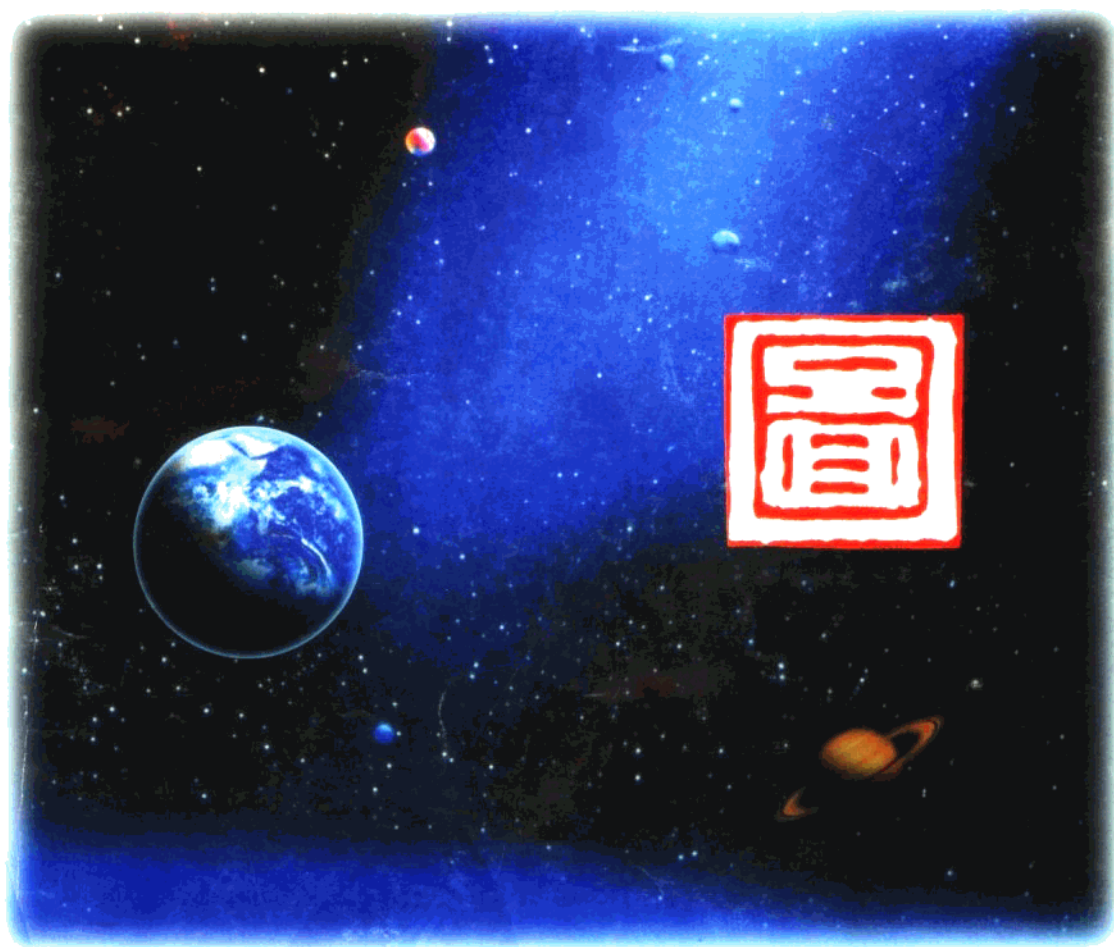


中国图象图形科学技术新进展

第九届全国图象图形科技大会论文集

9th National Conference on Image and Graphics

NCIG ' 98



中国图象图形学学会

1998年5月11日~13日 西安

中国图象图形科学技术新进展

第九届全国图象图形科技大会论文集

9th National Conference on Image and Graphics

NCIG'98

中国图象图形学学会

1998年5月11日~13日 西安

目 录

图像处理

基于小波的彩色图象匹配和整合的层次算法.....	徐 丹	石教英	鲍 歌(001)
基于内容的多媒体信息检索 CMIR 与国际标准 MPEG-7			章毓晋(009)
多模式自适应图象实时压缩系统.....	程子敬	周孝宽	周付根 王 飞(017)
熵编码在图象压缩中的应用.....	闫宇松	王春梅	石青云(022)
提高遥感图象分辨率的空域方法研究.....	卢 峰	葛成辉	史久浩(025)
参数识别法的消大气模糊复原.....	陈书海	朱国富	陈 华 安 斌 王宝慧 冯保红(028)
一种高效图象正交变换编码:DDD-T 及其应用.....			金东瀚 郑 雷(032)
基于二维子带分解方法的图象边缘检测.....		张 宇	郝重阳(036)
基于小波变换的图象无损压缩方法.....	朱桂斌	吴乐华	张海呈(040)
遥感图象自适应容错格式同步器的研究与实现.....	吴松平	葛成辉	谭耀麟(043)
基于 PIM 自适应量化编码算法	肖永隆	柳葆芳	平西建(047)
基于遗传算法的图象重建算法.....	伍晓平	谷士文	费耀平 刘应龙 李 杰(049)
一个改进的 AWBS 二值图象压缩算法及实现			杨洪兵 刘立柱(052)
基于二次规划的神经网络遥感图象的恢复.....	王耀南	王绍源	孙 炜(055)
指纹图象的自动识别.....	宋建中	陆 颖	朱 进(058)
肤色图象的颜色特性.....	金卫红	张良培	赵国传(061)
工业检测图象信号的异步传输.....		王柏艳	高毓麟(064)
最佳二维快速 DCT 算法:N-FDCT		金东瀚	郑 雷(067)
噪声图象的广义形态滤波.....			赵春晖(070)
一种基于自组织神经网络进行矢量量化图象编码的快速算法.....	王学军		陈贺新(075)
缸孔表面缺陷图象的分段检测.....	赵宏宇	赵红颖	王 楠(077)
无损探伤图象的采集与显示.....		和志强	高毓麟(080)
基于两组多层神经网络的电容层析成像研究.....	肖 化	胡广莉	何惠玲 保宗悌(083)
实时静态图象压缩解压的硬件设计.....	阎晓东		王树元(086)
一种基于哈夫变换的圆形物体检测技术.....	王丰元		孙壮志(089)
一种直接直方图修正方法.....	余日泰		蔡吸礼(093)
多进制小波及其在 DEM 简化中的研究	万 刚		朱长青(096)
一种新的图象压缩编码方法:矢量预测及其应用		郑 雷	金东瀚(100)
用于极低码率可视电话编码的脸部分割技术.....	王辉柏		张春田(103)
编码黑白图象的数字彩色再现技术.....	姜 晶	林 列	方志良 刘福来 母国光(106)
简述现代高效率编码技术.....			金东瀚(165)
序列断层图象的八元树构造.....	束学斌	赵海峰	黄 勇 汪炳权(542)

图象分析与分割

基于图表示方法增强图象分割算法的稳定性.....		罗 为(108)
模糊选择多分辨率 Kohonen 聚类网络用于灰度图象分割	胡世英	周源华(112)
城市航空摄影图像融合分割方法.....	王强军	王炳生(118)

基于多特征图象、模糊聚类的分层分块的纹理分割方法	赵 峰	赵荣椿(121)
两种改进的图象模糊散度阈值化分割算法	薛景浩	章毓晋 林行刚(125)
一种基于矢量量化的纹理匹配分类方法	付新文	张妙兰 李象霖(130)
基于模糊群聚技术的多回波脑部磁共振图象分割	江宝钊	张 森 胡兰青(133)
面向数据采掘的自适应图象分割技术	刘健勤	鲍光淑(137)
Walsh 函数的镜像复制	王能超	谢松法(140)
复杂边缘图象中弱边缘的检测方法研究	周志权	周建林 贾晓光(143)
基于小波的多分辨率贴面算法及实现	赵友兵	潘志庚 石敦英(145)
基于图象结构的门限自动选择及目标提取	王秋让	赵荣椿(150)
基于尺寸特性的图象边缘提取算法	冯晓毅	周凤岐(156)
几种边缘检测方法及其比较	王 岩	勒中鑫(158)
基于灰度秩数的非监控纹理图象分割	毕笃彦	毛柏鑫 马林华(200)
基于分形的红外图象目标自动检测	卢福刚	赵荣椿(204)
精子运动图象的多目标检测与分割	匡 宇	游志胜(209)
一种基于小波变换的红外热图象分割方法	任 彬	汪祖媛 黄 勇 汪炳权(537)
图象分割技术进展	赵荣椿	迟耀斌 朱重光(547)

图形学

基于图象序列分析的表面形状建模和绘制	谈 正	刘 辉(162)
机械图中标注字符粘连问题的解决	钟晓刚	李新友 唐泽圣(171)
基于分形的自然纹理图象生成方法	曹汉强	朱光喜 朱耀庭(176)
基于综合推理的计算机美术图案生成研究	李 宁	潘云鹤(179)
一种基于标记的自动矢量化方法	任金昌	赵荣椿 刘 哲(183)
基于直线段拾取的工程图矢量化方法	王明磊	李新友 王德英(187)
曲线形状的空间结构与谱估计分析方法研究	田 越	平西建(192)
一个大型实时数据自动检测系统图形界面的设计	刘其真 孙旭峰	王宗彩 何永保(195)
基于 RS-485 协议的智能图形图象显示技术	曹延华	赵卫国(212)
仿射不变的多边形相似性度量	刘俊义	王润生(236)
基于 PC 教员台三维动画实时显示研究	周 惠	于海峰(270)
复动力系统的分形集合和具有周期性的分形图象的生成	卢 洵	李世铮(272)
适于图案造型设计的分形表达	贾艾晨	魏小鹏 韩国城(275)
基于分形几何的花型设计方法	姜李玮	许鹤群(278)
Bezier 曲面和 Nurbs 曲面拟合在三维足迹表面重建中的应用	肖 菁 王 琰	平西建 邵美珍(280)

计算机视觉与模式识别

视觉导航中基于运动模型的道路识别算法的研究	苏开娜	任文君 易小琳(215)
雷达和红外成像双传感器信息融合目标识别研究	唐文彬	郝重阳(219)
应用运动矢量预测的自适应运动估计算法	柳葆芳	吕家忠 平西建(223)
一种基于色度信息跟踪移动物体的方法	葛庆平	贺聿江 刘京锐(226)
基于活动原点的仿射匹配方法	封举富	石青云 程民德(230)
三维激光彩色扫描仪中获取物体表面弱反射区深度信息的方法	金 刚	李德华(241)
基于结构光立体视觉恢复三维表面的误差分析	吴东晖	叶秀清(246)
一种复杂背景下基于知识的目标识别方法的研究	任文君	苏开娜(283)
一种图象结构化特征的提取和跟踪方法	贾云得	M. Hebert(287)

一个二维搜索的立体匹配算法.....	贾波	章毓晋	张宁	林行刚(291)
汽车车型图象的自动识别.....	王年	任彬	黄勇	汪炳权(294)
一个基于表观的弧立手势识别系统.....	祝远新	徐光祐	黄浴	林学闯 余志和(320)
双眼立体感知几何模型的研究.....			梁栋	韦穗(362)
一种稳健的人脸检测方法.....	李士进	熊辉	陆建峰	杨静宇(531)

多媒体技术

汕头海湾大桥防作弊支援系统.....				董铁军(244)
在互联网上开发多媒体业务.....	冯玉琨	徐为		侯建军(250)
《图象处理和分析》计算机辅助多媒体教学课件的研制.....	章毓晋	李勛		黄英(254)
网络上基于内容查询的图象数据库系统设计和实现.....			杨翔英	章毓晋(257)
多媒体 CAI 课件中的动画制作技术.....			郭启全	刘舒(261)
基于 CSCW 的多媒体协同编辑系统.....	郭德威	朱光喜		王汉民(265)
多媒体数据库的开发与应用.....				秦长贵(395)
甚低码率多媒体通信与模型基图象编码的研究.....		张明方		蔡德孚(399)
多媒体技术与军事仿真.....				孟章荣(501)
BDHAL——一个通用超媒体系统(BDCard)制作语言的定义及其编译实现.....				
	徐章伟	崔鹏	冯艺东	蒋宇全 董士海(527)

仿真与模拟

建筑工程多媒体仿真的可视界面设计技术.....	张士纲		吴炜煜(298)
在 CGCAI 中实现算法动态执行可视化的一种方法.....		蒋志方	曹磊(302)
时序电路测试中时间及测试序列识别问题研究.....	王仲	陈思成	魏道政(305)
CCD 成像星图的计算机模拟.....	张钧萍	林涛	贾晓光 钱国惠(309)
基于 Internet 的设计信息表示.....	赵继政	刘敏	魏生民 杨彭基(313)
可视化平台上的数据建模和数据封装.....	王健宁	董峰	石敦英(316)
分形学理论在人体三维数据仿真模型建立中的应用研究.....	余轮	周霆	叶东毅(325)
人体肺部器官计算机仿真模型的建立与研究.....	周霆	余轮	叶东毅(328)
用于 VR 的全景图象建模方法及实现.....	罗恒	朱志刚	徐光祐 熊成宇(331)
防化计算机仿真模拟研究.....	何江华	钟鸣	李晓军 丁忠杰(335)
计算机辅助制造中物流系统的动画仿真.....		蒋志方	许振宇(338)
基于计算机动态图形仿真技术的港口生产调度过程优化方法的研究.....			真虹(343)
虚拟现实造型语言 VRML 及其使用方法.....	马小虎	潘志庚	石敦英(350)
基于分形方法的艺术图案生成与设计.....			王小铭(354)
云模型的扩展及其在空间数据发掘中的应用.....	邸凯昌	李德毅	李德仁(521)

三维几何模型与地图

可在水利行业普遍推广的实用地理信息系统 FCWGIS 的研究和开发.....	徐杨	周年生	武靖源(367)
"4D" 技术的土地资源遥感动态监测.....	张继贤	林宗坚	孟文莉 洪炯(372)
等值曲面图的绘制.....			王勇(377)
一种适于实时显示的三维地景动态模型结构.....			廖朵朵(381)
三维立体地形环境仿真动画的制作与应用.....			夏青(386)
利用计算机制图技术进行地图的生产和制作.....	刘海砚		孙群(391)
提高大数据量矢量电子地图显示速度的算法与数据结构的设计与实现.....			陈刚(404)

OpenGL 软件在人体三维数据模型建立中的应用	叶东毅	周 霆	余 轮	(420)
三维实体的体几何模型及其应用	王利生	谈 正	(509)	

应用技术

协同神经网络在汉字图象识别中的应用	赵 同	戚飞虎	(359)	
ACPCS— I 计算机自动旅客计数系统的研制	毕笃彦	毛柏鑫	(408)	
基于微机的数字视频立体测图系统的研究	夏运钧	夏 青	(412)	
工业现场的智能图象分析报警系统	程体德	刘天宇	(417)	
基于 PCI 总线下的汉字识别加速硬件卡的实现	陆庆杭	张彬	盛立东	(423)
微机数字视频监控记录系统	马林华	毕笃彦	毛柏鑫	王晟达 (426)
光密度的由来及其有关问题			汤耀法	(429)
TMS340 系列图形图象加速板在船舶自动驾驶模拟系统中的应用	乐美龙	汪希龄	陆惠生	(433)
基于红外热图象电力设备热故障的自动诊断	孟 坚	任 彬	束学斌	汪炳权 (437)
地图出版系统中的输出控制技术	张 东	陈广学	(441)	
基于 TMS320C80 的 H. 263 运动估计优化算法实现	谭径微	周 璇	潘 健	余松煜 (445)
纸张匀度的计算机视觉检测方法的研究	王亚鹏	许增朴	于德敏	王永强 (450)
NIP1 邻域图象处理机	苏光大	左永荣	尚 焱	(455)
玻纤薄毡缺陷实时在线自动检测系统	谭立球	费耀平	谷士文	刘应龙 李 杰 (460)
装配序列的简化约束表示研究	张建标	魏生民	杨彭基	(463)
实现彩色空间变换的一种硬件设计	王 新	王树元	(466)	
基于图象分割的人工神经网络在两相流层析成像中应用研究	肖 化	胡广莉	何惠玲	保宗伟 (469)
视频点播系统中广播式视频服务器方案	陈丽莎	李岩松	陈国玺	(472)
交互式有线电视系统的实现方案	陈国玺	周 峰	姚庆栋	李岩松 (474)
基于 CL560 的活动 JPEG 实时视频编/解码系统	李文贤	游志胜	(477)	
基于 PCI 总线的微机控制线阵 B 超诊断系统	彭 浩	苏光大	刘 敏	(481)
晶粒度评级的改进算法			赵 唯	(486)
用图象处理的方法监测原煤块率的研究	王新红	钱建生	刘富强	栗 君 (489)
数字化技术——提高电视图象质量的方向			李锦秀	(493)
大地震现场动态图象传输系统	张 咏	周 峰	杨桂君	张 昱 吴 强 (496)
适用于视频文件系统的有效调度方法	仇建伟	郑 红	(504)	
计算机人像组合查询识别系统	苏光大	尚 焱	严 超	谢炳龙 张翠萍 杨淑兰 (513)
局限性癫痫脑电的维数拓扑图方法	刘 慧	丁北生	刘 容	吕扬生 万柏坤 (517)
图象处理技术在体育机器人中的应用	齐丙辰	大川善邦	(534)	

基于小波的彩色图象匹配和整合的层次算法

徐丹* 石教英* 鲍歌⁺

(浙江大学 CAD & CG 国家重点实验室浙江, 杭州, 310027)(⁺香港理工大学电子计算系香港, 九龙, 红磡)

摘要 图象拼合是目前IBR(基于图象的绘制)系统中通用的一种场景编码技术,被用来建立复杂的虚拟场景表示。图象拼合的核心问题是图象整合,即将图象相互重叠的部分缝合。本文基于复值小波多分辨率分解(complex wavelet multiresolution decomposition)及人体视觉感知模型提出了一种有效的图象整合方法。首先,采用复值小波变换不仅可以保证全局优化的结果,还能够满足图象整合的规模不变和平移不变性。其次,采用人体视觉系统来过滤小波系数可以使处理的图象数据大大压缩(15-30:1),同时又保证视觉的可靠性。通过改进图象匹配和整合算法中相似距离的测量降低了计算复杂性。另外,基于多分辨率小波分解可以实现层次的图象整合,从而使系统与传统的图象整合方法相比具有较高的性能。最后,系统可以直接拼合从数字相机捕获的图象,而无需知道相机运动及其它内部参数。

关键字 图象拼合, 复值小波变换, 图象匹配, 图象整合, HVS

1 简介

IBR是虚拟现实应用中一项引人注目的新技术,它不依赖于几何模型,而是利用预先获得的一组图象对环境进行编码,并通过适当地组合这些图象生成位于不同视点的新视图,最终实现虚拟环境的完全漫游。IBR方法根据不同的表示模型可简单地分为五类:视图插值(View Interpolation)[6]、利用立体视觉的视图合成(View Synthesis via Stereo Vision)[8,15]、全视函数构造(Plenoptic Modeling)[7,15]、分层图象表示(Layered Image Representation)[1,11]和图象拼合(Image Mosaics)[22,23]。其中,图象拼合是指将一个场景的各个相重叠的子画面对齐并拼合成一个较大的画面的表示绘制方法。图象拼合方法已成为许多IBR绘制系统中的一项基本技术,用于虚拟现实应用中高分辨率的及宽视场的图象生成,例如,360球面和柱面全景图,环境映照及高分辨率图象等。

图象拼合方法的核心问题是图象整合(Image Registration),即是寻找一个变换,使图象间相重叠的部分对齐、并缝合成一个新的图象。这涉及到图象间对应部分的匹配问题。众所周知,图象匹配一直是摄影测绘学、计算机视觉和人工智能中最

难解决的问题,这可能是因为图象匹配的问题域一般是不封闭的,也可能是缺少足够的数学和信息理论支持,以及缺少支持在整个匹配过程中对信息处理机制进行彻底理解的工具。

图象整合算法所面临的匹配问题是怎样直接从图象数据中寻找局部对应关系。最简单的方法是手工指定四个应点:一般的整合算法依赖于图象间的相互关系(cross-correlation)[4,7,15]。这些方法大都采用寻找局部对应或抽取重要特征的方法来进行匹配,很多都要求校正相机,而且只能得到局部优解,还面临遮挡和特征附近的噪声等问题。基于相关联(phase-based correlation)的方法利用Fourier变换虽然可提供全局解,但是正如我们所知,Fourier变换具有很差的空间变比性。与之相反,小波分析在信号处理中具有很强的频域和空域的局部性和规模不变性。更进一步地,如果采用复值小波,还可保证平移不变性,并且使插值计算较为方便。

图象整合技术中另一个待解决的问题是如何寻找将图象间对应部分对齐所需要的变换。现有的一些文章把这个问题独力出来讨论,假设已经知道对应关系[21]。但是直接从原图中拟合变换比较复杂,问题规模大 系统性能低。层次算法能快速得

到拟合结果,并能逐层精化[14, 17, 24]。我们提出的算法利用小波分解的特点把匹配和变换拟合这两个问题结合起来同时考虑。

在本文所提出的方法中,首先输入图像经复值小波分解,然后所得的小波系数根据人体视觉系统(HVS)模型被筛选。接着,本文提出一种基于复值小波的层次匹配和整合算法,算法在大多数情况下能自动完成,并且不需要知道任何相机参数(如焦距等)。另外该算法可以得到全局解,还可根据统计结果由粗到精的逐层优化。图1描述了两个图像整合的算法结构。

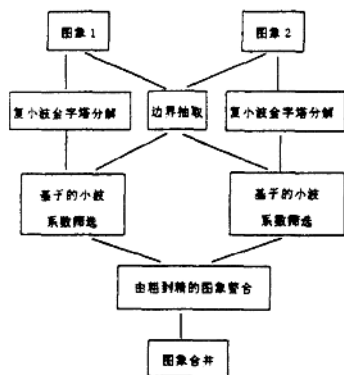


图1

2 复值小波分解表示

众所周知,小波分解是一项主要用于信号分解的技术[13, 19]。一般,给出一图像函数 $f(x, y)$,其实值小波分解定义为:

$$f(x, y) = A_1 f + D_{1,1} f + D_{1,2} f + D_{1,3} f$$

.....

$$= A_n f + \sum_{j=1}^n [D_{j,1} f + D_{j,2} f + D_{j,3} f] \quad (1)$$

其中 $A_j f$ 是在规模级 $s=1/2^j$ 上对 $f(x, y)$ 的拟合;

n 是最粗的分解级; $D_{j,p}$ 代表 $A_{j-1} f$ 与 $A_j f$ 间的差,称细节域。事实上,如已知图像 $f(x, y)$ 的高散表示,通常的实值小波金字塔分解是通过计算每一层上四个矩阵 $A_j, D_{j,p}, p=1, 2, 3$ 的滤波系数构成的,如图2(a)所示。

不同的小波有不同的特点。对于图像整合技术,往往要求小波分解满足规模、平移及旋转等不变性以便于图像的匹配,但是通常的实值小波分解

不能满足平移和旋转不变性[19],这正是其它基于小波分解的图像匹配算法所存在的问题。考虑到在采用图像拼合技术的IBR应用中通常待整合的两个图像或是视点相同、或是视点沿一个基轴平移、并只有很小的旋转,因此可集中考虑解决平移问题。在此前提下,复值小波是较合适的选择,因为空间域的平移可表示为复值相域中的旋转。而且,这种表示还便于对小波变换插值。图2(b)说明了复小波分解的过程,其中 $\{A_j, D_{j,p}, p=1, 2, 3\}$ 是系数矩阵集 $\{A_j, D_{j,p}, p=1, 2, 3\}$ 的共轭镜像。这两部分都包括了重要的信息。

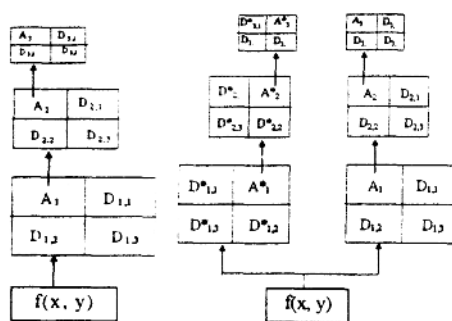


图2 小波多分辨率分解的实值及复值表示

在我们的算法中所采用的特定复小波是 Magarey & Kingsbury 复小波[13]。相应的小波和规模滤波器 g, h 定义为 Gabor 滤波[10]:

$$g_k = w(k + 0.5)e^{i\pi(k+0.5)}$$

$$h_k = \tilde{w}(k + 0.5)e^{i\pi(k+0.5)}$$

$$k = -N, -N+1, \dots, N-1 \quad (2)$$

其中 $\pi, \tilde{w}$ 是一对由 Gaussian 窗口构成的实值低通窗口:

$$w_1(k) = \exp\left(-\frac{k^2}{2\sigma_1^2}\right),$$

$$\tilde{w}_1(k) = \exp\left(-\frac{k^2}{2\tilde{\sigma}_1^2}\right) \quad (3)$$

最紧凑的窗口长度是4, 即 $N=2$ 。一种可选的 g 和 h 是 MKC-4:

$$f = [1-i \ 4-i \ 4+i \ 1+i]/10,$$

$$g = [-1-2i \ 5+2i \ -5+2i \ 1-2i]/14 \quad (4)$$

为了保证精度, 需要估计非整数下标的系数值。可插值性是复小波的一个优点。子采样的复值小波系数可以通过一个低通滤波器 o 与调制核的卷积来计算。

$$D_{j,p}(k+u, l+v) \approx D_{j,p}(k, l) e^{i\pi \Omega_{j,p} \cdot (u,v)} \quad (5)$$

3 基于HVS的小波系数筛选

本节将介绍基于人体视觉系统频率及空间模型的小波系数筛选机制, 详细讨论请阅读文献[2, 3]。

很多实验结果表明人眼对不同的频带有不同的反应, Mannos 和 Sakrison_ 将人眼的空间频率特性构造调制传递函数(MTF)[25]。它可以看作是空间带通滤波器。对于单色且静态的图象, MTF可表示为

$$MTF(\omega_r) = 2.6(0.192 + 0.114\omega_r) e^{(-0.114\omega_r)^{11}} \quad (6)$$

(6)式中, 线性操作符 w_r 假设是各向同性的, 并表示成辐射空间频率形式: $\omega_r = \sqrt{\omega_x^2 + \omega_y^2}$ 。

我们将小波基的空间频带特性应用到正规化的MTF空间域。在小波分解时, 带宽为 $B(0, \pi)$ 的信号被小波函数 y 连续投影到带宽 B_i 仅为基本域带宽一半的子空间 w_i 上。带宽 B_i 在第 i 层的分辨率为

$$B_i = \left(\frac{\pi}{2^i}, \frac{\pi}{2^{i-1}}\right), \quad \text{where } i \in Z \quad (7)$$

总带宽 B 为:

$$B = \sum_{i=0}^{\infty} B_i \quad (8)$$

小波变换系数的视觉敏感度 R_i 定义为MTF的空间频带与小波滤波器通带之积分比,

$$R_i = \frac{\int_0^{\pi} MTF(\omega) G_i(\omega) d\omega}{\int_0^{\pi} G_i(\omega) d\omega} \quad (9)$$

每个子域的量化步长为视觉敏感度的倒数, 即

$$S_i = \frac{1}{R_i}。$$

基于HVS的空间模型, 视觉感知灵敏度映射可按以下步骤由图象建立。首先, 从图象中抽取边界 $E(x, y)$, 然后将这些相邻的边界点连接组成边

界轮廓。边界轮廓定义为一个序列

$$S = \{(x, y, j) : x, y = 0, 1, \dots, n-1; j = 0, 1, \dots, m-1\} \quad (10)$$

其中 $n \times n$ 是图象的大小, j 是轮廓的长度。事实上, 不是所有的轮廓线段都是视觉重要的, 有的边界可能对人眼没有贡献, 尤其是那些较弱和较短的边界。因此, 把那些轮廓线段满足 $S^m < T^m$ 的边删去是合理的。 T^m 是构成视觉感知灵敏映射的轮廓线段长度阈值。在删除了哪些不重要的边之后, 剩下的轮廓线段就应该是高视觉感知的。我们定义视觉感知灵敏映射(VPSM)为 β 。

$$\beta(x, y) = \begin{cases} 1 & \text{if } (x, y) \in S^m \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

根据这个定义, 过滤小波系数的筛选函数 ξ 就可以建立了。因为小波分解保留了图象在不同子域的空间信息, 所以VPSM可用来跟踪各细节域从高层到低层的变换系数以识别出那些重要的高视觉感知系数, 及时地补充到基本域中。对小波分解的每个子图象, 出现在感知灵敏度映射中系数被量化为

$$c'_{k,l,j} = \begin{cases} c_{k,l,j} & \text{if } M > T \text{ and } (k, l) \in \xi(x, y) \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

其中 M 是最大的频带宽度, T 是由前面讨论MTF中得到的频带宽度阈值。

4 基于复值小波分解的图象匹配

图象匹配和整合算法的第一个任务是寻找参考图象对之间的部分匹配。我们采用层次式的匹配和整合策略, 由粗到精的逐步优化。值得强调的是, 在我们的机制中匹配算法和整合算法是互相配合, 互为结果的。也就是说, 每一层的匹配结果都用于该层的变换拟合, 而每一层的变换结果又用于下一层的匹配。

4.1 相似性测量

这里采用小波系数来进行相似性测量。算法与[18]近似, 但有两个重要的不同。首先, 我们根据每个子带的物理特性重新定义了模板矢量(Pattern Vector)。另外, 由于对各个子域进行了规范化处理, 因次在往下的较高分辨率层次上, 子域系数趋于零。因此可只考虑那些层次上的基本域部分对相似距离的贡献。因为相似距离计算是匹配过程中最

密集的计算,在每个象素处的执行次数可多达几百次,所以上述简化大大减少了相似距离计算的复杂性,提高了算法效率。

记 $PA_{j,q}(k,l)$, $q=1,2$ 和 $PD_{j,p}(k,l)$, $p=1,2,\dots,6$ 分别为基域和细节域中某象素中心的模板矢量,其被定义为

$$PA_{j,q}(k,l) = [A_j(k-r, l-c) : (r, c) \in N_q] \quad q=1, 2 \quad (13)$$

$$PD_{j,p}(k,l) = [B_{j,p}(k-r, l-c) : (r, c) \in N_p] \quad p=1, 2, \dots, 6 \quad (14)$$

其中,

$$B_{j,p}(k,l) = \frac{D_{j,p}(k,l)}{A_j(k,l)}, \quad p=1, 2, \dots, 6$$

$$\begin{aligned} &(\tilde{w}_j, w_j), (w_j, \tilde{w}_j), (w_j, w_j), (-\tilde{w}_j, w_j), (-w_j, \tilde{w}_j), (-w_j, w_j) \quad \text{for } p=1, 2, \dots, 6 \\ &(\tilde{w}_j, \tilde{w}_j), (-\tilde{w}_j, \tilde{w}_j) \quad \text{for } q=1, 2 \end{aligned}$$

其中 w_j 和 $\tilde{w}_j$ 满足:

$$w_1 = \frac{5\pi}{6}, \tilde{w}_1 = \frac{\pi}{6} \quad (15)$$

$$w_j = 3\tilde{w}_j, w_j = \frac{w_{j-1}}{2} \quad (16)$$

两个对应象素 (k, l) 和 (k', l') 间的相似距离定义为:

$$S_j((k,l),(k',l')) = \sum_{q=1}^2 SA_{j,q}((k,l),(k',l')) \sum_{p=1}^6 SD_{j,p}((k,l),(k',l')) \quad (17)$$

其中, $SA_{j,q}$, $q=1, 2$ 和 $SD_{j,p}$, $p=1, 2, \dots, 6$ 表示各个子带的相似距离, 定义为:

$$SA_{j,q}((k,l),(k',l')) = 1 - \frac{|PA_{j,q}(k,l) \bullet \overline{PA'_{j,q}(k',l')}|}{|PA_{j,q}(k,l)| |PA'_{j,q}(k',l')|} \quad q=1, 2 \quad (18)$$

$$SD_{j,p}((k,l),(k',l')) = |PD_{j,p}(k,l) - PD'_{j,p}(k',l')|, \quad p=1, 2, \dots, 6 \quad (19)$$

实验表明,在越下面的层次,即分辨率越高的层次,

称为小波分析的规范化细节域, $|\cdot|$ 表示取模。 $N = \{(0,0), (-0.5, -0.5), (-0.5, 0.5), (0.5, -0.5), (0.5, 0.5)\}$, $N_1, N_4 = \{(0,0), (-0.5, 0), (0.5, 0)\}$, $N_2, N_5 = \{(0,0), (0, 0.5), (0, -0.5)\}$, $N_3, N_6 = N$ 是象素中心的小邻域。模板矢量的定义考虑到了局部视差的连贯性,即在某象素附近的点应具有与该象素相同的视差。不同邻域的定义针对不同频率带滤波器的特性。例如,在水平低通垂直高通频率子带,将邻域定义为 N_1, N_4 ;而在垂直低通水平高通的子带,邻域定义为 N_2, N_5 。顶点的值可用插值公式(5)计算,其中 $\omega_{j,q}$ 和 $\omega_{j,p}$ 表示在 j 层上对子带 p, q 沿 x 和 y 方向的频率调制:

$B_{j,p}$ 的系数趋向零。于是,在这些层次上公式(17)中的第二个求和项可以忽略。这样做大大加快了高分辨率层上的匹配速度。

4.2 对应匹配

按照前面的相似距离定义,匹配算法在大多数情况下可以自动完成。如果存在大的视差,需要一些交互操作。匹配算法从两参考图象的最粗分辨率层开始,由里向外螺旋式进行。首先,利用(17)式寻找图象1的中心区域在图象2中的对应区域。第二,视差值由中心向边缘方向一圈圈传递,对每一圈都用(17)式再作视差微调。这个过程一直进行到对应边缘时停止。于是得到该层的对应匹配关系。第三,利用前面两步所得的对应关系可以拟合出一个较粗层次上的整合变换,这个变换又用于下一较精细的层,以求得其初始的对应关系,然后再用(17)式进一步调整。

注意图象整合与计算机视觉中的立体匹配问题[18]是不太相同的。立体匹配中左右眼的视差小,接近于完全匹配;而这里考虑的参考图象间的偏差可能很大,因此问题不仅仅是简单的完全匹配,需要考虑部分的图象匹配,即需要标记匹配区域的边界。我们用下面的公式来寻找边界:

$$|S_j((k,l),(k',l')) - S_j((k,l), N(k',l'))| \geq T_s \quad (20)$$

其中 $N(k', l')$ 是像素 (k', l') 的邻域。 T_e 是边界域值。当 (20) 式满足时, 当前检测的邻域像素就可判定为边界像素。

用上述算法可以求出图象 I_1 中的任意像素 (k, l) 在图象 I_2 中的对应像素 (k'_0, l'_0) 。为了获得更好的匹配, 可搜索 (k'_0, l'_0) 的一个小邻域以获得较精确的对应 (k', l') , 满足距离阈值 T_r ,

$$\min_{(k', l')} S_j((k, l), (k', l')) \quad (20)$$

$$\forall (k', l'): |(k, l) - (k'_0, l'_0)| \leq T_r \quad (21)$$

其中 $l \leq T_r \leq 2$ 。

5 层次图象整合

图象匹配和整合算法的第二个任务是寻找图象对间的整合变换。

一般地, 图象整合问题可以形式化为 $IR = \{G, M, I_1, I_2\}$, 它满足

$$I_2(k', l') = I_1(G(M; (k, l))) \quad (22)$$

其中 G 是一个平面投影变换, 完成从图象 I_1 到图象 I_2 的变换, M 是变换参数矢量: $M = (m_0, m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8)$ 。如果用齐次坐标表示, G 可以写为:

$$\begin{pmatrix} u \\ v \\ \mu \end{pmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \cdot \begin{pmatrix} k \\ l \\ 1 \end{pmatrix}, \quad \begin{pmatrix} k' \\ l' \\ 1 \end{pmatrix} = \begin{pmatrix} u/\mu \\ v/\mu \\ 1 \end{pmatrix} \quad (23)$$

其中 (u, v, μ) 是齐次坐标点, 它对应的图象笛卡尔坐标是: $k' = u/\mu$, $l' = v/\mu$ 。这种表达可以满足大多数 IBR 的应用要求。当两个图象有共同的视点时, 只需要恢复八个系数 ($m_8=1$), 于是 (23) 式重写为:

$$\begin{cases} k' = \frac{m_0 k + m_1 l + m_2}{m_3 k + m_4 l + 1} \\ l' = \frac{m_5 k + m_6 l + m_7}{m_3 k + m_4 l + 1} \end{cases} \quad (24)$$

如果已经知道两幅间的对应关系, 便可拟合从图象 I_1 到图象 I_2 的整合变换。显然, 这是一个非线性

性拟合问题。

我们由公式 (24) 导出误差:

$$e_i = \|(x'_i, y'_i) - G(M; (x_i, y_i))\|^2 \quad (25)$$

$$\min_M E = \sum_i e_i$$

公式 (25) 中 $\|\cdot\|$ 表示求欧几里德模数。为了最小化误差和, 我们采用 Levenberg-Marquardt 算法求解变换矢量, 其特点是可在很少的迭代次数后收敛 [16]。拟合结果根据 (17) 式可以由粗到精统计地优化。

实现 Levenberg-Marquardt 算法时, 要求计算 e_i 对每个 M 矢量中未知移动参数 m_i 的偏微分。例如,

$$\frac{\partial e_i}{\partial m_i} = \frac{x_i}{D_i} \cdot [(m_0 x_i + m_1 y_i + m_2) - (x_i + \delta x)], \quad (26)$$

其中 D_i 是 (15) 式中的分母。从这些偏微分, 可计算 Hessian 矩阵 A 及权重矢量 B

$$a_{ki} = \sum_j \frac{\partial e_i}{\partial m_k} \cdot \frac{\partial e_i}{\partial m_j}, \quad b_k = \sum_i \frac{\partial e_i}{\partial m_k} \quad (27)$$

最后, 迭代求解线性方程: $(A + \lambda \text{diag}(A)) \Delta m = B$, 可以逐步优化 M 的估计: $M = M + \Delta m$

算法:

① 从最粗的、也是最上面的层次开始, 用上一节介绍的螺旋式匹配算法找到最上层的对应关系:

$$x'_i = \delta x_i + x_i, \quad y'_i = \delta y_i + y_i,$$

并由此得到一个初始的整合变换估计。

② 用 Levenberg-Marquardt 非线性最小化算法拟合全局变换 (24), 使其满足公式 (25)。

③ 把第②步得到的全局变换用于下面的较精层, 获得该层初始的匹配。这些对应关系接着用 (17) 式进行调整以得到更优化的 (24) 式的解。

④ 重复迭代过程②、③ 直到误差小于指定的域值, 或达到一定的迭代次数

求得全局变换 (24) 之后, 就可以完成两幅参考图象 I_1 和 I_2 的混合。为了减少视觉走样, 给靠近图象中心的像素分配较大的权重, 而靠边缘的像素分配较小的权重。这样中心像素对组合的结果就有较大的贡献。权重函数可表示为双线性函数:

$$W(x, y) = w(x) \cdot w(y) \quad (28)$$

其中 w 是帽函数。它在图象边部分降为零。

图3是对拍自办公室的两幅图象进行整合的结果。

6 彩色图象的匹配和整合

本节讨论如何实现彩色图象的匹配和整合。众所周知,很多工作站和个人电脑都采用8位的帧存连同一个查色表来显示彩色,为此彩色图象必须被量化和表示为256种不同的颜色级。因此,要实现对彩色图象的匹配和整合,比如由红、绿、蓝三种色彩成分组成的RGB彩色图象,直接的方法是分别处理三个色彩成分的子图象,然后把三个结果结合成一个RGB的图象。但是,直接的方法存在两个明显的问题。一方面,时间开销大。因为匹配和整合算法对每一对图象要重复三次。而另一方面又需要大量的空间来存储每一个子图及相应的结果。

基于上述时间和空间开销的考虑,我们采用另一种简单且快速的方法。即是把RGB格式的图象根据R、G、B三成分间关系转换成亮度表示。这一转换表示成下式:

$$I = 0.30R + 0.59G + 0.11B \quad (29)$$

由于人体视觉对这三种色彩波谱有不同的感知。绿色成分最敏感,红色次之,蓝色的敏感度最低。因此在(29)式中绿色占有较大的权重。经上述转换后,彩色图象的匹配和整合算法就被解释为灰度图象的匹配和整合。这样做的好处是无论彩色还是灰度图象都用一致的算法处理。灰度图象的整合结果直接用于三个色彩子图就可获得匹配和整合后的彩色图象。

7 实验结果

我们对一组拍自室内和室外的图象数据进行了实验,整合的图象结果看上去几乎是无缝的(见图4、5)。注意图4(d)是(b)进行边界抽取后的结果。利用HVS模型过滤小波系数后边界抽取算法去掉了大量的不重要信息(压缩率为30:1),因而减少了计算量,提高了匹配和整合算法的效率。

所提出算法的另一个优点是重新定义模板矢量,使其自适应于不同的子域,这样可简化每一个象素处相似距离的计算。假设对一图象上的每一象素需搜索另一图象上的N个邻域点来寻找其对应象素,即要计算N次相似距离。用我们给出的相似距离定义来进行相似性测量,可在每个象素处理时比[18]中的方法减少 $N \times 16$ 次加法操作和 $N \times 8$ 次乘法操作。另外,我们发现在高分辨率层次上,规范化的细节子域系数趋于零。因此,可以忽略它们对

相似测量的贡献。这一选择提高了高分辨率层上的匹配速度。由于相似距离计算是整个匹配和整合算法中最密集的计算,在每个象素处可多达几百次,所以上述策略提高了整个匹配和整合算法的效率。

为了测量算法偏差,我们定义了代价函数:

$$err = \sum_i \|I(G(M(x_i, y_i))) - I'(x'_i, y'_i)\|^2 \quad (30)$$

其中, I 和 I' 是象素的亮度值,代价函数累计每对图象的对应点间的亮度误差。实际上,可选择一视觉容忍值 T_e , 并只累计大于 T_e 的误差。在我们的实验中,当选择 $T_e=10$ 时, (30) 式的误差均值小于 0.5。

8 总结

图象整合是基于图象的虚拟现实系统中建立复杂场景表示的一种基本方法。本文所讨论的图象整合方法为生成拼合图和全景图提供了一种全局优化的求解途径,它能同时完成自动匹配和整合图象,而不需要知道任何相机参数以及深度和对应信息。我们的算法利用复值小波分析保证了图象拼合是规模和平移不变的。另外,采用HVS模型,大大降低了计算复杂度并压缩了图象数据,而且保证视觉偏差很小。同时,由于简化了相似性测量,从而提高了算法效率。

参考文献

1. Adelson, E. H. Layered representations for vision and video. In *IEEE Workshop on Representations of Visual Scenes*, pages 3-9, Cambridge, MA, June 1995.
2. Bao, P., & Leung, B., "Wavelet Transform Image Coding based on Visual Perception Modeling", appear in the Proceedings of SPIE Conference, *Wavelet Application II*, April 1998, Orlando.
3. Bao, P., & Leung, B., "Image Compression based on Fuzzy Visual Perception Modeling and Wavelet", appear in the Proceedings of SPIE Conference *Transform Applications and Science of Computational Intelligence*, April 1998, Orlando.
4. Brown, L. G., A Survey of Image Registration Techniques, *ACM Computing Surveys*, 24(4):325-376, December 1992.
5. Chen, K., and Ramabadran, T., Near-lossless Compression of Medical Images through Entropy-coded DPCM, *IEEE Trans. on Medical Imaging*, 13, 538-548 (1994).
6. Chen, S. E. and William L. View interpolation for image synthesis, In *Computer Graphics(SIGGRAPH93)*, pages 279-288, 1993.
7. Chen, S. E. QuickTime VR - an image-based approach to virtual

- environment navigation. In *Computer Graphics(SIGGRAPH'95)*, pages 29-38, 1995.
- 8 Debevec, P. E. , Taylor C. J. and Malik J. Modeling and rendering architecture from photographs : A hybrid geometry_ and image-based approach. In *Computer Graphics (SIGGRAPH' 96)*, pages 11-20, August 1996.
 - 9 Gabor, D., Theory of Communication, *Journal of the IEE*, 93:429-457, 1946.
 - 10 Gortler, S. J. , Grzeszczuk R. , Szeliski R. and Cohen M. F. The Lumigraph. In *Computer Graphics (SIGGRAPH' 96)*, pages 43-54, August 1996.
 - 11 Lengyel, J. and Snyder, J. Rendering With Conherent Layers. In *Computer Graphics (SIGGRAPH'97)*, August 1997.
 - 12 Levoy, M. and Hanrahan P. Light field rendering. In *Computer Graphics (SIGGRAPH ' 96)*, pages 31-42, August 1996.
 - 13 Magarey, J and Kingsbury, N.G., Motion estimation using complex wavelets. *Technical report*, Cambridge University, Department of Engineering, 1995.
 - 14 Mallat, S. Zero-crossings of a wavelet transform. *IEEE T-IT*, Vol. 37(4):1019-1033.
 - 15 McMillan, L. and Bishop G. Plenoptic modeling: An image-based rendering system. In *Computer Graphics (SIGGRAPH ' 95)*, pages 39-46, August 1995.
 - 16 Numerical Recipes in C: The Art of Scientific Computing, 2nd edition, Cambridge, England, 1992.
 - 17 Olivo, J. C., Deubler, J. and Boulin, C., Automatic registration of images by a wavelet-based multiresolution approach. *SPIE Proceedings*, vol. 2569, pp. 234-244.
 - 18 Pan, H. P., General Stereo Image Matching Using Symmetric Complex Wavelets, In *SPIE Proceedings* Vol. 2825.
 - 19 Pan, H. P., Uniform Full-Information Image Matching Using Complex Conjugate Wavelet Pyramids, XVIII ISPRS Congress, Vienna, 1996. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI.
 - 20 Shapiro, J. M., Embedded Image Coding Using Zerotrees of Wavelet Coefficients, *IEEE Trans. Signal Processing*, 41(12), 3445-3462 (1993).
 - 21 Szeliski, R. and Coughlan, J., Hierarchical Spline-Based Image Registration, *Proc IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition(CVPR 94)*, IEEE CS Press, Los Alamitos, Calif., 1994, pp194-201.
 - 22 Szeliski, R., Video Mosaics for Virtual Environments, *IEEE Computer Graphics and Applications*, March, 1996, pp. 22-30.
 - 23 Szeliski, R., and Shum, H.Y. , Creating Full View Panoramas Image Mosaics and Environment Maps, *Computer Graphics Proceedings, annual Conference Series(SIGGRAPH'97)*, 251-258, 1997.
 - 24 Turcajova, R., and Kautsky, J., A Hierarchical Multiresolution Technique for Image Registration, In: *SPIE Proceedings, Mathematical Imaging: Wavelet Application in Signal and Image Processing I*, V. Denver, August 1996.
 - 25 L. Mannos, J. Sakrison, "The Effect of a Visual Fidelity Criterion on the Encoding of Images," *IEEE Trans. on Inform. Theory*, Vol. IT-20, no. 4, 1974.



(a)



(b)



(c)

图3 室内图象整合。(a), (b)是两幅参考图象, (c)是它们的整合结果。

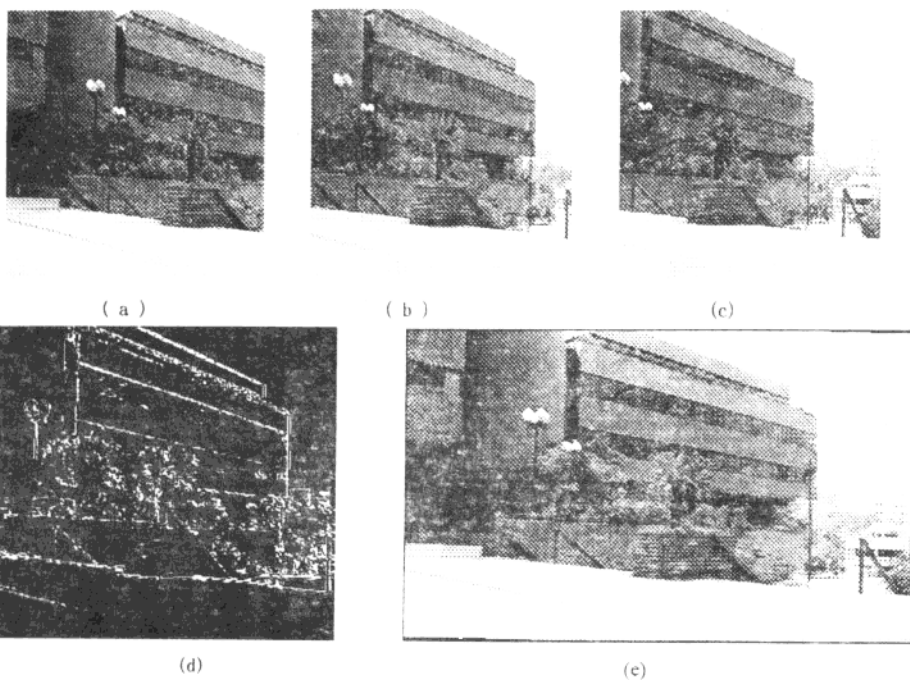


图4 基于平面投影的多幅图象拼合。(a), (b), (c) 是输入图象序列中的三幅，
(d) 是图(b)经边界抽取后的表示，(e) 是平面拼合的结果。

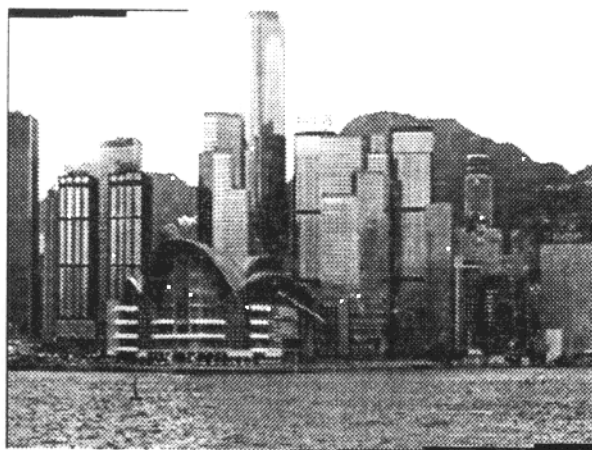


图5 香港会展中心的整合结果。由四幅图象经对应匹配和整合后缝合在一起。

基于内容的多媒体信息检索 CMIR 与 国际标准 MPEG-7*

章毓晋

(清华大学电子工程系 100084 北京)

摘要 为了有效地利用多媒体信息,国际上近年来广泛开展了基于内容对多媒体数据进行查询检索的研究工作。随着基于内容对多媒体信息处理、分析和理解研究工作的深入,国际上对多媒体信息内容描述标准化的需求也与日俱增,制定有关国际标准 MPEG-7 的工作也已启动。本文综合介绍了这两方面的情况和进展。

关键字 信息,多媒体,基于内容,查询,检索,国际标准,视频,音频,内容描述

1. 引言

随着科学技术的进步发展和推广应用,现代信息处理的对象和方法都有了很大的变化。这几年比较显著和重要的一点就是大量的各种类型的信息(特别是视觉信息)在全球得到了采集和应用,人们正在快速地进入一个信息化的社会。各种信息工具、技术、载体、方法和手段的综合集成产生了许多它们原来没有的新功能和新概念,例如出现了电子图书馆、交互电视、会议电视、远程教育、远程医疗等许多新的服务形式和项目。这些都对人们生活的方式节奏和社会结构的发展变化产生了重要的影响。由于各种信息的类别和载体的增加,以及它们的综合和集成,现在人们对把它们统而冠之的“多媒体”这个(新)名词已经很熟悉,尽管“多媒体”一词始终没有严格定义的[1],尽管“多媒体”一词至少在 1977 年已经出现[2]。

多媒体的产生和应用有社会发展和技术进步的基础和原因。从技术上分析,多媒体系统的能力主要来自于我们可将不同种类的信息都表达成一个码流(bit stream)。换句话说,无论是听觉信息还是视觉信息,或者还有嗅觉,味觉,触觉等不同形式的信息都可用统一的码流形式来表达。这种统一

的表达是多媒体广泛应用和传播的一个重要因素。因为这样不同形式的信息就可以用同一个设备——计算机来存储,处理和传输。将各种不同的信息形式都转化成码流还使我们能将注意力集中在掌握理解信息本身的内容上,而不是在获得信息的传感器和传递信息的通道上。

码流是一种非常适合于目前的计算机和通信设备进行存储和传输的形式。但是如果没有码流结构的知识和信息,码流只是码流,并没有用处。例如将一幅图用一串二进制码来显示是不能让人看懂的。我们必须把码流转换成一种我们的感觉可以理解的形式才能有效地利用码流。这种转换本质上是一个对信息抽象化的过程,也可以看作是一个对媒体进行处理,分析和理解的过程。随着研究的深入,人们逐渐认识到这个过程的重要性,并提出了不少具体的方法。这其中最有特色最有效的就是基于内容的多媒体信息检索(CMIR, Content-based Multimedia Information Retrieval)。而这方面的研究又促进了国际上对信息描述标准化(MPEG-7)的进展。本文将对这两方面的内容给予介绍。

2. 基于内容的多媒体信息检索

2.1 基本原理

本文工作得到国家自然科学基金和高技术计划(863)资助。

现代技术已能运用各种手段大量地采集和产生各种类型的信息数据。据估计全世界的信息量正以 20 万倍于人口的增长速度递增[3]。全球各地已建立了许多大型的多媒体数据库并向公众提供服务[4]。随着信息高速公路的建设,各种信息的流通和传输也将越来越方便。但在很多情况下,信息膨胀已给人类带来过多的信息量以至于要超过人的接受能力了。

多媒体信息中量最大最主要的一种就是图象信息。这里图象的含义是比较广泛的,既包括一般所说的静止图象,也包括视频,计算机图形等可作用于人的视觉器官的实体。由于图象信息数据量大,抽象程度低,上述信息膨胀问题更为严重。

为解决上述问题需要对多媒体信息(特别是视频音频信息)进行有效的管理和利用。我们不仅要能大量存储和快速传输多媒体信息,我们还要对多媒体信息进行自动选择查询,使人类只接受或能快速地获取需要的信息而不至于淹没在多媒体信息的汪洋大海之中。

传统的对视频音频等查询的方法主要是借助对它们编号、加标签来进行的。具体就是给视频音频加上一个对其进行描述的文字或数字标签,然后在索引时对标签进行检索。这样一来对视频音频的查询变成了基于标签的查询。这种方法虽然简单,但有几个根本的问题影响我们对视频音频信息的有效使用。首先由于视频音频内容丰富,事实上很难用文字标签完全表达。所以根据这种局限方法进行的查询常会出现错误。其次文字描述是一种特定的抽象,如果描述的标准改变,则标签也得重新制作。换句话说,特定的标签只适合特定的查询要求。最后目前这些文字标签基本是靠观察者选出来加上去的,因此受主观因素影响很大。不同的观察者或同一个观察者在不同条件下对同一组视频音频数据可能给出不同的描述。由于不够客观且没有统一标准,这些描述常会自相矛盾。

为克服上述问题需要全面地、一般性地和客观地提取视频音频的内容来。实际上人们利用视频音频并不是仅根据它们的视觉或听觉质量而更重要的是根据它们的视觉和听觉内容,所以只有根据内容进行视频音频检索才有可能有效地获得所需的视频音频信息,同时在掌握视频音频内容的基础上视频音频数据库中的信息才有可能有效地得到管理。

如何提取视频音频内容呢?在较低的层次上可借助视频音频的特征,例如图象中目标的几何形

状、颜色或灰度、纹理、相互位置及它们的组合;声音中目标的音响、音调、音色以及和声等。这些特征常可自动或基本自动地从视频音频中提取出来。在较高的层次则需要借助识别、分析、理解等技术,例如可利用人工神经网络训练学习,以帮助索引视频音频中各种模式。

对视频音频内容的提取可在建立视频音频数据库时进行。图 1 给出基于特征对视频音频进行归档和查询的原理框图[5]。在视频音频归档时,对输入先进行分析,提取视频音频或目标的特征向量。在将输入视频音频存入视频音频库的同时将其相应的特征向量也存入与视频音频库相连的特征库。在视频音频查询时,对给定的查询图,也先进行分析并提取其特征向量。通过将该特征向量与特征库中的特征向量进行匹配并根据匹配结果到视频音频库中搜索就可提取出所需要的视频音频来。

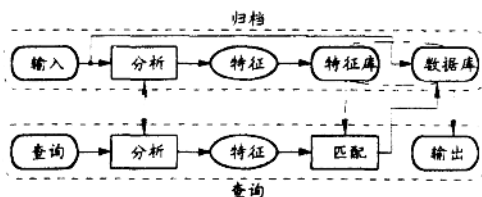


图 1: 视频音频归档和检索流程图。

对多媒体数据库的查询从查询方式来看可分为两类。一类是直接查询,即查询条件的种类和数据库内数据的种类相同。另一类是间接查询,即给定的查询条件与最后想查询的数据的种类不同。例如查询条件可以是文字或数字而查询对象可以是图象或声音,或者反过来。这后一种查询在多媒体系统中是常见的。这是因为在多媒体数据库中有许多不同类型的数据或媒体,我们有时会需要在同一时刻面对不同类型,而查询的条件为了方便常常只用某一类数据或媒体表示。间接查询比直接查询在指定具体查询条件时更为灵活[3]。

2.2 一些典型系统

鉴于基于内容的多媒体信息查询检索的重要性,有效性和优越性,近年来国际上已纷纷投入人力物力广泛开展了研究并已制成一些试验系统。例如 IBM 研制开发了基于图象内容查询的系统 QBIC (Query By Image Content)。它允许用户给出希望查询图象或目标的形状,颜色或纹理,然后系统以此为例可以从大型在线(图象)数据库中搜索、

浏览和提取库中满足条件的图像。

在 QBIC 中, 建立数据库有三个步骤: 采集图像, 制作缩微图 (thumbnail image), 加入文字信息。用户可有选择地用手或半自动地指定目标或区域。手工时主要是勾画边界, 半自动则可仅勾画一个范围, 系统在这个范围内确定边界。图像库建立后要计算特征, 特征可以是彩色特征、纹理特征、形状特征、草图特征等。QBIC 中的彩色特征主要有彩色直方图和平均值等; 纹理特征主要有粗糙度、对比度和方向性等; 形状特征主要有目标的面积、各阶矩、圆形性、离心率和主轴方向等; 草图特征也可称为边缘图特征, 如边界各点的切线方向等。

将特征计算好与图像一起存入数据库后就可以进行查询了。查询可对整幅图进行, 也可对图中目标进行。在用彩色特征查询时 QBIC 允许用户指定各种颜色的百分比。在用纹理特征查询时 QBIC 允许用户从给定纹理库中选取有代表性的纹理图案。在用形状特征查询时 QBIC 允许用户指定一个已有目标, 也允许用户勾画草图来表示形状。而对草图特征, 由于主要考虑边缘, 所以 QBIC 允许用户只画出主要边缘的位置和走向。QBIC 在设计时强调了用户界面, 用户既可使用以上任一特征, 也可利用它们的组合[6]。

该系统已在一些场合试用[7]。例如旧金山的艺术博物馆中采用 QBIC 查询艺术图片, 加州大学 Davis 分校的艺术和艺术历史系利用 QBIC 让学生和公众了解教师的信息。QBIC 采用美国邮票以及专利和商标进行的试验也给出了很好的效果。

再如由新加坡国立大学开发的计算机辅助人脸像推断和抽取系统 (CAFIR) [8]。它可用于帮助警察通缉寻找罪犯。具体就是警察根据目击者的描述合成罪犯脸像, 而系统根据脸像特征 (主要是五官的形状和相互位置) 在以往记录中搜查以确定罪犯。

视频信息的管理是多媒体信息利用中的一个重要问题。加州大学研究的 Virage 视频引擎[9]提供了自动处理分析视频码流产生其索引以及利用索引提取视频的手段和工具。它允许对视频进行包括图像序列和声道索引的多模式索引。

另外 MIT 的图像数据库管理系统 Photobook[10], 哥伦比亚大学的全自动图像查询系统 VisualSEEK [11], 美国国家医学图书馆的基于万维网的医学信息提取系统 WebMIRS[12]都是典

型的基于内容的多媒体信息查询检索系统。以上一些系统主要针对图像和视频的查询检索, 事实上基于内容的音频检索系统也得到了广泛的研究, 例如 Muscle Fish (见[13])。

2.3 通用框架和模块

以上列举的一些典型系统虽各有特点, 但它们也有共性。一般来说基于内容的多媒体查询检索系统是介于信息用户和 (多媒体) 数据库之间的一种信息服务系统。用户通过它可以按自然的方式从库中提取满足所需要内容的信息。这类系统的通用框架可见图 2, 主要由五个模块组成[14]:

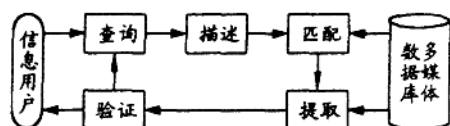


图 2: 基于内容的多媒体信息检索系统框图。

- (1) 查询模块: 对用户提多样的查询手段, 以支持用户根据不同应用进行各种类型的查询工作;
- (2) 描述模块: 将用户查询的外部表达转化为对多媒体信息内容的比较抽象的内部表达和描述;
- (3) 匹配模块: 将查询描述与库中被查询信息的描述进行匹配以确定它们内容的一致性和相似性;
- (4) 提取模块: 在内容匹配的基础上将所有满足给定条件的信息自动地从库中提取出来;
- (5) 验证模块: 对用户提方便的验证手段以评价提取的效果。如果不够满意, 新一轮的查询可通过修改查询条件而重新开始。

对每个模块都有许多技术可以采用, 结合我们的初步研究可给出如下一些例子:

- (1) 对图像的指定查询方法主要有: a) 利用采样图像, 即借助库中的图像[15]; b) 利用示例图像, 即借助与期望图像类似的图像; c) 利用画出的草图 (常是简单的素描); d) 某种 (概括) 描绘。
- (2) 描述可用的方法与特征有关。例如对颜色, 可利用直方图、累积直方图[16]、局部累积直方图[17]等; 对形状, 可利用目标边缘的方向直方图; 对区域分布可借助图像分块后的子块直方图。