

语音图像 通信信号处理

第七届全国学术会议论文集



1995.10.16-10.19 西 安

SICS'95

前 言

第七届全国语音图像通信信号处理学术会议于1995年10月16日至19日在西安召开。这个学术会议是由中国声学学会语言、听觉和音乐声学分会与中国电子学会信号处理学会语音图像与通信专业委员会联合主办的,每两年一届,本届会议由西安电子科技大学承办。

这个学术会议涉及信息科学中一些前沿领域的课题,引起了同行学者们的普遍关注。由于大家的共同努力,历届会议起到了沟通科技信息、启迪学术思想、促进友好合作的作用,并且形成了严肃、认真、团结、活泼的优良传统。

本届会议的论文征集工作得到了国内同行的热烈响应,共征集到论文114篇,经程序委员会集中评审,录用了其中91篇,(包括大会特邀报告3篇,摘要14篇),载入论文集正式出版发表,同时在本次会议上进行报告交流。登载在论文集内的论文包括以下几个方面:

语音识别 18 篇

图像处理 30 篇

语音分析 9 篇

语音合成 14 篇

语音编码 8 篇

通信与信号处理 12 篇

这些论文覆盖范围较广、内容丰富多彩,既体现了国内有关领域的最新研究成果,也在一定程度上反映了当前的国际学术动态。有些论文学术水平较高、有创新精神,有些论文介绍了高技术产品开发和新技术应用方面的最新成果。没有入选的论文也有相当多是有新思想和很有发展前景的,但我们相信这些论文在下一届会议上交流时将会发展得更加成熟和完美而未选入,当然也可能有个别好文章漏选了。在此我们向所有投稿的作者以及以其它方式关心和支持本届会议的同志表示衷心的感谢,欢迎他们到会进行学术交流。

人类步入了信息化时代,信息科学技术的发展日新月异,其中语音处理、图像处理、通信及其中的信号处理、一般信号处理理论和技术近年来一直十分活跃。它们有的本身就属于学科前沿,有的与信息科学中多个前沿学科有着十分紧密的联系。这些领域是出成果出人才的圣地。我们希望本次学术会议能对国内这些领域的发展起较大推动和促进作用。预祝大会成功。

袁保宗 张家骅

一九九五年九月

第七届全国语音图像通讯信号 处理学术会议

(SICS'95)

1995年10月17日~19日 西安

主办单位 中国声学学会语言听觉音乐学分会
中国电子学会信号处理学会语音图像通信专业委员会

承办单位 西安电子科技大学

大会主席 袁保宗 北方交通大学(100044)

副主席 张家骥 中国科学院声学研究所(100080)

程序委员会

主席 谢维信 西安电子科技大学(710071)

副主席 易克初 西安电子科技大学

委员 诸维明 北京邮电大学科学技术应用所(100088)

齐士铃 中国科学院声学研究所(100080)

黄泰翼 中国科学院自动化研究所(100080)

方棣棠 清华大学计算机系(100084)

刘村友 (成都)西南通信研究所(610041)

徐秉铮 华南理工大学(510651)

蔡德孚 中国科学院电子学研究所(100080)

方至 中国科学院心理研究所(100080)

顾瑞 (北京)中国人民解放军总医院

侯朝焕 中国科学院声学研究所(100080)

李昌立 中国科学院声学研究所(100080)

吴国威 清华大学电子工程系

赵荣椿 西北工业大学(710072)

刘增基 西安电子科技大学(710071)

阮秋奇 北方交通大学信息科学研究所(100044)

大会秘书组

秘书长 周世禄 陈克勤

秘书 张宏旗 张泽升 程俊

综合业务网理论及关键技术国家重点实验室简介

综合业务网理论及关键技术国家重点实验室位于古城西安,设在西安电子科技大学。1989年由国家教委、国家计委批准立项,1991年开始建设,目前已对国内外开放。

现代通信网正朝着数字化、综合化、宽带化、智能化和个人化方向发展。综合业务数字网已成为通信网发展的必然趋势和八十年代以来通信学科的国际前沿研究课题。本实验室的研究方向是综合业务网,包括综合业务局域网 ISLN、综合业务数字网 ISDN 和以异步转移模式 ATM 为基础的宽带综合业务数字网 B-ISDN,以及相关的信源编码、信道编码与密码、信息传输与交换等方面的理论与技术。实验室正在研究的许多课题对国家通信高技术发展计划及未来的“信息高速公路”建设均具有重要意义。

实验室十分重视贯彻“开放、流动、联合”的方针和边建设边开放的原则。实验室由三个专业研究部和综合室组成。综合室负责实验室的运行、维护和管理;通信网专业部侧重研究 ISLN、ISDN 及 B-ISDN 的关键技术等;信源编码专业部侧重研究语音、图像编码与数据压缩;信道编码专业部侧重研究新的调制解调技术、纠错码、密码及信息安全保密技术。研究队伍按照联合、流动的原则构成。第一批受聘的研究人员共 70 名,其中固定研究人员 38 名,客座研究人员 32 名;有博士生导师 6 名,正高职 19 名,副高职 25 名,博士后 4 名;40 岁以下人员占 55%。

为实现研究手段现代化,实验室引进了先进的电子仪器设备,如计算机工作站网络系统,程控数字交换机,先进的通信测量仪器及系统等。

实验室的研究方向由学术委员会把握,1993 年 7 月学术委员会第一次会议制定了“开放研究课题指南”,目前已资助开放课题 11 个。热忱欢迎有志于从事综合业务网理论与技术的科技工作者,特别是中青年学术骨干来我室从事合作研究与开发。

实 验 室 主 任: 刘增基教授

学术委员会主任: 樊昌信教授

邮 编: 710071

电 话: 5261024—2524

目 录

一、大会特邀报告

- 1.1 三维影像的研究及应用.....袁保宗 唐晓芳 (1)
- 1.2 从汉语的结构特点看汉语语音识别合成研究.....易克初 张家骥 (6)
- 1.3 立体显示技术和虚拟现实.....谈 正 (11)

二、语音分析与识别

- 2.1 一种灵巧的语音处理卡设计.....蒋烈辉 (15)
- 2.2 三种不同语速中的汉语语调分析.....石 锋 (19)
- 2.3 语音基调检测方法比较研究.....杨道淳 方 元 张 渊 (23)
- 2.4 语音信号的非线性预测.....余崇智 谭廷忠 袁中选 徐柏龄 (27)
- 2.5 一种汉语语音信号分析平台.....盛 晔 张 凯 严普强 (31)
- 2.6 汉语普通话大词汇语音数据库、参数库及其管理系统
.....陈 韬 孙金城 李 彤 莫福源 李昌立 (36)
- 2.7 人耳听觉感知知识用于语音识别的研究.....高 升 吴善培 (40)
- 2.8 基于多段 VQ 的 HMM 模型的语音识别系统
.....李向吉 刘承玺 王秉钧 (44)
- 2.9 音素分类用函数链网的训练和结构优化.....孟宪军 徐近需 (48)
- 2.10 基于最小误识率准则建立声韵母连接模型.....李功俊 黄泰翼 (52)
- 2.11 特征参数分维处理的语音识别研究.....袁中选 徐柏龄 余崇智 (56)
- 2.12 一种基于模糊统计的语音识别算法.....李胜利 侯朝焕 (60)
- 2.13 语音信号参数表示方法对语音识别性能影响的研究姚磊.....黄泰翼 (64)
- 2.14 基于动态变化信息的汉语连续语音的切分方法张凯 ... 盛 晔 严普强 (68)
- 2.15 特定人无调音节汉语语音识别系统.....王向军 胡起秀 吴文虎 (72)
- 2.16 一种基于特征帧和时长信息的神经网络快速语音识别新方法
.....朱 山 陈道文 (77)
- 2.17 基于声/韵母的上下文有关的声学模型
.....马 斌 徐 波 黄泰翼 张希军 (81)
- 2.18 模糊集方法在汉语孤立字识别中的应用.....宁 飞 宋兴彬 韩 骏 (85)

- 2.19 高噪声环境下语音识别技术研究.....韩 兵 易克初 (90)
- 2.20 基于 SUN 工作站的汉语大词汇语音识别系统
.....秦 勇 莫福源 李昌立 (94)
- 2.21 基于知识的有序码书 RQ 用于汉语声母识别.....杨 兵 (98)
- 2.22 汉语全音节实时识别系统研究.....王安良 易克初 程 俊 (334)
- 2.23 播音专业大学生噪音的多维声学分析.....鲍怀翘 (345)
- 2.24 噪声下大词汇连续语音识别的实现.....张 颖 (346)
- 2.25 大词汇量连续语音实时理解系统的研究和构造
.....王 刚 马 斌 高 强 黄泰翼 (347)

三、语音编码与合成

- 3.1 两种提高矢量量化器性能的设计方法
.....王惠丽 唐朝京 陈善思 (102)
- 3.2 G. 728 语音压缩算法应用研究.....易 波 林嘉宇 唐朝京 (106)
- 3.3 几种中低速语音编码器的音质评价实验
.....张知易 王 瑛 胡 昇 刘智广 (109)
- 3.4 PFS-1016 CELP 声码器的研究.....王炳锡 贺英华 (114)
- 3.5 基于小波包的语音信号压缩方法
.....何冬梅 沙宗先 刘寿民 秦 兵 (119)
- 3.6 语音信号的 MPLPC 参数矢量量化.....张晓东 杜 凯 (123)
- 3.7 语音多带激励模型的研究.....李 强 孙允恭 (129)
- 3.8 国内外语音合成技术的发展概况.....吕士楠 初 敏 李晓光 (133)
- 3.9 汉语识别合成型声码器研究进展.....易克初 程 俊 李 勇 (142)
- 3.10 用于汉语规则合成的波形叠加算法的研究.....陶建华 华一满 (146)
- 3.11 用线性预测法实现气声语音的重建.....刘 莹 李国锋 (150)
- 3.12 疑问句语调模型的研究
.....陆亚民 吕士楠 初 敏 贺 琳 (154)
- 3.13 自然语调的汉字语音合成初探.....张 波 朱思俞 (158)
- 3.14 基音同步的发音速度修改及标注的研究.....罗 恒 蔡莲红 汪 泳 (162)
- 3.15 普通话儿化和轻声研究及其合成试验
.....曹剑芬 颜景助 李爱军 华 武 (166)
- 3.16 英语语音合成过程中的单词正文→音素转换符号系统.....郑全录 (171)

- 3.17 声韵母混合编码的合成新方案.....李 彤 莫福源 陈 韬 (177)
- 3.18 汉语文—语转换系.....张 璞 易克初 程 俊 (348)
- 3.19 汉语音节间的协同发音在听觉感知中的作用
.....初 敏 司宏岩 田旭青 吕士楠 孔江平 (349)
- 3.20 中性语调语音合成的音高音长规则
.....倪 宏 孙金城 李昌立 莫福源 (350)
- 3.21 中性语调初探.....孙金城 倪 宏 何国健 (351)
- 3.22 一种用于改进语时编码器的听觉模型.....成立新 陈显治 (352)

四、图像处理

- 4.1 基于 FACS 的三维面部模型的建立及表情合成
.....姜艳辉 郭 峰 文成义 (181)
- 4.2 子波变换图象处理综述.....李 刚 王曙光 (188)
- 4.3 彩色感知分量检验模型.....邓海明 马宝华 于庆魁 (193)
- 4.4 H. 261 图象压缩算法在 ADSP—21020 上的实现.....石 伟 李昌立 (197)
- 4.5 部分畸变或遮挡的二维物体形状识别.....石旭刚 钱忠良 (201)
- 4.6 商用字符串识别的预处理方法.....李晓昆 梁德群 刘春阳 (205)
- 4.7 基于 DT 的人类面部基本特征的二维模型
.....卢朝阳 周幸妮 吴成柯 (211)
- 4.8 自适应 BFT 分形图象压缩.....樊国良 周利华 (216)
- 4.9 现代图象通信中关键技术的实现.....陶 凌 郁乃杰 张 煦 (220)
- 4.10 子波域中彩色图像编码的运动补偿研究
.....万建伟 皇甫堪 周良柱 梁甸农 (224)
- 4.11 图象自适应无失真压缩中的概率估计.....江黔夫 徐善驾 陈学淦 (228)
- 4.12 小波变换在图像分形编码中的应用.....王大凯 陈 炜 (232)
- 4.13 运用子波变换的乳腺 X 光片图象处理.....朱秋煜 莫玉龙 杜晓红 (237)
- 4.14 基于空间信息及灰度信息的聚类图像分割
.....裴继红 谢维信 王大勇 韩 琦 (241)
- 4.15 参数自适应图像压缩编码.....程存学 (245)
- 4.16 子波变换与局部化的图象不变性描述.....梁毅军 贺朋令 蔡元龙 (249)
- 4.17 分形自适应 PIFS 算法研究.....周立红 平西建 邵美珍 (253)
- 4.18 分数布朗运动及其在图象处理中的应用.....谢文录 谢维信 (257)
- 4.19 不变性目标识别研究.....万里青 赵荣椿 孙隆和 (261)

- 4.20 定量金相图像分析及图像数据库系统.....石 锐 李志敏 秦国友 (266)
- 4.21 人工视网膜中的神经网络学习算法
.....陈振湘 纪安妮 许志庚 郭东辉 (270)
- 4.22 子分层维纳滤波算法的子波改进.....李 刚 王曙光 (275)
- 4.23 CK—AVP—A 声音图像开发平台
.....陈 丹 平 胡 科 何国建 李昌立 (279)
- 4.24 多媒体通信中 MPEG 视频压缩技术的研究.....王春宁 (284)
- 4.25 一个图像序列运动估值的新算法.....邱光华 侯朝焕 (341)
- 4.26 遗传优化块匹配运动估值算法研究.....檀喜敬 平西建 李冬海 (361)
- 4.27 随机分形在三维表面重建中的应用.....程 辉 田金文 柳 健 (355)
- 4.28 具有主观视觉特性的子波系数量化.....余 越 邓元木 柯有安 (356)
- 4.29 分形技术在计算机仿真自然景物中的应用.....邓元木 柯有安 (358)

五、通信与信号处理

- 5.1 TMS320C40 信号处理系统主机接口的 FPGA 设计实现
.....王宝利 李昌立 关定华 (288)
- 5.2 复倒谱中相位卷绕问题的讨论.....李国锋 (292)
- 5.3 受加性噪声污染的信号增强和提取.....张颖辉 (296)
- 5.4 信号多尺度边缘重建算法的改进.....朱秋煜 韩锦成 莫玉龙 (301)
- 5.5 采用 SVE—NET 的时间序列预测方法.....江永红 徐秉铮 (305)
- 5.6 基于 BP 神经网络的弱信号提取方法研究
.....杨建华 侯 宏 孙进才 (312)
- 5.7 基于新的“判决—校正”方法的自适应均衡器.....冀 显 (316)
- 5.8 基于神经网络的 DS/SS 信号的扩频码估计
.....李鸿彬 夏 原 郭宗祥 李 杰 (321)
- 5.9 八维 TCM 译码与自适应均衡相结合的研究.....王 英 刘 康 (325)
- 5.10 复相角神经网络的存储容量分析
.....郑金成 叶建伟 陈振湘 刘瑞堂 (330)
- 5.12 一种改进的时频分析方法.....陈建锋 (359)
- 5.13 小波分析及其在语音置乱中的应用...王跃科 王湘祁 杨 俊 黄 锐 (360)

三维影像的研究及应用

袁保宗 唐晓芳

北方交通大学信息科学研究所

北京 100044

[摘要] 本文介绍作者于一九九五年初应邀于法国里昂国立应用科学学院工作期间,访问法国几个著名研究院和学院时,所了解到的关于三维影像技术的研究和应用情况。文中除了介绍一些典型的应用之外,还涉及三维影像的一些基本研究方向。鉴于在科学技术进入高度信息化的时代,高速计算机和宽带数字网络传输的发展,三维影像技术必然在近期研究中占极其重要的地位,此稿的目的主要是与读者共同了解该研究方向上若干热点课题。

一、引言

最近一年来在信息科学技术方面,有一大批专题热点题目: 1、信息高速公路(N I I); 2、多媒体技术; 3、Internet 国际互联网; 4、交互式电视、点播电视、高清晰度电视; 5、可视化技术; 6、虚拟现实; 7、神经计算机等,这些专题分别在不同的场合已经有过不少介绍。我今天的报告内容仅想限于一个比较专门但又是基础性、应用面很广的新课题,那就是三维影像技术的研究和应用。

影像技术,主要是图象技术,其中也包括了它与图形技术的结合。它所解决的问题都是与视觉密切联系着的,企图使机器能帮助人或部分代替人来处理视觉信息,达到人们更充分更便利地利用这些信息,达到各自的目的。由于近几年亿级运算速度的计算机、千兆级的存贮容量,以及几十兆级的宽带通信网等新技术迅速推广,使许多过去难于实现的三维影像技术,都日趋成熟,跨入实用化阶段。从而使我们从事图象和信息技术应用的人们不得不十分注视这一新技术的发展及推广。

就以上的专题,我们利用一九九五年初应法国里昂国立应用科学学院(INSA-LYON)副校长 R.Goutte 教授的邀请讲学机会,访问了法国信息与自动化研究所(INRIA) LIFIA分院、SOPHA-ANTIPOLAS分院、巴黎国立高等通信学院(INST)以及尼斯大学信号与系统实验(I3S),会见了几位法国著名的教授如: O.Faugeras, R.Mohr, H.Maitre, M.Barland, R.Goutte, G.Gimenez, I.Magnin等,收集了一些他们的研究工作。下面的介绍亦是以这次了解的情况为主。虽然不是很全面,但是亦能反映有关的主要方向。

二、三维影像的应用

我们从应用开始介绍,目的是说明这些研究大都从应用驱动的。

1、三维数据场的可视化

三维数据数量大,形态复杂,而且常常是随时间变化的,如果没有可视化技术,就难以被人们直观地加以理解和利用。

· 水声应用: 地质结构的三维可视化早有应用,但是在水声、特别在海洋声学方面利

用可视化还是比较少。在海洋声学方面,一方面通过对发射水声信号的时频分析,可以反演出海底地形分布,利用计算得到的海底地形等高线分布,即可以通过三维可视化,将地形的高度实体、分布图演示出来。另外,通过超声声源,及阵列接收,可以计算出海洋海水中声速的分布,通过对声速的分布,可以推算得到海水温度的分布。由计算获得这些三维分布,可以用图形学方法,将三维速度和温度分布,用体素方法直观地显示出来。这种显示对研究海洋对大气及地球的研究非常重要。

心脏和大脑三维数据的显示:大脑及心脏是当前医学中最主要的两种研究对象。对它们的状态的详细分析及对照,对于诊断和治疗各种脑血管病具有十分重要的意义。利用NMR(核磁共振)CT或X光CT可以获取脑和心脏的层析数据,然后将层析数据叠积起来,可以形成立体的关于脑和心脏的三维形态。如果对每一张层析照片,将血管加以分离,那么经过三维处理,可以形成三维血管分布的立体感图象。利用三维立体数据对病照进行定位之后,就可以用于治疗中的正确定位。这种工作,已经进入商品化阶段。近期的研究是通过高速立体CT,采用动态空域重建(DSR),可以对(实验是用狗做的)左心室进行动态重建。汉堡大学医学和计算科学研究所和里昂应用科学研究所合作,在Thomson-CGR磁扫描500核磁共振仪上,采用与心电图QRS波同步地录取一个心跳周期内的约10幅三维心脏数据,它们分别有10mm间隔,平面分辨率为1.64mm的层析图像组成。通过对这种数据的重建,可以对左心室外壁的运动位移和加速度测得,从而用于临床诊断。

立体B超:通常的B超只能看到断层反射波图象,当然也可以通过断层叠加,构成三维图象。但是这种显示方法,我们只能看到某种剖面,或投影。真正的立体B超,目前尚未见到过。里昂应用技术研究所建立了一种实验的立体B超方法,可以通过佩戴眼镜看到立体的B超图象。它的原理是将显示的图象由两幅视图组成,左眼仅能看到左视图,右眼仅能看到右视图。这样当双眼同时看时,在感觉上就可以看到立体的B超图象。利用以上原理,在实验室内已在模型组织上做了显示试验。

2、三维图象的分割及分类

对于任何图象的处理,分割和分类往往是必然要进行的!这是因为:如果以解剖学为目的,那么首先要对不同组织进行分割,然后再进行分类。如果作为诊断使用,那么正常组织和病变组织的分割及分类也是必要的。下面主要介绍两种三维数据的分割:

三维心脏血管的分割及显示:在里昂国家保健与医疗研究所(INSERM)的附属医院看到利用国际仅有的由通用电气公司生产的试验立体X光CT设备,称为Mophometer。利用该CT,获取的三维心脏图象,通过对其中心血管的分割分离,然后通过三维重建,在工作站上显示的三维心血管分布。由于这台仪器还属于试验机,未能对人的心脏进行试用,但是动物试验正在进行。

乳腺癌的自动分类:妇女乳腺的普查,以及乳腺癌的早期诊断,一直是医学界的一件大事。由于乳腺的医学图象灰度变化较小,并且纹理非常密集,对其癌变组织的分割比较困难。现在正在研究分割效果更好的算法以及三维的边缘检测算子,用于这类图象的分割。

3、三维物体的重建

三维重建是指对客观世界的物体,通过图象处理的办法,在计算机屏幕上将其立体形态,重新重建和显示出来,并且能使其作一定的位移或旋转。这种技术有着广泛的应用领

域,除了医疗之外,在工业制造,工程建筑,文物保护,军事侦察等方面都有大量应用前景。

· 博物馆古代雕塑的重建: 法国巴黎卢浮宫内珍藏了古代无数稀少雕塑。巴黎国家高等通信学院有一个三维数据采集实验室,该室拥有一套高精度三维数据采集系统。该室置于地下室,有防震的激光平台,所以获取很精密的三维数据,用此设备,在实验室内已复制了多个古希腊塑像的腊制人头。

· 珍贵稀缺动物的重建: 巴黎动物园内有一种非常奇异的动物,它的头部类似鸭类,但身部呈扁平扇形,无脚,全身羽毛。他们将此动物先麻醉之后,放在三维数据采集仪下,然后再在屏幕上将此小动物重建出来。当然这时可以让此动物呈任意姿态。

· 心脏搏动时的三维序列重建: 根据心脏外壁的三维数据,利用序列图象的原理,将其跳动的立体形态,在屏幕上重建出来,供医生诊断时参考。

· 大型建筑物的重建: 近年来由于计算机影视制作的兴起,一切现实环境下的大型建筑需要制入影片之中,参与拍摄。如何通过比较简便的方法,利用普通摄像机,即可录取其三维外形数据,再进行重建,已经成为三维计算机视觉研究中的一个新方向。这方面欧洲的研究有较大进展。目前已经有一些街道,桥梁及古典宫殿被重建而成。

4、机器人的景物理解

· 自动装配是机器人视觉研究中一个研究课题。一个带有手眼系统的机器人可以通过立体视觉,抓取放在空间不确定位置的工件。此项研究是在 INRIA - SOPHIA - ANTIPOLIS 的题为 PRISME 进行的。

· 自动走行车。INRIA, ISIA (巴黎矿冶学院) 和 I3S 实验与 ECA 公司合作正研制一种利用计算机视觉引导的,可以在结构化的环境下自动移动的小车,它们计划称为 ICARE。

· 虚拟手术。此项应用研究称为医学影像与机器人,它的目标是通过将 CT, 核磁共振 CT, 超声回波等图象的处理,再经过多模分析的信息融合,变形运动分析,诊断参量的自动提取,然后与医用机器人结合,可使先用虚拟现实的技术进行虚拟手术模拟。

5、主动视觉

过去开展的计算机视觉研究,由于摄像机参数都是固定的,所能觉察到的视野,常常也是固定的,因而可以称为被动视觉。近几年开始注意主动视觉的研究。这种系统,视觉系统可以根据外界景物中所感兴趣的部分来自动调节方向、倾角、聚焦距离以及视场大小。欧共体正在开展的 VIV 项目,已经提供一种可以灵活跟踪兴趣对象,同时还具有实时对简易景物理解的功能,具有友好的人机交互界面。

三、三维影像的有关技术

贯串以上三维影像应用的关键技术,可以归纳如下:

1、三维数据采集:

数据采集是第一位的要素。没有采集就无法实际获得应用,至今常用方法有:

结构光方法: 这是采用激光光面照于物体表面,通过光平面与物体表面的交界来计

算得到表面的三维数据。这是目前最常用的方法。也是最精确的方法之一。但采集速度比较慢,代价比较高。

·CT层析技术:这是医用三维图象的常用方法。

·新一代立体CT及立体超声:这种机器仅在实验之中,前面介绍了的Mophometer就是一例。

·普通摄像机采取的序列图象方法。这是一种新的方法,基础是立体视觉原理,但不只取两幅,而是许多幅组成的序列图象,从多幅图象中找出对应点,然后计算三维数据。

2、数据压缩:

图象分析进入三维之后,数据压缩是势在必行。特别是医用图象,医生对图象的质量提出了严格的要求,不允许经压缩后复原的图象有明显的质量损失。从而对数据压缩提出了新的要求。这次访问中看到所研究的新方法中主要有以下几种:

·3D子波变换 加矢量量化

·3D子波变换

·自适应余弦变换

看来这方面还有许多工作可做。

3、立体匹配:

立体匹配一直是视觉研究中的一个难点,为了得到较精确的三维坐标,一般对摄像机的内外参数进行计算并定标。这样的定标过程限制了立体视觉的运用。在一般场合,使用一般商用摄像机,在比较任意的条件下拍摄的立体图象,是否亦能获得三维数据,并开展重建工作呢?正是这种需要,近几年在INRIA研究所开展了所谓使用非定标镜头的立体视觉,同时利用物体的几何不变量来开展物体的识别研究。

·非定标镜头的立体匹配

法国学者 R. Mohr, O. Faugeras, 日本学者 Kanate 和 荷兰学者 Van Doorn 在这方面都做出过开创性的工作。它们的基本原理是通过对两幅或多幅在不同位置和角度用非校正镜头拍摄的图象上给定相当数量(4点到6点)的已知对应点,然后,通过镜头的投影模型公式,用最小平方估计的方法求出有关的参数。他们的做法是利用摄像机在物体周围连续拍摄几十到上百幅序列图象。由于序列图象之间的变化很小,因而对应点的问题可以简化为点的跟踪来加以解决。然后从此序列的对应点中选择所需参数与计算的四至六个点作为参考点,用以求取镜头投影模型的有关参数。利用这种方法已经给出了多种实验结果,说明方法是有效的。

·充分利用几何不变量

三维视觉中解决物体识别遇到的一个困难问题是“形状随视角变化而变化”。为了求取在不同视角下物体的不变参量已经做过许多研究。近年来利用代数几何,微分几何等手段,利用几何不变量进行视觉研究已经形成专题的研究潮流,已经召开了三次专题学术讨论会。今年四月在我国西安由法方和中方联合发起,成功地召开了“计算机视觉中几何建模和不变量欧-中研讨会”(GMICV'95),欧美著名的学者大多到了会。论文集反映了目前国际的学术水平。

4、图象解释:

图象解释是三维图象工作的基础,一般还对二维图象进行。这里想介绍一些新的办法:

- 时频分析方法: 时频分析用水平声信号分析以及医学图象分析已经见到一定成效。
- 高阶谱分析: 对医学图象的分割,利用高阶谱的特征可以得到更良好的边缘特性。
- 多尺度分析方法亦在图象的预处理方面得到应用。

5、三维形状的建模:

无论从造型、识别来说,对三维表面的闭合实体进行建模都是十分必要。这里专门介绍的一种建模方法是采用超二次曲面 (Superquadric) 对左心室进行建模的应用。超二次曲面是近期由美国学者 Pantland 所建议的能很灵活地表现闭合曲面的方法。这种方法的曲面表达能力很强,使用参数数目较少,特别适合于利用参数进行建模,并从而进行物体识别的场合。用超二次曲面描述左心室也是为了医学诊断。这种曲面表达方式将在图形学方面获得广泛应用。这里顺便提一下,我们研究所已有一位博士完成一个由超二次曲面对表面进行建模描述,并可以组成各种外形实体的三维图形和动画生成软件包。它运行在 SUN SPARC 工作站,可以提供实际应用。

6、数据和信息融合技术:

· 数据融合: 由于噪声的影响,图象的边缘分割往往不能达到予期的要求。采用多种方法对同一图象的融合处理,可以提高边缘正确的。

· 双眼立体视觉的融合: 近来有一种三维图形的视觉方法,即一张计算机产生的二维图案,当你用双眼凝视该图时,经过一小段时间,在你的视野中该图案即显象为一幅三维立体图案。这一双眼立体感的数据融合过程在人的视觉器官和大脑中是如何形成的呢?目前正在立题,从生理及心理以及图形学多方面来综合加以研究。它的研究成果可能有得于三维影像技术的新发展。

四、结束语

以上就是“三维影像技术的应用及研究”,非常不全面地介绍一些见闻。由于它发展的迅速和广泛的应用前景,特别提出来与读者共同加以注意,愿更多的研究者及开发者在领域做出更多的工作。

从汉语的结构特点看汉语语音识别合成研究

易克初

张家驷

西安电子科技大学

中国科学院声学研究所(100080)

ISN 国家重点实验室(710071)

摘要:本文用信息论观点分析汉语的结构特点、应用上的优势和弱点,探讨汉语识别合成技术研究的策略和发展前景。

1. 引言

语言是人类交流信息的工具,也就是人与人之间进行通信的手段。从通信系统和信源编码理论的角度来分析汉语的结构,有利于更加深刻地认识汉语语言的根本特点。通过这一分析我们发现:汉语是一种结构独特、应用上优势很强的语言,但它也存在一些弱点,这就是汉字的结构比较复杂,因而初学汉字书写或计算机录入都比较困难。近年来发展起来的汉字编码输入方法、汉字自动识别技术、汉语语音识别与合成技术为弥补汉语在应用上的弱点提供了有效的手段。特别是汉语语音识别与合成技术的研究,将为最广大的群众(包括非汉语母语的使用者、不经常阅读和写作的人、存在听觉或视觉障碍的人等)提供一种能方便地运用汉语进行听、说、读、写的手段。

汉语语音系统的结构非常科学,有利于计算机自动语音识别和语音合成的实现,特别是比较容易实现无限词汇的语音识别与合成,因此这方面的研究具有特别重要的意义。我们应该充分认识汉语的特点和主要应用需求,并以此为依据来考虑我们的研究策略,制订我们的研究规划。

2. 用信源编码理论分析汉语的结构特点

用语音或文字交流信息,从信源编码理论的角度来看,都可以看作是一种离散信源符号序列的编码和译码过程,其编码端和译码端都有一个共知的编码符号集,或称基元集。对于印欧语系来讲,这种基元集就是音素或字母。这两种符号集的大小都只有几十个符号,一个是语音的基本单位,一个是文字的基本单位。

汉语的结构比较独特,其语音和文字都拥有一个非常庞大的基元集,一个是1300个左右的汉语音节,一个是16,000个左右的汉字。

为什么不把汉字的笔划和汉语音节中的声韵调看作基元而要把汉语音节和汉字看作基元呢?这要从通信系统中编码符号集应该满足的两个基本条件来理解^[1,2]:①它必须是发方容易产生、收方容易辨识的有限个符号的集合;②它必须是能用于以串行分组形式构成内容不受限制的信息段,并且这样构成的信息段是在收端容易译码的。显然,汉语音节和汉字都是符合这两个条件的。以汉字或音节为基元可以构成无限的词语和句子,国标一至四级字表中16000个左右的汉字或汉语普通话中1300个左右的有调音节都是一个相当稳定的有限

个元素的集合。由笔划构成汉字和由声韵调组成音节都不是简单的串行组合,而是根据一些特殊规则构成的。正因为每个汉字或音节的构成存在一些特殊规则,使它们更加便于产生、辨识和记忆,更适于充当基元。汉字和音节这些特点与其它语言中的词汇的性质是不同的,而与印欧语中的字母和音素的一样可以看作是基元。

为什么汉语能形成如此巨大的编码符号集而仍能十分方便地使用呢?这主要是因为汉语较充分地开发了文字和语音这两种信源的潜力。文字是一种二维信息,人类要产生和快速辨认许多种较复杂的二维图形并不困难,因为人类视觉系统是一种高度并行的信息处理系统,它处理图像或图形的速度极快。据测试和估算^[2],人类看电视时平均每秒钟接受的信息量是 2×10^4 比特,这比阅读英文资料的处理速度(100 比特/秒)要高出 200 倍。汉语用较复杂方块形汉字代替简单的字符,就是利用了视觉系统并行信息处理的潜力。

至于汉语音节的结构就更是奇妙无比了^[3]。一般地说,构成一个音节,其中至少要有一个响音,这个响音通常是元音。除了这个响音之外,前后都可加上各种辅音,因此由几十个音素可以构成上万种不同的音节,英语就有一万多种音节。汉语则只从这些音节中挑选出四百多种,这些音节可以很有规则地排列在一个 23×35 的二维阵列中,同时又以调频方式给每个音节赋以四种声调的变化,然后删除其中少数发音或听辨不大方便的音节,得到 1300 多种有调音节。虽然每个有调音节都是由声母、韵母和声调紧密结合的一个整体,然而人类听觉系统却是很容易将三者分离的。因此每个有调音节可以看成是可分离的 $23 \times 34 \times 4$ 的三维离散空间中的一个元素。要从这 1300 个音节中找到某一个,只需逐维地在 23, 34 和 4 个元素中搜索,因而其发音、记忆、听辨都十分方便,用计算机进行自动语音识别也很方便。

根据信息论中信源熵的计算公式可知^[1,2],一个编码符号集大小为 N 的信源,平均每个信源符号所携带的信息量(即信源熵)为:

$$H(s) = - \sum_{i=1}^N P_i \log_2 P_i$$

比特,其中 P_i 为第 i 个信源符号的出现概率。假若各符号等概率出现,那么其信源熵为 $H(s) = \log_2 N$ 比特。可见信源符号集越大,信源熵一般也就越高。现在我们根据《现代汉语频率词典》^[4]所提供的统计数据来计算一下汉字信源和汉语音节信源的熵。将该词典中统计出现频率最高的 4574 个汉字及其发音(音节)分别作为信源符号集,算得汉字信源的熵为 9.36649 比特,相当于 660 个汉字等概率出现时的信源熵;汉语音节信源的熵为 8.50000 比特,相当于 362 个音节等概率出现时的信源熵。这当然比印欧语中字母或音素的信源熵高得多。我们用外推法进一步求 16000 个汉字构成的信源的熵,结果是 9.52996 比特,相当于 739 个等概率出现的汉字信源的熵。

从这些计算结果我们发现两个问题:第一,汉字虽然比汉语音节数量上多得多,但二者的信源熵却相差很少。其原因是汉语音节的概率分布比汉字概率分布要均匀得多,这是因为高频字的同音字通常比低频字同音字少一些而产生的后果。第二个问题是汉字字表增大,其信源熵增加不多,这是因为频率很低的字,虽然每个字的信息量非常大,但因其出现概率很低,对信源熵的贡献却很小。

当然,汉字信源熵不像人们想像的那么高,不会影响汉字的载信能力,因为低频字携带的信息量非常大,学问较高的人可以较多地选用低频字,就可以用非常简洁的文字表达含义丰富的内容。不同文化层次的人可以选用不同信息含量的文字进行信息交流,灵活性很大。

汉字和汉语音节数量虽然很多,但其信源熵却并不是非常高,这不仅不是缺点,反而是一个重要优点,因为熵越低,每收到一个信源符号时的疑义度就越低,就越容易辨识。从以上的计算和分析结果可以看出,汉语语音的听辨和汉语文本阅读过程中所需消除的疑义度的平均值是基本相似的,都只相当于从几百个等概率出现的元素中辨认出其中的一个,这是人类的感知能力可以胜任的。加之汉字的象形特性和汉语音节结构的规则性,进一步改善了其辨识效率。因此,汉语阅读和听辨都是不难学习的。

众所周知,信源编码符号集越大,其编码效率也就越高。汉语显然是全世界 3500 种语言编码效率最高者之一。我们所熟悉的许多种语言中没有一种像汉语那样拥有如此巨大的基元集。例如:如果用 16,000 个汉字排列组合构成二字词,就可以构成 2.56 亿个词,构成三字词、四字词的种类就更是多得惊人了,就连不同音的三字词也可达 20 多亿个。

总之,汉语拥有十分巨大的基元集(汉字和音节),因而具有极高的编码效率;另一方面汉字信源熵和汉语音节信源熵不是太高,因而辨识效率也很高。这是汉语结构上的根本特点。

3. 汉语的优势和弱点:

上节用信源编码理论分析了汉语的结构特点,下面我们从应用角度来讨论汉语的优势和弱点。

汉语在应用上具有五大优势:1)组词能力强、缩写非常方便,语言表现力极强;每个汉字常常就是一个独立的词,同时又可兼作词根,前缀或后缀,因而创造新词或猜测新词的含义都十分方便。例如:在“葡萄”后面加一字就可以派生出许多词,葡萄干、葡萄酒、葡萄园、葡萄糖等等词意十会明显,而英语中相应的词是 grape, raisin, wine, vineyard, glucose 等,对于后者我们很难从词形或发音上猜出它们是出于同一词源的。所以英语已有的 70 万词汇未必穷尽了世界上的万事万物,而汉语的 16000 个汉字只要组成四个字以下的词就可达到 10^6 的数量级,足以用简短而形象的词语表达无穷无尽的事物。

汉语还便于将一个很长的专有名词或含义丰富的短语缩写成两三个汉字,例如:“西电”、“打假”,可以把一个很长的故事用一个四字成语表达得生动形象、一目了然,甚至可以单用三字句、四字句、五字句或七字句写成本本读起来朗朗上口的书。这种我们习以为常的奇迹,实际上正是汉语优势之所在。

2) 汉语的文字和语音为文艺创作和发展提供了一个理想的平台:由于每个汉语音节都以舒展的响音结尾,都可以随意延长,因此便于歌唱,便于创作富于音乐感的诗歌和顺口溜,相声演员常常也可利用汉语的特点做点有趣的文字游戏。汉字的图形特征就是汉语书法艺术能引人入胜、代代相传的基础。

3) 汉语的基本发音规则简单、发音方法易学、好记,发音清晰响亮容易听辨理解。

4) 以音节为基元可以实现无限词汇的汉语语音识别与合成,这对于汉语听写机和汉语识别合成型声码器的研制和推广应用是非常有利的。人们早就认识到语音压缩的率失真下界约为 50 比特/秒,这就是利用识别与合成技术进行语音编码^[2,5,6]。但是,到目前为止,尚未见到其它语言已实现了这种声码器的报导,而无限词汇的汉语识别合成型声码器已初步研制成功^[7,8],正是利用了汉语语音结构的特点。

5) 汉字便于永久记忆,因而没有失读症患者:人脑存贮图形的容量是非常巨大的,几十年前见过的人的面孔也能一下就回忆出来,何况一万多个形状各异,含义形象的汉字呢?相比之下,拉丁文字的单词是一些串行编码的码组,它在大脑中的存贮部位不同于汉字,必须结合读音才容易记住。不难想象,“哑巴英语”很难学得精通,而哑巴学习汉语阅读却是不难做到熟练自如的。

汉语与其它语言相比,在应用上有许多明显的优势,但是汉语也有弱点,其主要弱点是:学会书写或机器录入比较困难,特别是那些非汉语母语的学习者要学会书写汉字是十分困难的,即使是有一定文化素养的中国人,如果不是经常写文字材料,也很难做到准确无误、快速完美地进行写作。但是,汉语的上述弱点完全可以应用语音识别与合成技术来弥补。

4. 汉语语音识别与合成研究的发展前景:

从上节汉语的优势和弱点分析,我们已可看出汉语识别合成研究具有特别重要的意义和十分广阔的应用前景。随着汉语听写机技术上的突破,人们对听写机是否有推广应用的价值的疑惑很快就会有消除,汉语识别合成型声码器也会随着多媒体终端一起进入千家万户,它将以约百分之十于电话信道费用的信道费用为人们提供准实时的双向语音通信或非实时的语音邮件服务。

人类已进入信息化社会,未来的通信方式必然是多种媒体并存和互相补充的关系。电话、传真、电子邮件、语音邮箱、BP机传呼等,都有它们各自的特点和发展余地。由于汉语具有如上所述的一些独特的性质,未来中国社会的通信方式也必定会有一些与众不同之点。其中语音识别合成技术将会在其中起十分重要的作用,例如:电子邮件中的汉字输入要借助语音识别;无限词汇的汉语识别合成型声码器,由于它具有极高的数据压缩比,将为语音通信提供一种效率极高的手段。

语音识别合成技术在汉字信息处理、办公自动化、数据库自动查询服务、汉语语言学习、辅助阅读等方面也大有用武之地,其应用潜力是众所周知的。

目前的主要问题是还有些关键技术有待突破,就是运用已公开的技术构成一个基本好用的系统,也还有不少工作要做。

5. 关于汉语语音识别与合成研究的策略

汉语的结构不同于英语。对于汉语来讲,解决了全部汉语音节的识别就可以实现无限词汇的识别。加之汉语音节在语流中的独立性较英语中音素的独立性要大得多,因此前几年大家的研究的重点放在全汉语孤立音节的识别上,是可以理解的,也是必要的,因为单音节词在汉语语音中的出现概率就有20%左右,孤立音节的识别是回避不了的。况且汉语全音节识别是检验我们的语音识别系统中关键技术是否过关的最好手段。目前,非特定人孤立音节的识别还没解决好,连续词语中全音节识别的问题也没有解决好。这些问题,对于汉语识别合成型声码器的研究是特别重要的。因此,在今后很长一段时期中汉语全音节识别技术的研究仍然有重要意义。可以说,将来真正好用的汉语听写机和汉语识别合成型声码器的成功,其中的关键技术多半会产生于汉语全音节识别技术的研究。

当然,现在人们从实践中认识到:基于孤立音节识别的汉语听写机不受用户欢迎,要想