

,95

智能计算机接口与应用进展

——第二届中国计算机智能接口与智能应用学术会议论文集

主 编 吴泉源 高 文

主办单位 国家863计划智能计算机系统主题专家组

承办单位 中国科技开发院威海分院

清华大学出版社

前 言

并行化、小型化、网络化、智能化是当前计算机发展的主要趋势。人工智能技术与计算机主流技术结合，并融合在主流技术之中，旨在使计算机更聪明、更有效，与人更接近。智能接口的最终目标便是利用人工智能的理论与方法，建立一种拟人化的或自然化的人机交互的和谐环境，其最高境界是使人与计算机交互时“与人更接近”，能够像人与人信息交流时一样方便、自由；而智能应用的最终目标是融人工智能于计算机主流技术之中，建立一种具有人类丰富知识和高级思维与能力的计算机应用系统，其最高境界是使计算机求解问题时“更聪明、更有效”，能够像人一样解决各类复杂的应用问题。

智能接口与智能应用有着极其诱人的应用前景，同时，它们的许多课题都属于开放性研究课题。根据国家 863 高技术研究发展计划确定的“有限目标、突出重点”的选题原则，以及邓小平同志关于“发展高科技，实现产业化”的指示精神，我国智能接口与智能应用研究在突出计算机智能化的中国特色和走社会主义市场经济的道路上已经取得了一系列可喜的成果，并正在向更高、更广、更深以及更实用的方向发展。

为进一步推进我国智能接口与智能应用的研究开发工作，及时交流和检阅这方面的最新研究成果，自 1993 年以来，国家 863 计划智能计算机系统主题专家组决定每年召开一次全国计算机智能接口与智能应用学术会议。参加会议的不仅有承担 863 课题的研究人员，而且有许多来自全国从事其他课题研究的专家学者。本届会议来稿踊跃，得到了从事有关研究的全国广大科研人员的热情支持，论文的总体水平基本上反映了近几年我国智能接口与智能应用研究的国家水平。愿我们的同仁，特别是我们的跨世纪人才，在计算机智能接口与智能应用的研究开发中不断取得新成果，并为我国高技术的产业化做自己的更大贡献。

衷心感谢全国同仁对本届会议的积极参与和热情支持，衷心感谢国家科委高技术司和中国科技开发院威海分院对本届会议的关心和鼎力支持。

我们相信，本届会议一定能够在和谐、团结和充满朝气的气氛中取得圆满成功！

吴泉源 高文

1995 年 5 月

目 录

前言

特约论文

- 为实现和谐的人机环境,开展灵境系统的研究.....汪成为 (1)
多功能感知机的框架结构.....高文 (7)
汉字识别策略的探讨.....丁晓青 (21)
农业智能系统中多项新技术的应用.....熊范纶 等 (27)

第一部分 文字识别与符号分割

- 脱机手写汉字文本识别.....夏莹 马少平 常新功 等 (33)
联机手写符号识别的一种方法.....叶志远 李宝荣 顾小凤 (39)
汉英混排文本的识别.....郭宏 丁晓青 张忠 等 (44)
手写体数字的矢量化算法研究.....李学群 李天牧 (50)
汉字分类的实验研究.....胡家忠 (56)
中文 PDA 的设计与实现.....王福四 李铁才 舒文豪 (61)
大字符集手写汉字在线识别系统.....刘家锋 舒文豪 (67)
一种自由手写体数字识别方法研究
.....潘保昌 张平 汪同庆 等 (72)
工程图识别中粘连字符的分割.....杨立瑞 周济 (76)
一个具有良好健壮性的图文分割算法.....武卫红 (80)

第二部分 语言信号处理

- 实现汉语文语转换系统的新方法.....莫福源 李彤 李昌立 (85)
语音帧间相关信息对基于 HMM 系统识别精度的影响.....戴加宁 (89)
多模糊状态的非特定人语音识别研究.....袁中选 徐柏龄 余崇智 (95)
基于正则互相关的参数补偿方法在含噪语音识别中的应用.....俞栋 黄泰翼 (101)
普通话两音节间 F0 过渡的实验研究.....林茂灿 (105)
多音节词组语音输入中音节分割与清/浊音的切分技术
.....张凯 盛晔 严普强 (111)
连续语音训练词表的构造.....孙甲松 王作英 王侠 等 (116)
基于半音节 HMM 的汉语连续音节识别.....计天颖 杨浩荣 王作英 等 (122)
段长信息在语音识别中的应用.....李柏 王作英 (128)

采用 HMM 方法的 50000 词书面语连续语音识别基本系统——HUST1

.....	谢锦辉 陈浪舟 李 晖 等	(133)
汉语语音识别中声学模型研究的几个基本问题.....	方隼棠 李树青	(140)
基于 VQ-HMM 的非特定人语音识别与说话人自适应		
.....	张希军 徐 波 黄泰翼	(144)
提高合成语音表现力的研究.....	周俏峰 蔡莲红	(149)
噪声中语音的基音估测.....	吴元清 郑 榕 郝 杰 等	(155)
基于小波变换和零树模型的变速率语音编码		
.....	吴元清 郝 杰 郑 榕 等	(160)
一种汉语 F0 曲线函数模型.....	倪晋富 王仁华	(165)
语言模型复杂度的研究.....	吴 军 工作英 余嘉联	(169)
汉语双音节合成中的韵律调整.....	李 勇 程 俊 易克初	(175)
基于电话通道的特定人语音识别系统.....	施海航 俞铁城 张恩耀	(181)
一种多级判决的汉语说话者确认系统.....	张保轩 王连军 田 岚	(185)
基于小波变换的语音信号识别.....	刘金涛 黄岷雪	(191)
基于神经网络的汉语语音识别的研究.....	宋兴彬 宁 飞 马 驰	(196)
特定领域自然语言的关键词检出策略.....	李国勇 任为民 徐近霏 等	(203)

第三部分 计算机视觉与图象处理

图象中人体姿势的跟踪与识别.....	姚建华 俞志和	(209)
运动目标的自动捕获与跟踪.....	杨 威 冯秉瑞 张田文	(215)
表格识别系统.....	代远义 武裕朴 张佩芬	(221)
基于多尺度因子小波变换的图象压缩方法.....	向阳朝 戴汝为	(226)
基于机器视觉的实时交通流检测.....	宣国荣 夏炜新 储 震	(232)
基于模板匹配的多姿态人脸特征检测.....	游素亚 徐光佑 张永越	(237)
面部表情分析与识别的研究.....	高 文 金 辉	(243)
采用时间 BP 的运动目标识别.....	洪家荣 李星原	(249)
手势元遥控系统.....	王双林 高 文 刘 伟 等	(253)
一个实时的人脸识别系统.....	高 文 周吴煜 刘 岩 等	(259)
基于骨架校正和约束求解的矢量化方法.....	陈月林 朱建新 王平江 等	(265)
工程图纸矢量化过程中角度的测量和线条的跟踪.....	陆宗琪 张秋萍	(271)
基于分形特征的景象匹配方法之研究.....	田金文 高 谦 张继贤 等	(277)
层析成像分割与可视化联想策略及应用目标.....	杨真荣 董占球 阎树龄 等	(282)
自动识别航天图像中目标的研究.....	王润生 刘 方 周军红 等	(288)
坐标网格工程蓝图实时分割算法.....	岳春生	(294)
基于自适应网格的图象序列编码方案及其应用		
.....	张茂军 胡晓峰 熊 华 等	(298)
表格图象数据抽取方法的设计与实现.....	刘江宁 杨 嵘 张 剑	(303)

改进的剖分立方体算法及其在三维医学图象处理中应用	徐美和 唐龙宛 铭 等 (309)
支持三维医学图象处理的并行加速器 PA' 95 的体系结构设计 和系统软件构成	董迎飞 郑纬民 汤志忠 等 (315)

第四部分 多媒体、图形学与虚拟现实

GBH 的链——如何超越媒体空间	刘连芳 吴玉军 罗春晔 等 (321)
多媒体终端的设计与实现	钟玉琢 王凯 (327)
医学多媒体专家会诊系统中的网络通讯	关燕林 卢汉清 马颂德 (333)
分布式超媒体系统 HDB 的设计思想与实现	胡晓峰 吴玲达 老松杨 等 (337)
基于 Client/Server 计算环境的超媒体系统的设计思想	吴玲达 胡晓峰 老松杨 等 (343)
超媒体系统中扩展关系数据库的处理方法及实现	王炜 胡晓峰 老松杨 (349)
多媒体超文本系统中的一种图形对象数据模型	李国辉 库锡树 武德峰 等 (355)
一个基于对象时空描述的信息演播软件生成系统的设计与实现	汪诗林 张晨曦 李琼 等 (360)
多媒体证券交易知识信息系统	张建昌 (366)
多媒体知识工程软件环境 KESE	李卫华 张亮 韩波 等 (371)
自动会议安排系统 GKD-SCHEDULE 的设计与实现	贾焰 高洪奎 王怀民 (377)
一个基于 Client/Server 结构的多媒体合著系统的设计	朱海滨 王朴 吴泉源 (383)
CSCW 中的群体协作及其实现	沈清 (389)
CAD 集成框架中的超文本浏览器的研究	卢先兆 米强 李思昆 等 (395)
超媒体 ICAI 系统	陈卫芹 陆汝钤 (401)
超媒体手语计算机辅助教学系统的研究与实现(字典部分)	舒宏先 王申康 章高清 (407)
计算机动画系统及其开发平台的研制	方伟 俞昌 (413)
计算机动画技术在农业科学中的应用	冀松杨 张方 (420)
计算机动画制作中的几何设计与纹理映射技术	黄心渊 古梅 戈建涛 等 (425)
面向工程档案自动输入与存储管理的多媒体数据库系统	王岩 (431)
人脸表情模型虚拟实现方法的研究	傅强 谈正 (435)
人脸表情的合成方法研究	曾建超 金兰京 徐光佑 (442)
THMOS ——多媒体办公系统	苏军 钟玉琢 (448)
地图图例的层次结构与自动绘制方法	吴建敏 许卓群 (454)
二维地质构造图的智能处理	赖宏茂 胡宏涛 黄全舟 (460)
油气藏图件智能绘制系统数据库智能接口设计	胡宏涛 黄全舟 赖宏茂 (465)

第五部分 机器翻译与自动文摘

- 机器翻译的歧义与语义知识的关系.....王明昕 史晓东 陈火旺 等 (469)
- RVG 文法研究.....王 挺 史晓东 陈火旺 等 (477)
- MATRIX 英日机译系统的研究与实现.....张 静 史晓东 陈火旺 (483)
- 基于模式的汉-英机器翻译系统中自动分词算法的设计与实现
.....王海峰 李 生 张 民 等 (489)
- 汉语全文文本检索系统开发平台.....刘开瑛 (495)
- 汉藏机译系统中以词组为主的藏语词语分类初探.....德盖才郎 李延福 (501)
- 面向实例、基于模式的机器翻译方法与实现.....赵铁军 李 生 王海峰 等 (507)
- 论中文文献文摘的自动编制.....王永成 王继才 (513)
- HIT-836 I 型非受限域中文自动文摘系统.....王开铸 李俊杰 吴 岩 (519)

第六部分 知识工程与智能应用系统

- 文献智能检索系统的研究.....葛家翔 钱 放 刘耀东 等 (523)
- 图书采购辅助模糊决策.....蒋 琢 胡运发 钱 放 (529)
- 基于实例的推理在智能诊断中的应用.....魏蛟龙 周曼丽 朱耀庭 (535)
- 金融领域战略战术决策模型的设计与实现.....杜长征 应启瑞 杨 帆 (539)
- 一个通用的交通系统管理决策支持系统 CCTP-TSMDS
.....童学锋 吴永明 杨佩昆 等 (545)
- 中国银行智能决策支持系统.....赵保华 潘少钦 (550)
- 启发式搜索策略在资源调度中的应用.....毕忠梅 (556)
- 多功能智能化刺绣编程系统.....谢 琪 谢俊元 高 勇 等 (562)
- 综合应用智能技术进行油气藏描述的基本思想.....张学工 刘业新 李衍达 (568)
- 基于统计知识的相似性评价.....韩 柯 李德毅 原晓玲 (573)
- 计算机辅助服装款式设计.....邹定国 卢习林 林德静 (578)
- 中文科技文献自动标引系统概述.....朱 立 胡运发 (584)
- 人工智能在地震综合预报中的应用.....朱岳清 梅世蓉 张国民 等 (590)
- 一个实时的基于并行的专家系统的研究及实现.....胡运发 邓胜辉 戚成功 (596)
- 储集层伤害诊断专家系统 DFDES 中推理机的设计与实现.....徐英卓 郭建明 (602)
- 不确定算法中的灰色模型及其应用.....郭建明 徐英卓 李 棋 等 (608)
- 塔里木保护储集层软件智能决策模型的研究与应用
.....张绍槐 郭建明 李 棋 等 (614)
- 一个智能化的高层次决策训练系统.....李乃奎 田渭涛 范美成 等 (620)
- 一个基于分布式客户/服务环境的 DIS 原型系统构想
.....李梦汶 张最良 许瑞明 (625)
- 面向作战决策仿真的神经网络专家系统.....林 东 张最良 (632)

用于密钥管理方案分析的专家系统.....	王茂玲	(639)
城市煤气中压管网抢修计算机辅助决策系统.....	陆宗骥 冯琪龙 高敦岳 等	(645)
洛河发电厂电气设备操作票专家系统的研制.....	陶维青 鲍道良 桂汉生	(651)
智能地震识别目标技术研究.....	王伟策 陈晓波 温蓉	(657)
智能决策支持系统的推理机制及其并行实现.....	陈华平 陈国良 孙志霞 等	(662)
宏观经济智能决策支持系统 MEIDSS	田园 李彤 冯珊	(668)
大型减速器 CAPP 专家数据库系统中的知识表达与处理	陈维克 张赛珍 等	(675)
一个基于知识的模型库管理系统设计	杜晓明 赵强 王兴旺 等	(680)
地震预报智能决策支持系统的总体结构.....	周胜奎 汪成民 黄金丽 等	(685)
油气开发智能信息综合集成系统设计.....	成绥民 薛中天	(691)
引入构造解释信息减少油藏描述中的不确定性.....	邵旭辉 李衍达	(697)
FoxPert: 一个增强型 FoxPro/Clips 松耦合结构的研究与实现	杨文婷 邓铁清 何久田 等	(702)
导航源智能管理.....	宋兆虎 范跃祖 宁文如 等	(708)
军事后勤保障决策模型中的知识工程方法.....	李德彩 周堤基	(715)
地理信息系统在 G3 1 领域中的应用	陈东雨 张捷	(721)
定性约束分析法及其应用.....	刘江宁 张华滨 吴泉源	(727)
新型信息管理与决策支持系统中的关键技术研究.....	周立柱 朱滇	(733)
维吾尔语语音识别与合成中的语音理解处理	地里夏提·义福组拉 吾守尔·斯拉木	(739)

为实现和谐的人机环境，开展灵境系统的研究

汪成为

中国工程院院士

国防科工委科技委常任委员

863计划智能计算机专家组组长

摘要：本文讨论了灵境（Virtual Reality）技术和多维信息空间（CyberSpace）的基本概念及其在实现和谐的人机环境中的作用。

关键词：灵境技术，多维空间，人机环境。

应该承认，最近数十年来计算机技术的进展已大大地减小了人与计算机之间的隔阂，但尚未能从根本上解决现有的计算机的不合理性。人们已经认识到，为了适应未来信息社会的需要，必须提高人与信息社会的接口能力、提高人对信息的理解能力，因为这是发挥信息的共享性和增值性的首要前提条件。为此就驱使我们去探讨什么是表示、传送和处理信息的最佳技术途径？

比较一致的认识是：能较完整地表示概念、能较迅捷地传递概念、能以符合人的感知和认知过程的方式对概念进行加工的方法就是较理想的信息表示和处理的途径，它就能使人完成某个智力任务的行为过程得到较大的改善，就使人与计算机的关系变得更为和谐了。人们普遍认为，21世纪的信息技术将在建立和谐的人机环境方面会有许多创新性的发展，从而促使信息应用系统在深度和广度上发生本质性的变化。

总之，为了适应于21世纪信息社会的需要，人不仅仅是要求能通过打印输出、或显示屏幕上的窗口，从外部去观察信息处理的结果，而且要求能通过人的视觉、听觉、触觉、嗅觉、以及形体、手势或口令，参予到信息处理的环境中去，从而取得身临其境的体验。这种信息处理系统已不再是建立在一个单维的数字化的信息空间上，而是建立在一个多维化的信息空间之中了，建立在一个定性和定量相结合、感性认识和理性认识相结合的综合集成环境中。VR（Virtual Reality -- 灵境或虚拟现实）技术将是支撑这个多维信息空间（CyberSpace）的主要关键技术。

为了使计算机不仅仅是人进行逻辑思维的有力工具，也是人进行形象思维的帮手，首先就要求计算机能适应于人所惯用的信息获取形式和思维过程。例如：人并不是仅仅靠听和读文字（或数字）材料获取信息的，而是通过他与所处环境的交互作用、利用人本身对所接触事物的感知和认知能力、以全方位的方式获取各式各样表现形式的信息。

怎样从技术的角度去说明上述这种适应于人的信息环境的特点呢？Burdea, G. 在 Electro'93 International Conference 上所发表的 "Virtual Reality Systems and Applications" 一文中，曾提出了一个“灵境技术的三角形”，它较简捷地说明了灵境系统的基本特征，即：三个“1”，它们是 Immersion - Interaction - Imagination（沉浸 - 交互 - 构想）（如图1-1所示）。应该指出，这三个“1”既是灵境系统的主要技术性能，也是和谐的人机环境的基本特征，这三个“1”强调了人在以计算机系统为主的环境中的主导作用。

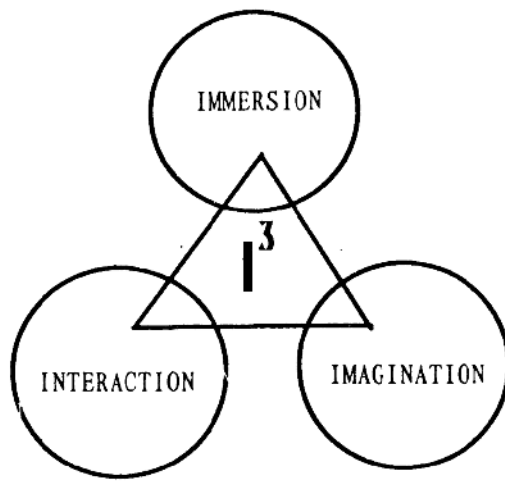


图 1 灵境技术三角形

构建以上所描述的一个灵境系统的基本手段和目标是：利用并集成高性能的计算机软硬件及各类先进的传感器，去创建一个使参与者处于一个具有身临其境的沉浸感、具有完善的交互作用能力、能帮助和启发构思的信息环境。图 1—2 表示出构建这样一个灵境系统的根本目的、应克服的主要技术瓶颈和所需的基本计算机软硬件支撑技术。

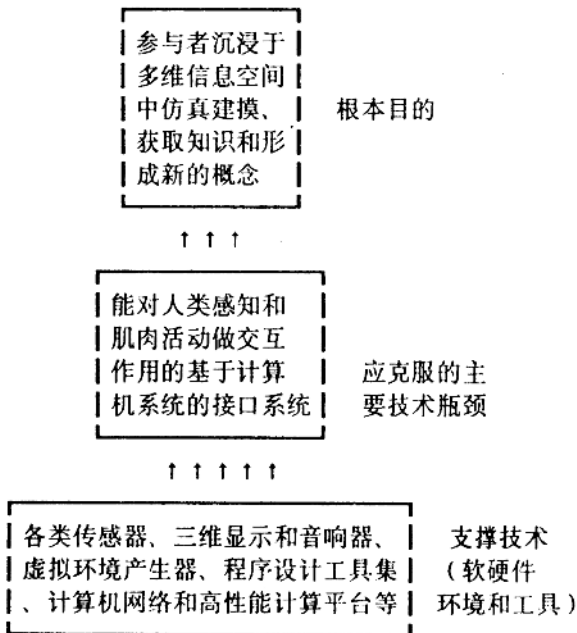


图 2 构建灵境系统的目的和所需的支撑技术

经过世界科技界十余年的努力，目前已能提供一些为构建一个实用的灵境系统所必需的软硬件设备和工具了，如：

1. 跟踪系统：以确定参与者的头、手和身躯的位置；

2. 触觉系统: 提供力与压力的反馈;
3. 音频系统: 提供立体声源和判定空间位置;
4. 图象生成和显示系统: 产生视觉图象和立体显示;
5. 高性能计算处理系统: 具有高处理速度、大存储容量、强联网特性。

在软件方面,除一般所需的软件支撑环境外,主要是提供一个产生虚拟环境的工具集或产生虚拟环境的"外壳".它至少应具有以下的功能:

1. 能够接受各种高性能传感器的信息,如头盔的跟踪信息;
2. 能生成立体的显示图形;
3. 能把各种数据库(如地形地貌数据库、物体形象数据库等)、各种CAD等软件进行调用和互联的集成环境。

到目前为止,几乎所有这些软件系统都是建立在面向对象方法学基础上的。灵境的实现需要面向对象技术的支撑,灵境的实现又促进着面向对象技术的改进。从技术实现层次而言,灵境系统对软件技术的特殊要求是:

- 能对多维信息所描述的对象进行表示;
- 能对多维信息所描述的对象进行实时处理;
- 能对多个多维信息所描述的对象进行并发处理。

为此,对面向对象技术进行了一系列的改进,主要是从以下几方面入手的:

1. 从多维信息表示的需求出发所提出的改进

由于在多媒体、超媒体和灵境系统中,所处理的已不是仅仅用单维的数字化信息描述对象的,而是具有以下特征对象:

- 以多种媒体为信息载体、以多维信表示的 (Media to be mixed) 对象,并且要求“对象”具有对不同媒体作统一的表示的能力;
- 各种媒体间和各维信息间可相互引用、和维与维间的连接(即所谓的超连接 cross-referncing of media and hyperlinking);
- 多维信息的信息共享;
- 多维信息的联网和开放存取 (networking and open access)。

为了实现以上的特征,至少应对原有的面向对象中的信息封装 (Encapsulation) 的概念进行扩展,使之能通过一个公共的接口对不同的媒体和不同维的信息进行存取,同时为了能实现网络化、分布式、多用户的多媒体、超媒体和灵境的应用系统,目前已初步研制成多种可运行在Client-Server环境上的并发的面向对象语言(如COOC——concurrent object oriented C语言)、网络化的超媒体工具集、各种可共享的超媒体(多维信息)组件(Groupware)、对智能体的更新及在系统中的生命周期的控制软件等软硬件系统,但除了少量的已能作为产品提供使用外,大部分还处于试验阶段。(图3)是适应于这种多维信息处理的一般原理性框图

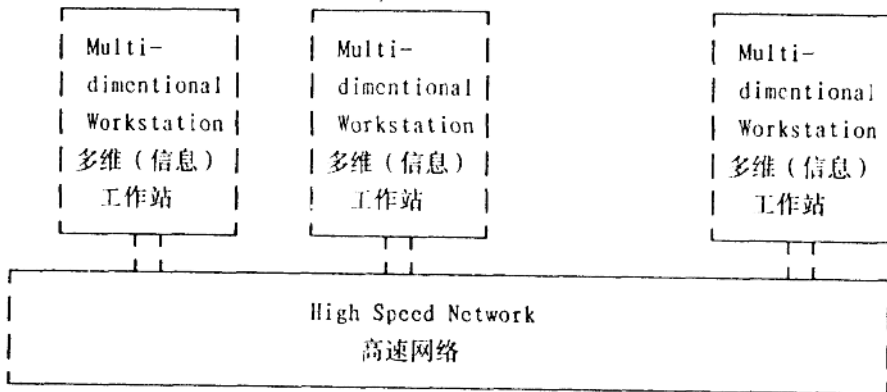
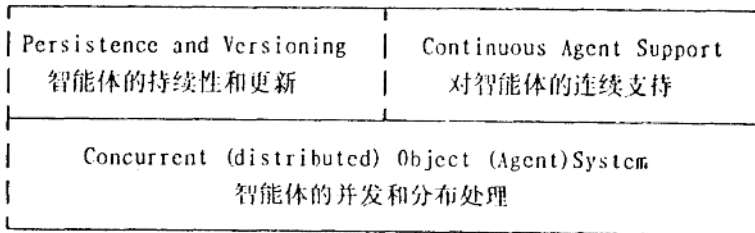
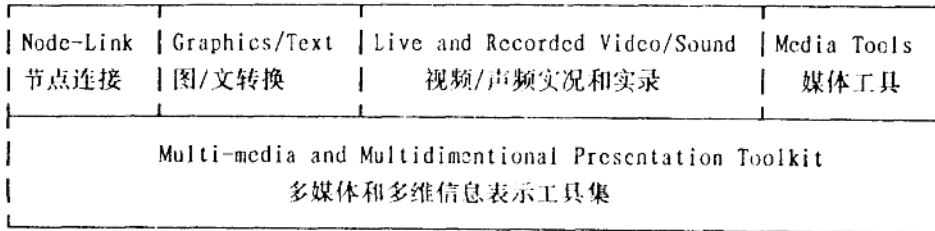
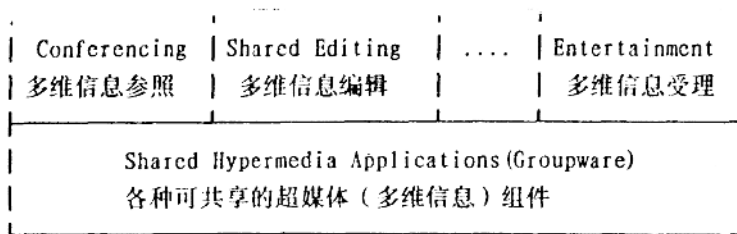


图3 多维信息处理系统的原理框图

2. 从实时处理需求出发所提出的改进

如目前已提出的实时C++ (即RTC++) 是对原有C++的扩展, 是建立在实时对象的概念上的. RTC++增加了能显式地把时间特性也封装 (encapsulate) 在对象内的功能. 这样就使对象更为活跃, 更为主动和活跃了 (更为active了). 其具体方法是在对象内封装一个或数个线程 (Thread), 一个线程是由数个进程组成的一个调度单位.

3. 从并发处理需求出发所提出的改进

在并行程序设计中, 经常使用的是三种方法, 即: 共享存储系统、消息传送系统和数据流系统. 为满足这些方法的需求, 目前已对原有的对象的功能作了扩充, 如Actor和

Agent即为一例。Actor既是实现系统中对这种新的“对象”的称谓，又是一种具有并发消息传送功能的语言。它的主要概念是：系统中的任何对象被称为Actor，用在Actor间传送消息的办法来实现系统中的每一件工作。每个Actor的属性（behavior）是用脚本（script）定义的。每个Actor还有一些关系（acquaintance），即能和本Actor进行消息传递的其他Actor的名字。

4. 从人工智能技术出发所提出的改进

问题求解是人们在从事科学研究和工程设计时的基本活动，也是建立灵境系统的最终目的。而归纳推理、演绎推理和推断推理又是问题求解的基本类型。传统的信息处理技术能进行与归纳推理和演绎推理有关的操作，但较难进行与推断推理有关的操作。因为推断推理是在基于知识、基于经验的基础上进行的，是感性认识和理性认识相结合的产物。推断推理的结果可以增加、改变原有的知识。因此，就进一步要求对象不仅能表示知识，更重要的是对象应是开放的，是能够从外部吸收新的知识的。我们把这种更“聪明的对象”表示为：

智能体 = 知识表示结构 + 处理方法。

为了能使用户在网络环境下开展灵境技术的研究和应用，除了提供必要的软硬件设备和工具外，更重要的是为用户提供一个虚拟环境管理器、一个完整的灵境系统开发环境。它应该具有以下功能。

1. 虚拟环境管理器除了能提供各种高级的传感装置（如头盔式显示器、数据手套、三维鼠标等）的驱动程序外，还能向用户提供了一个称为“脚本（Script）”的描述文件，用户可很方便地用它来定义、启动和运行一个虚拟的世界。多个用户还能通过网络实现网络配置、通信管理、并发控制、资源共享等功能。根据用户在描述文件中所规定的要求启动一个用OML所描述的虚拟环境。这样就大大地减少了用户为构造一个虚拟环境所花费的大量时间和繁琐过程。在“脚本”描述文件中包括以下内容：

- 设备配置：规定本地的设备配置；
- 对象原型文件：OML解释执行程序将根据此文件进行解释执行；
- 各种实例：规定每个对象的各种实例，包括对象名、实例名及与实例有关的参数等；
- 对象状态：各类状态参数；
- 网络配置：规定网络特性、并发控制信息、对象及实例共享信息等。

2. 能向用户提供一个构造对象的语言，开发环境内的编译器能对用这种构造对象语言所写成的文件进行解释执行，从而生成对象的原型。在系统的运行时间内，对于同一个对象原型可以衍生出任意个带有不同参数的实例，因此每个实例可以具有各自的几何实体和状态。

图4就是灵境系统开发环境的主要软件构成部分的示意图。

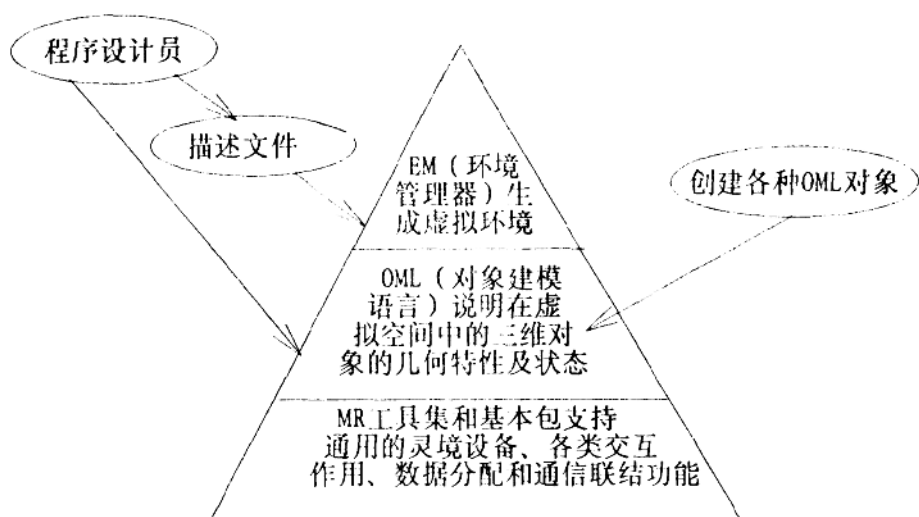


图4 灵境系统开发环境的主要软件组成部分

建立和谐的人机环境将是一个长远的目标。在本世纪内计算机技术中的两个M和两个O（即极大规模的并行处理、多媒体、面向对象、和开放系统技术）已为此作出了贡献，但仍然不能满足人们的要求，人们仍然感到自己与计算机之间存有间隙。当本世纪内在灵境技术方面取得一些进展时，人们对它寄予很大的希望，但也认识到尚有许多困难等待着我们去克服，客观的需求在呼唤我们尽快地把灵境技术在实验室内的成果移植到实际的应用系统中去。

多功能感知机的框架结构

高文

哈尔滨工业大学计算机系,150001

摘要:本文讨论面向人体语言感知的多功能感知机的机理与框架结构。可以认为人类语言是由自然语言和人体语言组成的。在人与人面对面的通讯中自然语言与人体语言同时或联合传递信息,因而研究人体语言的感知模型以及与自然语言的信息融合模型对于提高计算机的自然语言理解水平和加强人机接口的可实用性将是有意义的。人体语言是由包括面部表情,手势,体势,说话时所伴随的唇动,说话时所伴随的语调等等部分组成的。本文所提出的多功能感知机将力图分层次实现这些功能部分并提供信息融合的功能以便最终实现支持人类语言感知的计算机系统。

1 人体语言研究与多功能感知机

在日常生活中,当人与人进行面对面的通讯时我们不但使用包括口语(语音),书面语言(文字)等等被称为自然语言的语言通道,同时我们还经常使用手势来强调某些重要的部分或表达某些特殊的意思,使用特殊的语气或声调表达特殊的含义,使用特殊的表情或动作表达和传递许多只可意会,不可言传的信息。为了与基于文字的自然语言相区别,我们不妨称包括手势,表情,体势等等语言部分为人体语言现在在利用计算机认知的领域和研究智能人机接口领域,人们大多注意自然语言理解的研究,较少注意人体语言认知的研究。在人与人面对面的通讯中自然语言与人体语言同时或联合传递信息,因而研究人体语言的感知模型以及与自然语言的信息融合模型对于提高计算机的自然语言理解水平和加强人机接口的可实用性将是有意义的。

研究人体语言认知的意义在于,首先有可能为提高语音识别的识别率提供帮助。因为我们人在进行交谈的时候不仅要听对方的话,还要看着对方,这会帮助我们理解对方表达的意思。其次,在许多特定的场合下使用手势进行会话或控制是方便的,例如篮球比赛中的裁判员向记录员和主席台观众报告犯规队员号码的时候使用手势;再次,利用人体语言的检测来自动判断当前会议室中谁是发言人等等,在电视会议中的图象压缩或视景自动切换的应用中是十分有用的;还有,新的计算机接口技术希望能够实现无接触的三维鼠标或虚拟触摸屏,这些都可以利用人体语言认知技术进行研究与实现;最后,人体语言研究的一个最直接应用是为聋哑等残疾人提供更好的服务和与其它正常人进行交流的工具。

我们选择研究人体语言认知的一个重要原因是,自然语言是人体语言的高级阶段,是人体语言进化的结果。如果我们从基本神经元开始去研究智能,或者从昆虫开始研究[Brooks90,Brooks91],那麽我们将需要许多许多年才能到达人类智能的水平。但许多人利用专家系统等使用知识表达的高级系统去研究自然语言的理解问题总是逃脱不了只能在有限领域解决有限问题的约束。笔者认为这是因为我们对于自然语言中的内部模型和了解的不够多。比较起来,人体语言要比自然语言简单的多,直观的多,因而选择人体语言作为研究人类智能的突破口也许是合适的。

本文从人体语言的概念和功能入手,讨论如何利用现有的人工智能和计算机接口技术实

现对人体语言的认知,提出了多功能感知机的框架结构,提出了多功能感知机中的融合模型和个子功能的结构模型。提供了子功能系统的主要实现方法。所提出的多功能感知机将力图分层次实现这些功能部分并提供信息融合的功能以便最终实现支持人类语言感知的计算机系统。

2 人体语言的概念与多功能感知机基本构成

本文将人体语言定义为是由包括面部表情,手势,体势,说话时所伴随的唇动,说话时所伴随的语调等等功能部分组成的独立或伴随自然语言发生的一种信息传递与交换行为。如前所述,我们认为人类语言应该分为自然语言和人体语言两类,其结构与包含关系如图1所示。在自然语言类中,其功能又可以进一步划分为口语与书面语。如果从文字识别的角度看,书面语有可以分为手写与印刷语言。这是从模式识别和人工智能角度研究自然语言理解的人们所熟悉的。在人体语言类中,其功能可以划分为手势和手语,头势,体势,面部表情(包括唇动)等等。表1列出了人体语言中的一些实例。根据达尔文的生物化学的研究[Ekman75],自然语言随着民族和国界的不同而不同。同样,人体语言通常是与人所成长和生活的文化背景有关的,很多手势或面部表情在不同的国家和地区会有不同的含义。

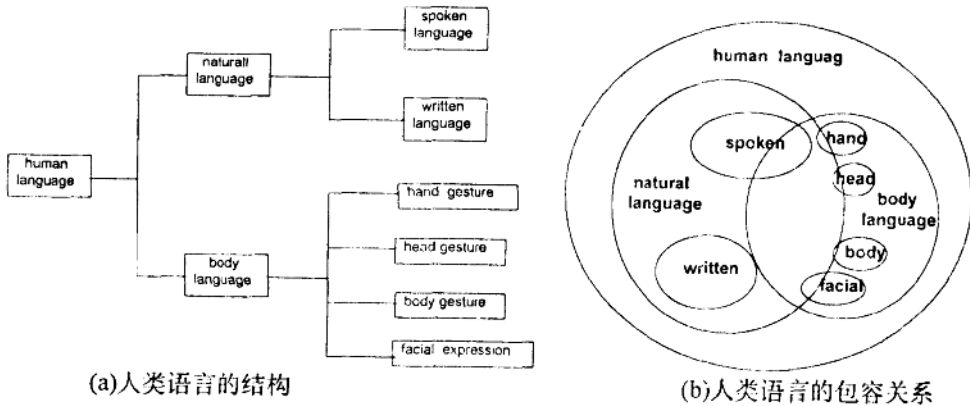


图1 人类语言的结构与包含关系

由图1(b)的包含关系可以看到,自然语言中的语音部分与人体语言的关系比较密切。

为了对人体语言进行研究,我们设计了一个多功能感知机系统。系统由硬件平台(见图2)和若干软件模块组成(见图3)。该系统的初期软件工作是分别在Sun工作站和SGI工作站上进行的,最终完成的系统是在以PC或工作站为主机的,有图象输入和语音输入的,有专用硬件加速部件所组成的系统结构之上的。软件的设计采用自顶向下(Top-Down)和自底向上(Bottom-Up)相结合的原则,各模块所使用的与具体算法有关的库函数先进行编程挑试,系统功能和各子系统的界面设计规划。在系统组织和软件设计中,采用面象智能体技术和面象对象技术。

表1 人体语言的部分用途举例

子类	动作	含义
手势	计数	条理

手势	挥拳	强调决心
手势	挥手	信心
手势	摇手	再见
手势	拇指指	了不起
手势	倒拇指	不值一提
头势	点头	同意,肯定
头势	摇头	不同意,否定,怀疑
体态	耸肩	不值可查,无可奈何
面部表情	扬眉,瞪眼,张嘴	吃惊,惊奇
面部表情	扬眉,撇嘴	恐惧
面部表情	压眉,闭嘴	厌恶
面部表情	皱眉,闭嘴	愤怒
面部表情	舒眉,咧嘴	高兴

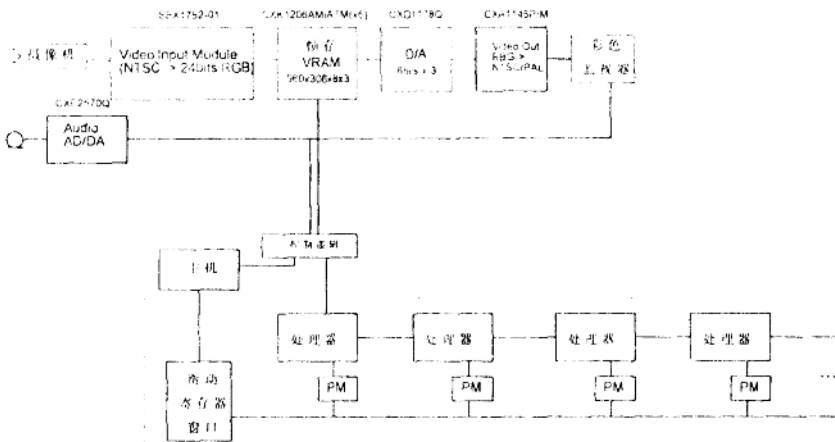


图2 多功能感知机的硬件平台

3 手势识别与合成

手势识别与手势合成是手势理解问题的正反两个方面,分属于模式识别和计算机图形学两个范畴。然而,这两个方面又是相互联系的,因为手势识别将为合成手势的准确性修正提供帮助,而手势合成的动力学模型又可以为手势识别提供参考。

3.1 手势识别

手势识别主要要解决以下两个步骤的问题: (1)实现手势目标的实时获取; (2)对手势目标的正确识别。手势目标的实时获取是指在任意复杂的背景条件下从图象流中分割出目标来,这是模式识别主要研究的课题之一,目前已有许多针对专用计算机视觉系统的较为成熟且易于实现的技术。例如,利用目标窗与背景窗的直方图分割目标的方法[Gilbert80],基于多图象信息的目标分割方法以及二维熵的阈值分割法等[Brink92][Liu91]。我们假定图象

的背景在短时时间段内是相对稳定的。在此假定下,针对所采集图象的背景不变性以及目标在视区的任意性采用了目标图象与背景图象的差分二值化的方法提取目标。对于手势目标的识别而言,我们将考虑:(1)孤立单手势目标的识别,(2)连续双手势的识别。首先,提取出能够正确表达该目标且具有几何不变性的特征是正确的识别目标的前提;目前所使用的特征主要有视觉特征、统计特征、变换特征以及代数特征等[Spirkvuska92],这些特征在一些具体的任务中体现出一定的优越性。经过比较分析,在经过大量实验的基础上,我们确定了使用基于边缘特征的差分链码表达方法。该方法针对边缘特征抗干扰性差等特点,对基本链码施加了差分以及多通道、多阈值滤波等处理提取出能够表示目标的不变性特征以供识别。

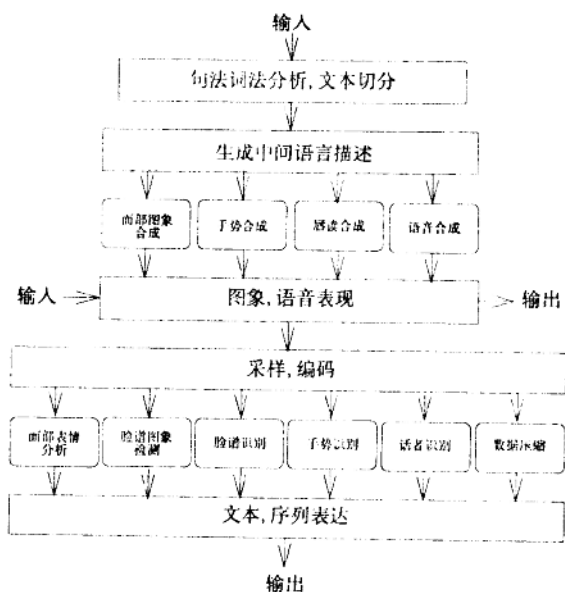


图3 多功能感知机的主要模块

对于手势目标的识别我们分两步走。我们首先考虑孤立单手势目标的识别问题。在这个问题很好解决的基础上,我们将孤立单手势目标识别技术与隐马尔可夫模型(HMM)[Rabiner89]相结合,实现第二步即连续双手势的识别。本小节的以下部分介绍的主要是我们第一步的部分工作结果。

3.1.1 手势目标边界的获取

手势目标的捕捉是实时地用摄像机对视区进行监视,检测目标的存在性、稳定性以及完整性。对目标背景的复杂程度、手势伸入视区窗口的角度以及摄像机的焦距等都没有具体的要求,不失一般性我们有以下限制条件:

- 手势必须以正对着摄像机的方向伸入且左右手不限;
- 背景图象与目标图象在灰度空间上或其它空间上有明显的差异(例如在灰度空间上有