统的人机交互方式相形见绌, 已经成为移动计算的瓶颈。Smart Sight 是一个智能的移动导游系统, 系统装备了 GPS 接收器、麦克风、耳机, 视频相机和头盔显示屏等一整套交互设备, 不仅实现了 GPS 定位感知功能, 并且支持自然语言处理、语音识别、视频输入输出和手势输入等多种交互方式的多模式融合<sup>[5]</sup>。Lancaster Guide 是另一个基于位置上下文感知的移动导游系统, 系统使用无线通信技术向用户广播动态消息和发送定位信息。系统提供用户可定制的游览路线, 用户根据自

己的兴趣爱好、所在位置、时间安排、费用限制, 甚至当地的天气状况来定制属于自己的特定的旅行路线<sup>[6,7]</sup>。北京大学计算机系的 TGH 导游系统也是一个主要采用语音和笔输入的多模式交互的智能导游系统。多模式交互和上下文感知技术已被研究证明是提高人机交互自然性、智能性和效率的有效途径<sup>[2,3]</sup>。在借鉴了 Smart Sight 系统多模式交互和 Guide 系统上下文感知技术, 给出了基于上下文感知的多模式交互系统的设计。

多模式交互作为人机交互中的一项核心技术在国内外受到了普遍的关注。2002 年, 国际标准化组织 W3C 成立了多模式交互标准的制定小组, 其目的是开发支持手持移动计算多模式交互的协同并发机制, 它详细描述了多模式交互的框架结构、输入构件和输出构件。我们开发的基于手持设备的智能导游系统原型, 以它的框架结构为标准, 实现了语音、笔和键盘输入的多模式交互。

上下文感知技术在计算机科学的许多领域都有其应用, 位置信息是上下文早期研究中关注的主要内容, 哥伦比亚大学的 Schilit 以 GIS 服务为应用背景, 提出了移动计算系统的上下文感知模型<sup>[8]</sup>。该模型采用独立的感知器感知上下文信息, 并通过网络传给服务器, 服务器根据感知的位置信息采取相应的上下文处理, 提供用户不同的服务。本文提出的系统原型通过手持设备的 GPS 接收器感知用户位置信息, 然后手持设备提交位置信息到人机交互模块, 由人机交互模块处理位置上下文信息。这种系统结构的设计, 与 Schilit 提出的移动计算系统的上下文感知模型完全吻合。

虽然手持移动设备在处理能力和存储容量上比以前有了很大提高, 但是仍不足以满足语音数据处理的需要, 因此在

**作者简介:** 张磊(1980-), 男, 硕士生, 主研方向: 多媒体技术; 俞建新, 副教授

**收稿日期:** 2006-03-27 **E-mail:** michaelzhanglei@126.com

我们设计的多模式交互框架中,采用 Guide 服务器来专门负责语音部分的数据处理。Guide 服务器最主要的功能之一就是独立的语音识别功能,其中预定义了一系列的语音指令,这些语音指令作为快捷方式执行特定的行为。

## 2 关键技术分析

上下文信息可以简单地划分为主动上下文和被动上下文,主动上下文直接改变系统行为,而被动上下文计算则需要通过用户来改变系统行为。上下文也是一类特殊的交互通道。充分考虑上下文参与的多模式交互符合移动计算环境中的用户认知特征,设计了如图 1 所示的手持移动计算中的多模式交互框架。在这个框架中,将上下文作为一类特殊的交互模式,同 GPS、语音、笔式输入和键盘输入一起参与到人机交互中去。上下文参与的多模式交互框架如图 1 所示。

多模式交互框架定义一个多模式交互应用由以下几个基本组成部分构成:输入模块,输出模块,交互管理模块和系统环境模块。输入模块支持多种输入模式,比如音频、语音、手写和键盘等,每个输入模块又都由识别器、解释器和组合器组成,识别模块最后生成的结果提交到交互管理模块。交互管理模块管理各种输入输出模式在系统中的执行路径,维护交互会话,它是多模式交互的枢纽。输出模块支持语言、文本、图片和动画等多种输出模式,输出模块由输出生成器和输出模式接口组成。系统环境模块管理设备能力、用户参数和环境变量等参数,提供给交互管理模块查询使用。

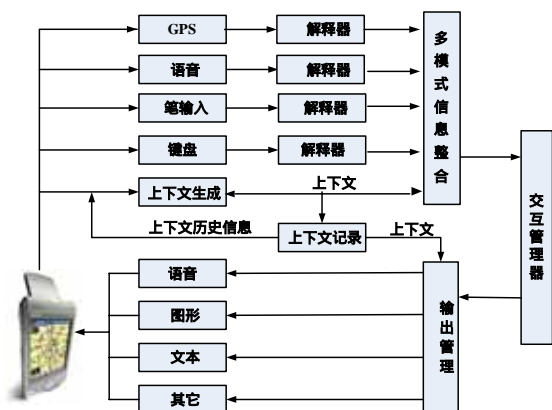


图 1 导游系统中的多模式交互框架

2005 年 4 月, W3C 组织定义了多模式交互的体系结构,在多模式交互框架上对多模式交互管理模块给出进一步的设计,详细描述了与用户输入输出模式之间的接口。这是一个具有良好封装性、易于扩展的、模块化的分布式体系结构。Multimodal Interaction(MMI)体系结构由 5 部分构成,分别是运行时框架、Dynamic Properties Framework(DPF)模块、数据模块、交互管理器和模式模块。

运行时框架提供系统的运行环境,装载和初始化系统的各个模块,生成底层的事件循环,各个模块之间通过事件进行通信。DPF 模块负责管理环境条件、设备配置和用户参数,提供其他模块需要的各种参数值以及动态接收参数更改通知来更新参数值。MMI 体系结构借鉴了 Model-View-Controller(MVC)设计模式的设计思想,数据模块、模式模块和交互管理器分别对应 MVC 设计模式中的模型、视图和控制器。在 MMI 体系结构中,每一种输入和输出方式都对应不同的名空间和标记语言,称为模式,它定义了用户的输入输出与系统交互的接口。交互管理器负责与各种输入模式协作,调度和

管理系统的执行路径,相当于 MVC 设计模式中的控制器。数据模块负责管理需要在各个模块之间共享的数据以及对数据的运算。

## 3 系统设计方案

在我们设计的系统原型中,系统由 3 部分组成,即手持移动设备、Guide 服务器和 Web 应用服务器。Guide 服务器接收移动设备发送的语音数据和 HTTP 请求地址,完成自然语言处理和语音识别以及请求转发功能,最后将语音识别的结果和 Web 服务器的内容返回到移动设备上。移动设备和 Guide 服务器之间通过无线媒介通信(如 GPRS, WLAN)。同时,在 Web 服务端,Guide 服务器向 Web 开发人员屏蔽了移动设备的特殊性,使得 Web 开发人员可以专注于 Web 应用的开发,Guide 服务器可以根据 Web 内容进行有针对性的格式转换和压缩,以便于 Web 内容在移动设备上的显示。我们设计的导游系统原型框架如图 2 所示。Guide 服务器和 Web 服务器的设计,本文不作赘述。

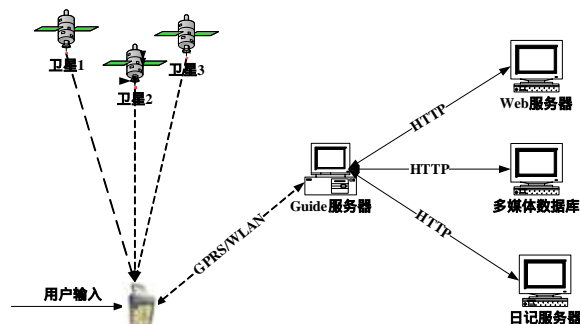


图 2 导游系统原型设计的框架

在我们设计的城市导游系统原型中,多模式交互模块的实现是非常重要的部分,也是本文介绍的重点。该部分主要在手持移动设备上实现,部分功能由 Guide 服务器实现,比如语音处理的部分,由 Guide 服务器来完成。导游系统原型中的多模式交互模块的体系结构如图 3 所示。

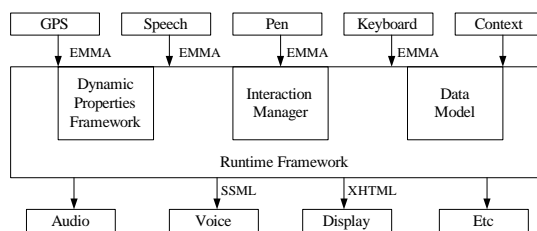


图 3 导游系统中的多模式交互模块的体系结构

导游系统原型包含的输入输出模式:GPS, 语音, 笔, 键盘和上下文输入模式以及语音, 图形, 文本等输出模式。基本的用户输入信息都采用 Extensible Multimodal Markup Language(EMMA)编码的形式提交到交互管理器。EMMA 是 W3C 定义的多模式交互接入 Web 的规范,它是多模式交互系统众多规范中的一部分,它给出了用户输入的标记语言的详细说明。用户还可以自定义上下文信息与交互管理器之间通信的接口。语音输出模式包含 Speech Synthesis Markup Language(SSML)标签的文本字符串组成,SSML 标签用来描述单词的发音。语音输出模块还可以输出预先录好的音频文件。交互管理器是多模式交互模块的重要组成部分,是与用户交互的枢纽。它接收用户输入的 EMMA 编码,并调度不同的输入模式编码到适当的生成模块。比如 GPS 接收器每隔

(下转第 282 页)