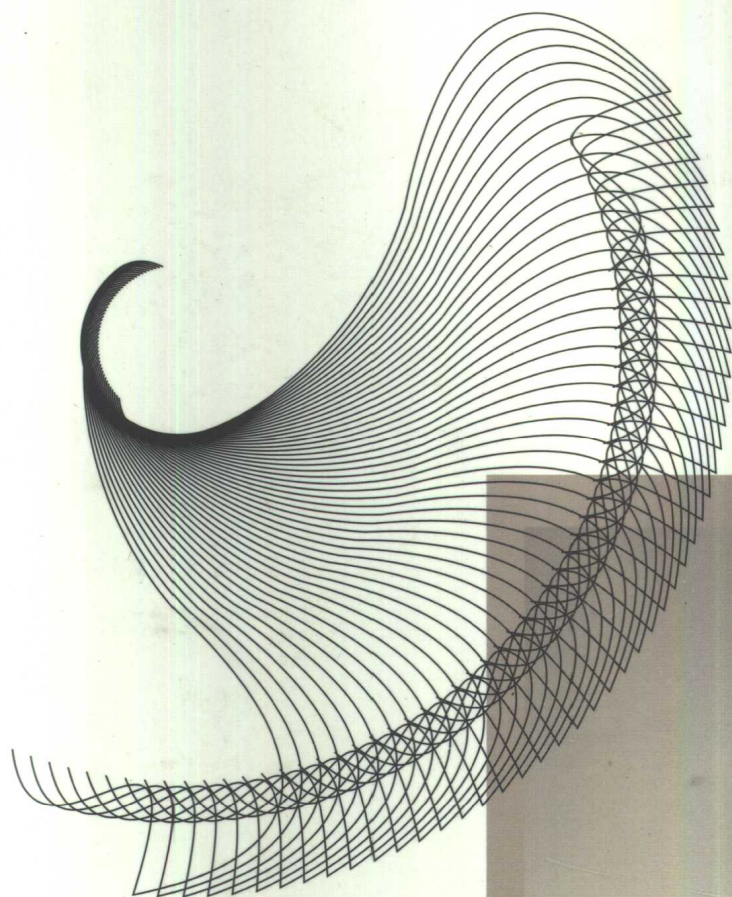


国家九五重点图书

计算机应用技术前沿丛书

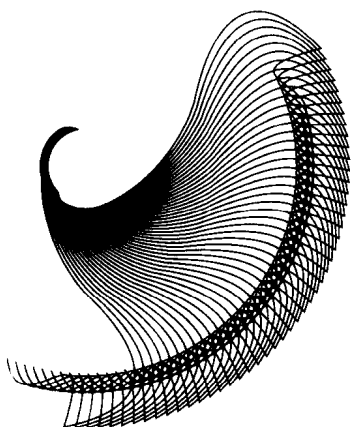


图象与声音压缩技术

杨长生 编著

浙江大学出版社





计算机应用技术前沿丛书

国家九五重点图书

图象与声音 压缩技术

杨长生 编著

图书在版编目 (CIP) 数据

图象与声音压缩技术 / 杨长生编著. —杭州: 浙江大学出版社, 2000.4

(计算机技术前沿丛书 / 潘云鹤主编)

ISBN 7-308-02318-4

I. 图... II. 杨... III. ①视频信号-频率压缩②音频-信号-频率压缩 IV. TN911.7

中国版本图书馆 CIP 数据核字 (2000) 第 21578 号

总 责 编 陈晓嘉 梁 兵

责任编辑 杜希武

封面设计 俞亚彤

版式设计 孙海荣

责任绘图 张作梅

责任出版 李慧华

出 版 浙江大学出版社

(杭州浙大路 38 号 邮政编码 310027)

(网址: <http://www.zjupress.com>)

(E-mail: zupress@mail.hz.zj.cn)

排 版 浙江大学出版社电脑排版中心

印 刷 浙江印刷集团公司

经 销 浙江省新华书店

开 本 889mm×1194mm 1/16

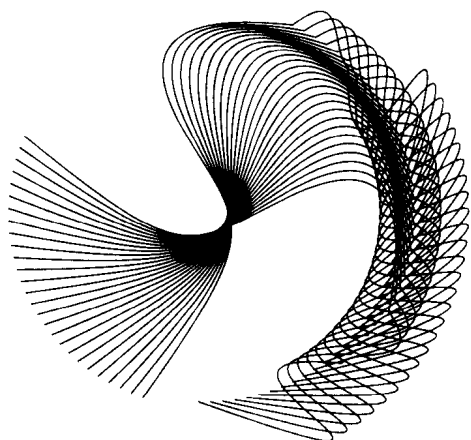
印 张 26.25

字 数 615 千

版 印 次 2000 年 4 月第 1 版第 1 次印刷

书 号 ISBN 7-308-02318-4/TN·047

定 价 52.00 元



主 编 潘云鹤

副主编 陈 纯

顾 问(以姓氏笔划为序)

王 选 李三立 李国杰 李衍达

沈昌祥 张 钹 张效祥 汪成为

何志钧 杨芙清 金怡濂 陈火旺

胡启恒 唐泽圣 戴汝为



序

20 世纪是一个科技、经济空前发展的时代,从世纪初相对论、量子理论的创立到今天以信息产业为龙头的高科技产业成为经济发展的第一支柱,人类社会发生了根本性的变革。而在这场以科学技术为社会发展直接动因的变革中,意义最深远、影响最广泛的就是计算机及其相关技术的发展和运用。

在过去的 50 年里,计算机已从最初的协助人类进行精密和复杂运算的单一功能的运算器发展成为能够模拟人类智慧、感觉,可以适应环境、具有多种功能的设备;计算机及其技术的应用已从尖端科学、国防工业延伸到了人类活动的各个领域。它不仅充实和革新了传统的科学与技术研究手段,带来了新的方法和理念,成为科学技术谋求新发展的沃壤,而且作为信息处理的核心工具和技术支撑起庞大的信息产业,进而改变了全球的产业格局。

目前,一个由遍布全世界的计算机连接起来的跨越国界、巨大而高速的多媒体信息网络正在形成,因此,不仅信息产业本身,而且各行各业都将面临计算机应用技术研究和发展的重大课题。我们针对国内外有关计算机应用技术的最新进展,根据浙江大学及国内其他学者在计算机辅助设计系统和方法、计算机动画、设计思维模拟、自主式智能系统、智能决策系统、软件智能化技术、工程数据库管理系统、图象与声音

压缩技术等方面取得的成果,组织编撰了《计算机应用技术前沿》丛书。我们希望,通过这套书,为从事相关领域研究工作的专家和学者提供一些有参考价值的方法和技术,为普及、推广进而开拓新的应用技术作出一点贡献。

潘云翔

一九九九年十月

前 言

在当前这样一个充满信息的社会中,人们在交换图象、声音等媒体信息方面的需求越来越大,对交换的质量要求也越来越高。随着微电子、计算机和传感器等技术的高速发展,图象、声音等媒体信息的记录、存储和传输正在朝着数字化方向前进。通常,图象、声音经数字化后形成的数据量非常大。例如,一幅数字化了的单色卫星遥感图象由 $10\,000 \times 10\,000$ 个像素(pixel)组成,每个图象点灰度用 12 比特(bit)表示,那么这幅图象就要用 1 200 000 000 比特表示。直接存储和传输如此庞大的数据不但开销很大,而且有时设备也承受不了如此大的负荷。然而,在这些数据中存在很多冗余。采用目前的压缩技术,对于上述图象可以毫不失真地压缩 4 倍;如果允许部分失真,那么就可以压缩几十或几百倍。显然,这种压缩技术具有重大的实用价值和广阔的发展前景。《图象与声音压缩技术》就是研究如何去除数据冗余,从而高效表示图象、声音数据的技术。

从有了电话的年代开始就有人研究声音压缩技术了,而对图象压缩技术的研究至今已有五十多年了。在最近的一二十年内,微电子和计算机技术得到了高速发展,并取得了长足的进步。因此,数字图象、声音压缩技术再次成为信息学科领域内的研究热点之一。经过多年的研究,人们已经成功地开发出了各种各样的压缩技术和方法。本书旨

在向高等学校有关专业(如计算机应用、电子通信等)的研究生,介绍一些对图象、声音媒体数据进行压缩的基本技术、研究热点,以及被证明是行之有效乃至成为国际标准的压缩方案。通常认为,图象信号是二维以上的多维信号,而语音信号比图象信号简单得多。实际情况表明,许多对数字图象数据压缩有效的技术和方案,对数字声音数据的压缩也很有用途。为此,本书重点介绍数字图象压缩技术,而对数字声音压缩技术只作一般的典型介绍。通常也认为,单色、静止图象的压缩技术是彩色、运动(序列)图象压缩技术的基础,所以本书以介绍单色、静止图象的压缩技术为主。考虑到当前视频编码技术正在得到越来越广泛的应用,本书还专门介绍了视频编码技术和美国大联盟 HDTV 系统。另外,由于采用同样的压缩方案对不同来源、不同类型的图象压缩的效果往往大不一样,而进行信息交换的图象是大量来自于现实世界的、具有连续色调的图象,因而,本书介绍的图象压缩技术主要用于处理这类图象。为便于读者对不同压缩技术与方案进行比较,本书采用了一些常用的、来自于现实世界的、具有连续色调的、单色的、静止的图象,例如 LANA, BOOTS 等,作为实验图象。当然,应当注意不同分辨率、不同的输出设备产出的图象的观察效果是不一样的。

全书由七章和一个附录组成。第 1 章主要介绍数字图象的基本知识,以及国际标准化组织在数字图象压缩方面的活动。第 2 章介绍了信息论的一些基本概念,包括图象模型、熵和有损压缩的失真度理论;另外,还介绍了一些常用的可变长度编码技术: Huffman 编码、算术编码和 IBM Q 编码器。本书的第 3 章是关于数字图象无损压缩技术,介绍的主要技术有:位平面编码、无损预测编码,以及有损编码与无损残差编码的组合编码技术。在本书的第 4 章和第 5 章中,详细讨论了当前流行的、各种各样的数字图象有损压缩技术:预测、变换、块串、向量量化、子频带、层次化、神经网络法、几何模型化、分形和小波变换等编码技术。在第 5 章的第 5.5 节不仅介绍了视频图象编码的基本技术,而且简单介绍了美国 HDTV 大联盟系统。第 6 章则介绍了人们很关

心的一些图象压缩的国际标准: JBIG, JPEG, MPEG-1, MPEG-2 和 MPEG-4 等, 以及正在酝酿中的新标准。在本书的最后一章(第 7 章)介绍了一些语音和音频压缩的基本的、流行的技术, 如 LPC(MP-LPC), RPE, CELP 等, 其中特别详细介绍了 CELP 算法以及国际标准的 AC-3 算法。另外, 在这一部分还介绍了当前语音和音频压缩的研究动向与热点。本书各章末尾大量地列出了在数字图象、声音压缩领域最有影响的参考文献。最后, 在本书的附录中列出一些在压缩技术中最常用算法的 C 语言程序清单。综上所述, 本书充分反映了当前数字图象和语音编码学科的最新技术和研究动向, 可以供高等学校计算机应用、电子通信等专业的研究生、教师和技术人员参考。

本书第 5 章中的第 5.3 和 5.4 节, 即分形和小波编码, 分别由宋广华、张三元编写。本书的其余部分由杨长生编写。清华大学李衍达院士对本书的编写大纲提出了许多有益的建议; 浙江大学顾伟康教授对本书进行了全面的审阅, 提出了不少建设性的意见, 值此机会表示感谢。

由于作者水平有限, 书中错漏难免, 不当之处, 请读者批评指正。

杨长生

2000 年 4 月

目 录

1 背景:数字图象和图象压缩	1
1.1 数字图象形式	1
1.2 数字图象及处理技术的应用	2
1.3 数字图象处理的主要研究内容	5
1.4 数字图象压缩的必要性	7
1.5 图象压缩技术的分类	9
1.6 数字化参数对压缩的影响	10
1.7 图象压缩的标准化组织	11
文献索引(一)	12
2 信息理论的基本概念	14
2.1 数字图象信息源模型与熵	14
2.1.1 离散无存储器的信源 DMS	14
2.1.2 离散无存储器信源的扩展	15
2.1.3 马尔可夫(Markov)源——带存储器的源	15
2.1.4 马尔可夫源的扩展和相交源	17
2.1.5 无噪声信源的编码理论	17
2.2 可变长度码	18
2.2.1 编码效率和信源的扩展	18
2.2.2 霍夫曼(Huffman)码	19
2.2.3 改进的 Huffman 编码	20
2.2.4 Huffman 编码技术的限制	21
2.2.5 算术编码	21

2.2.6	IBM Q 编码器	23
2.3	熵的估计和无损压缩	24
2.3.1	英语的结构和熵	25
2.3.2	英语的可预测性和熵	26
2.3.3	自然图象的可预测性和熵	27
2.4	失真度理论和有损压缩	27
	文献索引(二)	29
3	图象无损压缩技术	30
3.1	位平面编码	30
3.1.1	格雷码(反射码)	31
3.1.2	位平面的游程编码	32
3.1.3	位平面的算术编码	36
3.2	无损预测编码	37
3.2.1	DPCM 预测器的阶、分类和计算方法	38
3.2.2	差值图象的 Huffman 预测编码	38
3.2.3	差值图象的算术编码	42
3.3	有损编码加无损残余编码	43
	文献索引(三)	44
4	图象有损压缩技术(一)	46
4.1	有损预测编码	48
4.1.1	差值脉冲编码调制(DPCM)	48
4.1.2	自适应	54
4.1.3	DPCM 的结果	57
4.1.4	ADPCM 算法的实现及复杂度	60
4.2	变换编码	61
4.2.1	坐标轴旋转的变换	62
4.2.2	基本函数分解的变换	63
4.2.3	图象变换	64
4.2.4	变换编码的策略	67
4.2.5	JPEG DCT 算法	69
4.2.6	JPEG DCT 结果	74
4.2.7	JPEG DCT 算法的实现和复杂度	77
4.3	方块编码	78
4.3.1	量化器的设计	79

4.3.2	位图源编码和重建值	81
4.3.3	图象大小自适应的 BTC	83
4.3.4	BTC 结果	84
4.3.5	图象块大小自适应 AMBTC 算法的实现和复杂度	85
4.4	向量化编码	86
4.4.1	编码书的产生	88
4.4.2	编码书的设计:树形结构编码书	90
4.4.3	编码书的设计:产生式结构编码书	91
4.4.4	编码书的设计:分类 VQ	93
4.4.5	编码书的设计:有限状态 VQ	94
4.4.6	VQ 结果	95
4.4.7	M/RTVQ 和 I/RTVQ 算法的实现和复杂度	99
4.5	子频带编码	100
4.5.1	1-D 信号的分析/综合滤波器	101
4.5.2	2-D 信号的分析/综合滤波器	103
4.5.3	子频带编码技术	106
4.5.4	子频带编码和变换编码间的关系	107
4.5.5	SBC 结果	108
4.5.6	SBC/VQ 算法的实现和复杂度	111
4.6	层次编码	111
4.6.1	累进传输	112
4.6.2	多用途环境	112
4.6.3	图象层次化	113
4.6.4	可变分辨率分层化	115
	文献索引(四)	119

5	图象有损压缩技术(二)	125
5.1	基于神经网络的编码	125
5.1.1	基于神经网络的图象压缩	126
5.1.2	基于神经网络的图象预测编码	127
5.1.3	基于神经网络的图象变换编码	131
5.1.4	基于神经网络的图象向量量化编码	137
5.1.5	算法评估	141
5.2	基于几何模型的编码	143
5.2.1	基于几何模型的图象编码	143
5.2.2	面部图象的三维模型	146
5.2.3	人脸的三维模型编码的一个范例	149

5.2.4	模型/波形混合编码	154
5.2.5	应用和实现	157
5.2.6	三维模型编码的遗留问题	158
5.3	分形编码	159
5.3.1	真实世界图象的特性	160
5.3.2	真实世界图象的数学模型	164
5.3.3	分形图象压缩的数学基础	166
5.3.4	交互分形图象压缩	173
5.3.5	自动分形图象压缩	175
5.3.6	基于四叉树分解的分形图象编码	181
5.4	小波编码	182
5.4.1	小波的定义及基本性质	182
5.4.2	小波变换与图象压缩	194
5.5	视频编码	209
5.5.1	视频编码器结构	210
5.5.2	离散余弦变换	212
5.5.3	量化	214
5.5.4	具有视觉特性的加权矩阵	217
5.5.5	运动补偿	221
5.5.6	美国高清晰电视(HDTV)大联盟系统	228
	文献索引(五)	250
6	图象压缩的国际标准释例	259
6.1	引言	259
6.2	二值图象压缩标准——JBIG	259
6.3	静止图象压缩标准——JPEG	264
6.4	运动图象压缩标准——MPEG	269
6.5	发展中的 MPEG 压缩标准——MPEG-7	301
6.6	视频会议标准	307
	文献索引(六)	309
7	语音和音频压缩技术	310
7.1	基于线性预测的分析/综合编码	312
7.2	CELP 算法	314
7.2.1	历史	314
7.2.2	闭环搜索	315
7.2.3	激励编码书	315

7.2.4	音调周期的表示	317
7.2.5	LP 频谱参数的编码	317
7.2.6	多模式编码与音色分类	318
7.2.7	内插值编码	318
7.2.8	后滤波和预滤波	319
7.2.9	其它 CELP 技术	319
7.3	低延时语音编码	320
7.4	可变速率语音编码	320
7.5	声码器	321
7.5.1	LPC 声码器	321
7.5.2	正弦编码器	322
7.6	音频和宽频语音编码	322
7.6.1	变换编码技术——AC-3 编码简介	324
7.6.2	AC-3 编码的位分配策略	326
7.6.3	滤波器组	329
7.6.4	频谱包络	329
7.6.5	位分配的细节	332
7.6.6	联接技术	334
7.6.7	位流语法	335
7.6.8	特性	335
7.6.9	结论	338
7.7	音频和宽频语音编码的热点	338
	文献索引(七)	339
附录 图象压缩常用算法程序清单		349
(1)	PCM 编码	349
(2)	DPCM 预测编码/DPCM 预测解码	349
(3)	图象采样	352
(4)	基于 MSE 最小原则设计量化器	356
(5)	Walsh 分析/合成滤波器	359
(6)	图象边界检测	361
(7)	Huffman 编码/解码算法	365
(8)	LBG 训练算法	373
(9)	向量量化和动态位分配的压缩算法	385
(10)	分形图象编码模型源程序	392



1 背景:数字图象和图象压缩

随着微电子、计算机和传感器等技术的高速发展,图象、声音等媒体信息的记录、存储、传输正朝着数字化方向前进。通常,图象、声音等媒体信号经过数字化处理之后形成的数据量非常庞大。例如,一幅 $24\text{mm} \times 36\text{mm}$ (即通常所说的 35mm) 的彩色胶片,若以 $12\mu\text{m}$ 的间距进行扫描,则形成三幅彩色数字图象。每一幅彩色数字图象由 $3000 \times 2000\text{pixel}$ 象素组成;如果每个象素用 8bit (比特) 数据量表示,那么三幅数字图象需用:

$$3000 \times 2000 \times 8 \times 3\text{bit} = 144 \times 10^6\text{bit}$$

如此庞大的数据量不仅存储开销很大,而且在传输时使得通讯设备的负荷很重,甚至承受不了。然而,采用压缩技术,可以把这些数据压缩几倍、几十倍,甚至几百倍。显然,研究如何去除冗余数据,高效地表示图象、声音等媒体信息,即进行“数据压缩”,具有重大的实用价值。因而,数字图象、声音压缩技术已成为目前国内外的研究热点之一。^[1~21]

1.1 数字图象形式

数字图象可以是来自于真实世界的 3-D (即 3 维) 景色;也可能是由 3-D 景色产生的 2-D 图象 (例如照片)。这些图象源的信号本身都是模拟量,它们在空间和幅度上都是连续的。为产生数字图象,就要采用传感器对信号源在离散位置进行抽样。这些抽样的样本就称为图象的象素 pixel (即 picture element)。样本的值通常与样本所处位置发射的能量密度线性相关;而它们的数值落在一个有限的范围内。请注意,由于采样所用的传感器也许只对人类视觉范围外的波长敏感,所以有时人们感觉到的光线的强弱与采集到的‘发射的能量密度’并不完全一致。

对图象部位进行采样可以采用各种各样的方法,而大部分是采用正方形的网

格。在理想情况下,每个样本的值属于一个有限的数集。然而,由于传感器的光学物理特性,使得样本的值为有限范围内的一个实数。单位面积内采集的样本个数就称为采样率。采样率应服从于 Nyquist 理论,即采样率必须不小于信号源空间最高频率成分的 2 倍。如果达不到这样的抽样率,那么当信号源的空间频率高于 $1/2$ 抽样率时就会产生低频镜像。为防止产生镜像,通常在采样之前用滤波器进行滤波,从而压缩系统的频宽(注意,传感器的积分特性也会起预滤波的作用)。在图象数字化系统中,常常提到扫描分辨率(scanning resolution)这一术语。扫描分辨率实际就是采样率的倒数。例如,扫描分辨率是 $50\mu\text{m}/\text{pixel}$,那么采样率就是 $20\text{ pixel}/\text{mm}$ 。

为产生数字图象,需要把采样得到的连续值量化为离散数值中的一个值。众所周知,当亮度均匀变化时,人类视觉系统 HVS(Human Visuals System)所感觉到的变化并不是均匀的。由于在大多数应用场合,图象被人们主观地观察,所以按照人类的视觉系统 HVS 的感觉效果进行量化是必要的。在许多场合,通过对图象作非线性量化(如:指数或立方根函数量化),使得量化后的效果接近于人类的视觉系统 HVS 的感觉效果。对图象量化的级数取决于应用的需要。例如,对于黑白文本图象,由于文本图象中的任何一个象素不是白就是黑,所以只要两个量化级就行了(即 $1\text{ bit}/\text{pixel}$)。对于自然景物图象,或连续色调的图片,通常采用 256 个量化级(即 $8\text{ bit}/\text{pixel}$)。然而,当图象信号源的动态范围比较大、传感器的输出精度比较高时,则可以采用 $10\text{ bit}/\text{pixel}$ 或 $12\text{ bit}/\text{pixel}$ 进行量化。实际上,为了克服传感器的非线性引起的对模拟量直接量化后的误差,比较合理的图象数字化方法是首先采用 12 位或 14 位对线性空间的象素值进行量化;然后对量化的数据作非线性校正;最后,用查表法把这些数据再量化为 8 位的二进制数。

1.2 数字图象及处理技术的应用

数字图象处理一般是指通过数字计算机处理二维图象。更广义地讲,它可包括对任何二维数据的数字处理。这门学科的产生和迅速发展主要受三个方面因素的影响,一是计算机的发展;二是数学的发展,特别是离散数学理论的创立和完善;三是广泛的军事、工业和医学等各方面的应用。本世纪 20 年代,图象处理首次应用于改善伦敦和纽约之间的海底电缆发送的图片质量。直到本世纪 50 年代数字计算机发展到一定水平后,数字图象处理才真正引起人们的兴趣。1964 年,美国喷气推进实验室(JPL)用计算机对“徘徊者七号”太空船发回的大批月球照片进行处理,收到明显的效果。60 年代末,数字图象处理已经形成了比较完整的体系,形成一门新兴的学科。60 年代到 70 年代,由于离散数学的创立和完善,使数字图象处理技术得到迅猛的发展,理论和方法进一步完善,应用范围更加广泛。

这一时期,图象处理主要与模式识别、图象理解的研究相联系,如文字识别、

医学图象处理、遥感图象处理等。70年代后期至今,各个应用领域对数字图象处理提出越来越高的要求,促进这门学科向更高级的方向发展,特别是在景物理解和机器视觉方面,图象由二维处理变成三维解释。近几年来随着计算机和各个领域研究的迅速发展,科学计算机可视化、多媒体技术等的研究和应用,数字图象处理从一个专门领域的学科变成了一种新型的科学研究和人机界面的工具。

概括起来,数字图象处理主要应用于以下领域:

(1)宇宙探测:由于太空技术的发展,需要用数字图象处理技术处理大量的星体照片。

(2)遥感:分航空遥感和卫星遥感,这些图象需要用数字技术加工处理并提取有用的信息,可用于地形地质、矿藏、森林、水利、海洋、农业等资源调查,自然灾害预测预报,环境污染监测,气象卫星云图处理,以及用于军事目的的地面目标识别。主要进行图象的分割、几何校正等。

(3)生物医学领域:首先应用于细胞分类、染色体分类和放射图象等。70年代数字图象处理在医学上的应用有了重大突破;1972年X射线断层扫描(CT)得到实用;1977年白血球自动分类仪问世;1980年实现了CT的立体重建。

(4)工业生产:在工业生产领域的应用从70年代起有了迅速的发展,主要有产品质量检测,生产过程的自动控制,CAD/CAM等。在产品质量检测方面,如损伤、焊缝质量或表面缺陷的探测。又如金属材料的成分和结构分析,纺织品质量的检查,光测弹性力学中应力条纹的分析等。在电子工业中,可以用来检验印刷电路板的质量、监测零件部件的装配等。在工业自动控制中,主要使用机器视觉系统对生产过程进行监视和控制,如港口的监测调度、交通管理、流水生产线的自动控制等。在计算机辅助设计和辅助制造方面,已获得越来越广泛的应用,并和基于图形学的模具、机械零件、服装、印染花型CAD结合。目前,二维图纸的自动输入和理解,根据3-D实物建立CAD模型等越来越引起重视。

(5)军事:主要应用于军事目标的侦察、制导和警戒系统、自动武器的控制;公安部门的现场照片;指纹、手迹、印章、人像等的进一步处理和辨识,以及大量的历史文字和图片档案的修复和管理。另外,还包括其他方面图象信息的显示、记录、处理和文字自动识别等。

(6)机器人视觉:这是目前还处于研究之中的课题。主要应用于执行军事侦察、危险环境下的自主机器人,完成装配线工件识别、定位、太空中的自动操作的机器人,用于邮政、家庭服务的智能机器人。图象预处理是高级视觉不可分割的一部分。

(7)视频和多媒体系统:如目前电视制作系统广泛使用的图象处理、变换、合成,多媒体系统中静止图象和动态图象的采集、压缩、处理、存储和传输等。

(8)科学可视化:图象处理和图形学紧密结合,形成了科学研究各个领域新型的研究工具。

(9)新闻图象的传输:传输系统由压缩/解压终端、通信终端、卫星信道以及