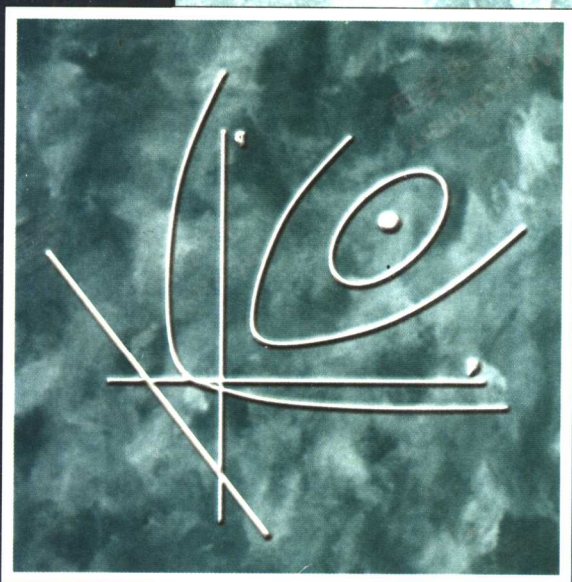




研究生系列教材

数字视觉 视频技术

李玉山 编著



西安电子科技大学出版社
<http://www.xduph.com>



研究生系列教材

数字视觉视频技术

李玉山 编著

西安电子科技大学出版社

2006

内 容 简 介

本书讨论了数字视觉视频处理技术, 主要内容包括数字视觉低层图像预处理技术与门类多元的中层及应用层视觉技术。其中, 运动视觉技术, 包括运动检测与分割、运动测量与分析、运动识别与跟踪、自主导航等是本书研讨的重点。其他内容有: 成像模型, 几何标定与不变量, 测距与立体视觉技术, 物体的二维和三维形状表征、纹理与彩色处理, 建模与识别技术等。此外, 本书还介绍了数字图像压缩以及与视觉处理相通的视频处理共性技术等。

本书借鉴了国内外涉及视觉和视频两个交叉领域的最新资料和相关研究成果。尽量突出物理概念, 简化数学推导, 导论性介绍研究思路, 其目的是使研究人员尽快熟悉研究领域的前沿。

本书可以作为信息与通信工程、电子科学与技术、计算机科学与技术、控制科学与工程等学科研究生及高年级本科生学习数字视觉视频技术的教材, 也可作为工程技术人员的参考书。

图书在版编目(CIP)数据

数字视觉视频技术/李玉山编著. —西安: 西安电子科技大学出版社, 2006. 1

(研究生系列教材)

ISBN 7-5606-1602-X

I. 数… II. 李… III. 视频信号-数字图像处理-研究生-教材 IV. TN941.3

中国版本图书馆 CIP 数据核字(2005)第 133043 号

责任编辑 潘恩祥 云立实 臧延新

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 邮 编 710071

http: //www. xduph. com

E-mail: xdupfxb@pub. xaonline. com

经 销 新华书店

印刷单位 西安文化彩印厂

版 次 2006年1月第1版 2006年1月第1次印刷

开 本 787毫米×1092毫米 1/16 印 张 19

字 数 447千字

印 数 1~4000册

定 价 26.00元

ISBN 7-5606-1602-X/TN·0321

XDUP 1894001-1

*** 如有印装问题可调换 ***

本社图书封面为激光防伪覆膜, 谨防盗版。

前 言

数字视觉视频技术是机器处理视频信息的主要方面,是一个一直处于动态变化中的学科领域。这一学科的内涵和外延尚未定型,技术走向也在形成和逐渐明朗之中。但是,它的应用前景非常诱人。专门从事高科技市场咨询的美国 Frost & Sullivan 公司部门经理 S. Ramachandran 预测,工业界机器视觉的市场将从 2005 年起迅速成长,其规模将会从 2002 年的 12 亿美元猛增到 2009 年的 26.2 亿美元。此外,还有多种从不同角度对这一领域市场变化的预测。显然,在充满活力和日趋明朗的发展势头中,北美、西欧和日本等发达国家则占有较大的先机。

编写本书的意图是尽量阐明数字视觉视频的技术范围和要点,希望这对国内从事本学科的科学研究和技术开发人员都能有所裨益。作为一本教材,本书的定位是面向工科研究生和科研人员的技术基础性导论读本,编写难度和压力都比较大。作者在努力理清视觉技术脉络的前提下,更关注内容的基础性、深度、广度和时效性。作为前沿性的学科研究,理所当然地应该涵盖在知识创新方面的最新进展。但是,最难处理的问题在于内容的取舍。面对浩瀚的新文献,必须区分哪些将会是具有划时代意义的基础性技术节点,哪些是应用型的技术分支。为了凸现研究成果的实际工程物理含义,书中把许多理论推导和公式演算都加以淡化处理。笔者恪守抛砖引玉之旨,诚存取长补短之心,冀收切磋砥砺之效,此外并无太高的奢望。我们力争本书出版后能有较长的生命周期,其中包括适时进行必要的修订更新工作。

毫无疑问,创新是推动学科发展的原动力。课堂讲授的目的应以培养本领域的创新思维为主。本书在讲述技术上比较成熟的内容时,尽量突出介绍研究人员最初的原创性精华和探索思路。同时,为后面预留了一定的讲座研讨题才,便于开放式地交流数字视觉和数字视频技术两个相互交迭领域的新成果,以期本课程成为一个“百家争鸣,启蒙启迪”的学术前沿舞台,吸纳不断推陈出新的研究动态新思想。

本书在原西安电子科技大学研究生教材的基础上进行了充实和删减。以西电的研究生教学实践为例,安排的课程总学时为 46 学时,包括最后一章较长的课堂研讨时段。学时的分配情况如下(供参考):

数字视觉视频技术引论(4 学时);第 1 章 生物视觉与人工成像(4 学时);

第2章 低层视觉图像预处理(6学时);第3章 图像分割与中层视觉处理(4学时);第4章 摄像机标定与几何不变量(2学时);第5章 运动视觉(6学时);第6章 测距与立体视觉(2学时);第7章 纹理与彩色处理(2学时);第8章 形状分析与模型描述(4学时);第9章 数字视频处理与编码(4学时);第10章 数字视觉视频技术进展(8学时)。

本书的出版获得了国家自然科学基金(No. 60172004)和教育部博士点基金(No. 20020701003、20050701002)及西安电子科技大学研究生精品教材系列项目的资助,也得到了西安电子科技大学出版社、教务处、研究生院、中国科学院西安光学精密机械研究所研究生部的大力协助。参加本书编写以及为写作提供帮助的教师和博士生有初秀琴、杨莉、刘洋、张大朴、李春明、张弘、侯彦宾、占志海、邱家涛、董巧玲等。此外,美国 IEEE Fellow 杨再元(T. Y. Young)教授、刘文泰(Wentai Liu)教授和英国 Teesside 大学的许东来(Donglai Xu)教授对本书的写作也给予了热情的指导,在此一并表示真诚的感谢!书中难免有不妥之处,殷切希望得到国内外同行和读者的批评指正。

作者

2005年12月于西安

通信地址:西安电子科技大学电路CAD研究所376信箱
电子邮箱:yshli@xidian.edu.cn

目 录

第 0 章 数字视觉视频技术引论	1	1.3.3 成像噪声模型	40
0.1 视觉视频研究概述	1	第 2 章 低层视觉图像预处理	41
0.1.1 数字视觉的内涵与外延	2	2.1 引言	41
0.1.2 视觉技术研究回顾	5	2.1.1 滤波分类	41
0.1.3 D. Marr 与视觉计算理论	6	2.1.2 滤波操作的对象	41
0.1.4 黑白与彩色图像	7	2.1.3 滤波的目的	42
0.1.5 数字视频技术	7	2.1.4 边缘检测预处理	42
0.2 视觉系统与处理技术	8	2.2 图像空间域过滤	43
0.2.1 数字视觉系统基本组成	8	2.2.1 模板匹配	43
0.2.2 视觉处理及系统设计技术	8	2.2.2 直方图变换	44
0.2.3 视觉基础理论和工程背景	9	2.2.3 中值滤波	49
0.3 视觉研究应用及展望	11	2.2.4 局域加权求和	51
0.3.1 主要应用领域	11	2.2.5 微分锐化	53
0.3.2 研究现状与前景	12	2.3 图像空频域滤波	54
0.4 教学计划与阅读参考	13	2.3.1 高通滤波	55
0.4.1 主要外文参考书	13	2.3.2 低通滤波	56
0.4.2 教材章节与课堂学时安排	14	2.3.3 同态滤波	57
0.4.3 考评方式与内容	14	2.3.4 其他滤波简介	59
0.4.4 参考阅读范围	15	2.4 边缘检测与估计	60
第 1 章 生物视觉与人工成像	17	2.4.1 边缘模型与检测准则	61
1.1 生物视觉机理	17	2.4.2 梯度与方向导数——空间	
1.1.1 视觉心理生理学	17	导数法之一	63
1.1.2 生物成像及预处理模型	22	2.4.3 拉普拉斯算子——空间	
1.1.3 格式塔原则	24	导数法之二	66
1.1.4 视感知的一般理论	26	2.4.4 LOG 算子	66
1.1.5 实验视觉心理学研究途径	27	2.4.5 DOE 算子	70
1.2 人工光学成像与基本模型	27	2.4.6 模板匹配法	73
1.2.1 人工光学成像	28	2.4.7 曲面拟合法	74
1.2.2 几何模型	29	2.4.8 非参数统计分类法	75
1.2.3 辐射与照度模型	31	2.4.9 一维边缘检测的假设检验	76
1.2.4 线性系统模型	35	2.4.10 子像素精度边缘估计	77
1.3 成像噪声	37	第 3 章 图像分割与中层视觉处理	79
1.3.1 CCD 成像过程	37	3.1 概述	79
1.3.2 成像噪声来源	38		

3.2 基于边界的分割	81	4.4.4 运动变换与坐标变换的合成	128
3.2.1 Roberts 边缘合成边界法	81	4.5 摄像机标定	128
3.2.2 边缘—边界的松弛算法	82	4.5.1 多种坐标系	129
3.2.3 动态规划边界生成	84	4.5.2 外参数标定举例	129
3.3 基于区域的分割	87	4.5.3 摄像机内参数标定	131
3.3.1 区域合并	87	4.6 投影变换不变量	133
3.3.2 区域分裂	88	4.6.1 几何不变量	133
3.3.3 分裂与合并	92	4.6.2 共线点的交比	134
3.3.4 区域生成的松弛算法	92	4.6.3 交比的“铅笔”形式	136
3.4 哈夫变换	94	4.6.4 不变量的讨论	137
3.4.1 检测直线的参数空间聚类技术	94	第 5 章 运动视觉	139
3.4.2 实用型哈夫变换	95	5.1 引言	139
3.4.3 哈夫变换的实现	96	5.1.1 运动视觉的定义	139
3.4.4 哈夫变换的推广	96	5.1.2 研究内容	140
3.4.5 哈夫变换直线检测实例	97	5.1.3 回顾、现状与展望	141
3.4.6 哈夫变换的利弊	98	5.2 三维运动的点对应法求解	141
3.5 角点检测	99	5.2.1 运动表征和坐标系选取	141
3.5.1 角点检测算法分类	99	5.2.2 非线性求解算法	143
3.5.2 SUSAN/MIC 算法	100	5.2.3 线性算法	144
3.5.3 筛选检测算法 RSUSAN	101	5.2.4 共面点情况	145
3.5.4 RSUSAN 中的滤波预处理	102	5.3 光流解析	146
3.6 图像形态处理技术	104	5.3.1 光流释义	146
3.6.1 集合的基本概念	105	5.3.2 线性约束方程	147
3.6.2 腐蚀和膨胀	106	5.3.3 光流求解之一	148
3.6.3 开启和闭合	107	5.3.4 光流求解之二	149
3.6.4 结构元基本类型	108	5.3.5 光流求解讨论	149
3.6.5 击中击不中变换	108	5.4 基于光流的行车临近分析	151
3.6.6 二值形态处理应用	109	5.4.1 光流图解	151
3.6.7 灰度图像的形态处理	111	5.4.2 基于散焦点预测前行擦撞	151
第 4 章 摄像机标定与几何不变量	113	5.4.3 临近物体的及时坐标分析	152
4.1 引言	113	5.5 基于光流的刚性运动求解	153
4.2 坐标系与齐次坐标	114	5.5.1 引理——刚体的距离守恒性	153
4.2.1 笛卡尔坐标系	114	5.5.2 刚性运动定理	154
4.2.2 仿射坐标系	115	5.5.3 刚体上点间的运动方程	155
4.2.3 齐次坐标表示	115	5.5.4 刚性全局运动速度	156
4.3 几何变换与成像几何	117	5.5.5 求解过程总结	157
4.3.1 射影变换、仿射变换、刚性变换	117	5.6 平面运动的测量与跟踪	157
4.3.2 透视、仿射及正射投影模型	120	5.6.1 算法分析说明	158
4.4 三维运动与坐标系的齐次变换	123	5.6.2 跟踪的实现	160
4.4.1 简单物体运动与坐标变换	123	5.7 Snake 模型和 Greedy 算法	161
4.4.2 三维运动的线性变换形式	124	5.7.1 Snake 模型的概念	161
4.4.3 三维坐标系变换的一般形式	127	5.7.2 能量函数的定义	161

5.7.3 动态规划及 Greedy 算法	162	8.3.1 形状结构与拓扑结构	216
5.8 运动物体外轮廓紧贴逼近	164	8.3.2 形状模型	217
5.8.1 引言	164	8.4 二维形状设计与描述	218
5.8.2 运动区域和初始轮廓生成	164	8.4.1 B-样条曲线绘制与拟合	218
5.8.3 多运动物体动态轮廓逼近	166	8.4.2 傅立叶描述子	222
5.9 非刚性运动分析	167	8.4.3 二维全局性描述	223
第 6 章 测距与立体视觉	169	8.5 三维形状设计与建模	226
6.1 测距及距离图像	169	8.5.1 三维曲线	226
6.1.1 测距	169	8.5.2 三维曲面	227
6.1.2 距离图像处理	172	8.5.3 广义圆柱体	229
6.2 双目视觉	172	8.5.4 三维全局性描述	230
6.2.1 立体图像几何	174	第 9 章 数字视频处理与编码	232
6.2.2 对极线方程与本征矩阵	178	9.1 视频概述	232
6.2.3 特征匹配	179	9.1.1 视频技术	232
6.2.4 相关问题	181	9.1.2 压缩标准列表	233
第 7 章 纹理与彩色处理	183	9.2 图像压缩编码技术	235
7.1 纹理分析与处理	183	9.2.1 图像压缩编码系统的组成	235
7.1.1 引言	183	9.2.2 编码中信息的度量——熵	236
7.1.2 空间灰度级相依矩阵	185	9.2.3 符号编码	237
7.1.3 空间灰度线段长矩阵	188	9.2.4 量化技术与无损编码	239
7.1.4 纹理度量中的其他测度	189	9.2.5 基于 DCT 的编码技术	242
7.1.5 纹理分割	191	9.2.6 预测映射与编码	244
7.2 彩色知觉与彩色基	191	9.3 静止图像压缩编码	245
7.2.1 彩色知觉与物理三基色	192	9.3.1 图像帧格式	245
7.2.2 计算三基色及色度图	193	9.3.2 静止图像压缩标准(JPEG)	246
7.2.3 彩色基及其转换	194	9.3.3 静止图像编码标准(JPEG2000)	248
7.3 彩色研究与处理	196	9.3.4 JPEG ASIC 硬件实现	250
7.3.1 彩色成像及不变性	196	9.4 运动图像压缩编码	251
7.3.2 彩色直方图及其变换	197	9.4.1 帧间压缩	251
7.3.3 彩色图像的低层和中层处理	197	9.4.2 视频编码中的运动补偿	252
7.3.4 假彩色与伪彩色	200	9.4.3 H. 261	253
第 8 章 形状分析与模型描述	201	9.4.4 MPEG-1	254
8.1 三维形状方向求解	201	9.4.5 MPEG-2 (H. 262)	255
8.1.1 由图像影调求解方向	202	9.4.6 MPEG-4 (H. 264)	256
8.1.2 由场景几何约束求解方向	206	9.4.7 MPEG-7 及 MPEG-21	258
8.2 线画标记与斜度定理	208	9.5 基于内容的图像检索	259
8.2.1 标记线画	208	9.5.1 CBIR 研究提要	259
8.2.2 斜度定理及应用	213	9.5.2 整幅图的外观表征	260
8.3 形状结构建模	216	9.5.3 图像的局部语义	261
		第 10 章 数字视觉视频技术进展	262
		10.1 视觉原理再探索	262

10.1.1	视觉研究的着力点	262	10.3.4	基于 IEEE 1394 标准的视频视觉系统	280
10.1.2	视觉处理的技术目标	263	10.4	视觉处理的实用化和实时性	283
10.1.3	场景恢复/描述的研究回顾	264	10.4.1	实用化研究平台	283
10.1.4	主动视觉和针对性视觉	266	10.4.2	实现实时性的难点与途径	284
10.1.5	小结与展望	268	10.5	视觉视频教材述评	285
10.2	成像芯片与并行处理技术	268	10.5.1	数字图像处理	285
10.2.1	数字视频成像芯片	268	10.5.2	计算机视觉	286
10.2.2	并行处理技术	270	10.5.3	机器视觉	287
10.2.3	SIMD/Systolic 架构	272	10.5.4	数字视频处理	288
10.2.4	数据通路	274			
10.2.5	视觉视频处理 ASIC/SOC 设计	275	参考文献		290
10.3	基于标准接口的视频视觉系统	276	一、中文专著		290
10.3.1	商用视频系统	276	二、中文相关论文		291
10.3.2	基于 PCI 的视频视觉系统	277	三、英文专著		292
10.3.3	基于 USB 标准的视频视觉系统	278	四、英文相关论文		294

第 0 章 数字视觉视频技术引论

0.1 视觉视频研究概述

我们把一般概念上的视觉细分为两个视觉感知阶段：感觉（perception）和知觉（cognition）。感觉就是对视觉信息数据的传感、采集、转换和变换。知觉又由两个过程组成，即对视觉信息内容的处理（processing）与推理（reasoning）。在知觉的处理与推理过程中，处理包括对感兴趣信息的提取和求解等视觉信号信息处理技术；推理则主要是根据已有的和新获取的多方面知识信息进行的高层次逻辑推理等智能活动，又称为高层视觉。我们的研究高度和重点将放在处理上，即视觉视频的信号与信息处理技术。

从本质上看，人工视觉视频处理的对象是数字化图像序列，所用的处理手段也是数字化的，即广义的数字信号处理技术。基于这项技术的对象和技术手段都是数字化的，所以称其为数字视觉视频（digital vision and video）处理是比较合适的。数字视觉仍然处在发展之中，该学科的建立打开了通向未知视觉领地的一扇门窗。每年，在这一领域都有数以万计的论文和成果面世。

数字视觉，这一名称具有更大的包容性。它强调用先进的数字技术分析数字化数据实现视觉感知，凸现多样化的技术实现途径。数字视觉可以涵盖计算机视觉，二者没有本质的区别。数字视觉既可以依赖计算机和处理器，也可以独立于计算机之外用各种数字系统和技术来实现。机器视觉更偏重于从应用角度来思考和实现视觉功能，也可以算作一种数字视觉载体。例如，对于独立的机器人来说，就是为其提供有视神经功能的眼睛。简言之，各种视觉研究的核心都是理解图像。

视频处理，指的是对以人类视觉观赏和传媒为主要目标的图像，特别是序列图像的信号信息处理。它主要涉及图像的存储与传输、数字电视、视觉图像通信等相关领域。

站在纯技术的角度，视频研究与数字视觉交迭的内容确实太多，我们可以将其看做是属于泛视觉技术的范畴。本书主要阐述视觉信号信息处理，在章节安排上以数字视觉技术为主线。由于许多内容同样适用于视频处理，因此我们将本课程称之为数字视觉视频技术。

数字视觉视频这门学科的迅速发展是近年来需求牵引和技术推动相结合的产物。

所谓需求牵引，指的是人们特别渴望实现视觉视频信息处理的自动化、数字化、计算机化。因为人类 80% 的知识来源于视觉信息，“百闻不如一见”，放眼一瞥，兆位量级的数据涌入眼帘。连续观察，其数据率则高达 1000 万位/秒，需要处理的信息实在是太多、太快了，而其他四种知觉（第六知觉除外）：听觉、嗅觉、味觉、触觉的信息量就相形见绌了，将后面这四种知觉的信息数据量加在一起还要比视觉信息少两个数量级以上。颇受好评的

“大唐芙蓉园”，号称是提供五种感觉的享受。显然，包括“梦回大唐”在内，主要还是“看一看”的视觉享受。这就意味着信息化社会和信息处理技术产业已离不开视觉视频技术，而其处理和推理的难度之高又非同一般。

技术推动主要与数字视觉视频数据的俯拾皆是和电子信息技术的迅猛发展息息相关。超大规模集成电路(VLSI)和计算机等软/硬件功能的迅速完善，数字信号处理器(DSP)及有效算法的实时实现为人工视觉视频功能铺平了道路。作为核心技术的数字视觉信号处理，已经从以数值计算为中心转向对数字化数据的处理。DSP理论和技术的超常进步正在改变着电子信息技术的走向。

0.1.1 数字视觉的内涵与外延

数字视觉(DV, Digital Vision)的完整含义是：传感设备对客观世界进行采集变换；实时数字系统通过算法对场景的视觉信息进行表征描述、建模重构、压缩存储、网络传输、处理分析(含理解识别)及逻辑推理，从而实现人类生物视觉(包括眼睛和大脑等)所具备的决策功能。它从所获得的二维图像中提取并生成对三维场景的明确、有意义的描述和解释等控制决策用视觉信息。它所涉及的采集、表征、存储、传输、处理、推理等六项技术都在不断地发展，但目前研究较多的还是视觉信号信息处理技术。

1. DV 与图像信号处理

数字视觉的基础是数字图像处理技术。

客观世界中包含时间在内是一个四维空间。各种变换(包括成像)将其不可逆地退化为三维、二维和一维的时间-空间数据。

有时，人们可能会自觉地作降维处理。例如，将四维空间信号生成电视信号，实质是将其转化为沿时间轴的一维信号来处理。即使一般的二维图像信号，也常常会退回到一维空间来处理。

一维、二维甚至多维信号处理，立足于信号理论、信号处理理论，是一种狭义的信息处理。库利·图基于1965年发明的快速傅立叶变换(FFT)对信号处理的发展有较大的贡献。它使得一维FT的计算量由 $O(N^2)$ 降低到 $O(N \lg N)$ ，效果非常明显。对于二维图像信号，计算量则由 $O(N^3)$ 降为 $O(N^2 \lg N)$ 。不过，在采用频域处理图像时，由于需要正变换(FFT)加上反变换(IFFT)，因此靠FFT技术改善速度反而并不明显，常常达不到应用中所需的实时性。

图像信号处理是一种特殊的二维信号处理，它把输入的图像信号根据需要处理成输出图像信号，起初的主要着眼点是为人们提供改善后的图像信号。其处理包括数字化、压缩、编码、分割、增强、恢复、重建。在图像处理深入发展的后续过程中不断出现新的更高追求，这些一再推动着该领域的技术拓展，进一步涵盖了图像描述、图模匹配和识别等，这就是图像理解。

可以认为，数字视觉是在图像理解概念上的进一步外延与升华。它明确提出用计算机或者机器实现类似人类的视觉功能，这种名称界定使得学科内涵比较具体、科学和深刻，也更加准确和概括。它所包含的理解、处理、思维、推理能力已经开始进入人工智能的范畴，成为一种多学科交叉的边缘学科。

总之,信号处理和图像处理是位于萌发和催生数字视觉技术的底层和起始端。但 DV 技术的继续发展使它并没有停留在一般意义上的多维信号处理。除信号处理之外, DV 更多地包含了特殊的决策用视觉信息处理技术,例如运动检测和跟踪等。数字视觉与图像处理也有一个明显的差别,就是视觉的输出量常常不是图像,而是描述性参数或者结构性模型。

2. DV 与计算机图形学

图像处理重在“像”,图形学重在“形”。数字视觉的许多应用需要将图像与图形结合起来,因此视觉和图形学都要研究形状建模和投影的几何关系。

计算机图形学研究图形的几何关系,研究图形的产生、存储、处理、显示、测量等技术。它根据几何原理,采用图形数据库技术,通过基元,例如直线、曲线、平面、曲面、立方体、圆柱体等来生成图形和图像。这种图像尽在人的主动掌握和理解之中,它主要用于 CAD、动画等领域,侧重点在于综合。

数字视觉则侧重于形状分析,因为它的图像被动地源自不合作的场景变动世界。人们往往通过图像再到图形来实现物体识别,例如对图像进行边缘检测后生成二值图像,与图形数据库中的二值图形进行匹配运算以实现识别。

3. DV 与仿生视觉理论

显然,就机器视觉技术实现途径而言,数字视觉≠摄像机(机器眼)+处理器+推理机。对于纯粹的视觉功能和机制,重点在于眼睛后面的处理和推理的结构和机理。对此,人们认知得并不充分。

人们潜心研究生物视觉机理,一个主要的目的就是希望能借鉴到数字视觉的研究中来。从前因为需求较小、难度又大,所以进展缓慢。20 世纪 80 年代初,从事研究生物视觉机理的人员荣获了诺贝尔奖(1981 年)。21 世纪出现三个最有活力的研究领域:生物生命科学、信息技术和材料科学,视觉技术涉及到其中的两个,而且又是其中的重点内容。人们常说世界上有三大科学之谜,即物质、生命和大脑。可以说大脑中的处理与推理,包括视觉功能在内,称得上是谜中之谜。对于人脑的认知,有点像是“悖论”。不禁使人联想到人们能较多地认知天文,可对于人类栖息的地球却缺乏足够的认知。人对于自己的智能活动,有点像“不识庐山真面目,只缘身在庐山中”。

下面粗浅地介绍一下关于人类视觉的几个外在表现:

(1) 视觉必定是一类处理(process)和推理过程,它要从图像中发现外部世界中自己感兴趣和需要关注的信息。

(2) 视觉必然是以某种方式来表征(representing)多种不同的信息,包括位置、形状、运动、类别、美感和其他等。

(3) 视觉具有学习记忆、存储和数据库管理机制。同样的景象看了几遍之后就能记住;几十年的老朋友、老同学能够迅速认出。我们好奇的是这些视觉信息在脑子里是怎么记忆和组织存储的。

(4) 视觉的联想推理能力。人类可以从某个图像联系到其他看似不相关的图像。例如陈毅等人从下棋可以联想到战争,或者从战争联想到下棋。

(5) 视觉中最奇妙的是创造能力。所谓“眉头一皱,计上心来”,人类具有这一独有的能力,机器智能却对此望尘莫及。现有的图形 CAD 技术,极大程度上辅助了人的脑力劳

动,但是仍然代替不了人们在图形方面的创造性能力部分。

人们对视觉机制的认知显然还比较肤浅,大多数仍然说不清楚。对于其中的处理过程,例如直线检测、角点检测等是属于没有察觉的本能类型。人们已经在用核磁共振技术探究大脑的思维过程及其伴随发生的生理活动。显然,我们首先需要借鉴和有可能理解的是视觉处理类技术,例如“感受野”等。

关于其中的视觉推理过程,人们似乎从中获得了比较多的可行性感受。但是,人们仍然是几乎不明白大脑中算法的奥妙。

归纳起来,以往的数学是研究机器的科学,而不是关于人脑的科学。维纳曾一度试图研究人脑,进而统一对自然和社会的理性认知,在人机控制论(cybernetics)方面取得了长足的进展。至今,“生物控制论”和“模糊数学”可以算是往这方面发展的有力分支,但其具有的功能仍然远远不够。

在人们研究数字视觉系统感到无助时,常常不自觉地与人眼的功能相对照进行借鉴。例如人们根据生物神经元的启示,已经着手研究“轴突晶体管”。它是多输入单输出,而且它的输出与过去的输入,或者说当前的状态有关。

生理学家 Gibson 在 1979 年就指出,感知是一种行动(action),而不仅仅是一种被动式的反应(response)。这是否更多地意味着在信息处理中主观主动、交互参与、联想机制、自适应随机应变的重要性?

研究仿生学指导人工视觉的原则之一是眼能做到的,处理器或数字化机器也应该设法做到;指导原则之二是处理器应该做到一些普通人所做不到的事,例如精确测量等。可以想像人脑虽然经过上亿年的进化,但作为基本上是一个固定连接的硬件系统,难免有些局限性。例如人们在口述相互间距离时,常用描述性的模糊概念,而不是用厘米和毫米等精确概念来描述。

4. DV 的多层次视觉处理

在数字视觉的学术界,人们已经习惯将 DV 分为高层(high level)和低层(low level)。一种流行的分类法是,再加上中间层(intermediate (middle) level);更有甚者的一种说法是把广义中间层再细分为狭义中间层和面向应用层(application level),本书也采纳了这种提法。

一般来讲,现有信号处理理论适用于低层处理,而高层处理需用数字人工智能理论与逻辑推理来实现。中间层和应用层处理介于低层和高层之间,主要是一种视觉信号信息处理技术,是目前研究的主流,着眼于研发大量急需的能够解决实际问题的技术途径。

low level 视觉主要研究涉及像素一级的运算操作,包括成像、滤波、边缘检测和阈值区域分割等。

狭义的 intermediate level 主要是研究线、圆、椭圆、洞、多边形、角点等形状的检测,比较突出的技术是哈夫(Hough)变换。广义中间层再细分出的 application level 处理研究的是二维形状模板匹配、影调和遮挡,纹理、彩色、距离、运动,以及几何不变性等。广义中间层根据区域的性质和特征,找出区域间的关系,推导出相应的三维场景描述。随着应用场合的不同,要特别关注并应用其隐含的不同先验知识。

high level 视觉主要研究普遍适用的三维场景的解释、理解和物体识别理论技术,研究通用的逻辑推理等智能推理甚至决策过程。它与模式识别、人工智能融为一体,神经网络

络就是其中的一个分支。现在的专家系统、神经网络就是模式识别理论的进一步深化和发展，触及到了 DV 的高层次内容。

研究信号处理的人们习惯于用统计信号处理技术来研究图像，在低层进行的研究已经给出了许多优秀成果。但是在高层，对于一个“自然人”很容易实现的事，比如儿童都具有的关于“狗”的定义、识别，用信号处理理论则无能为力，因此必须采用更有效的逻辑推理技术来实现。

0.1.2 视觉技术研究回顾

从 20 世纪 60 年代 Roberts 的研究开始算起，对人工视觉的研究至今已经过去了大约半个世纪。

比较有代表性的早期对图像理解的研究，是美国军方长期支持、资助的结果。众所周知的国防高科技计划局(DARPA, Defense Advanced Research Projects Agency)催生出丰硕的高档科研成果。例如 20 世纪 90 年代初，美国国防部曾出资 7000 万美元资助 350 个神经网络(neural network)等基础研究单项计划。由国防部资助南加州大学完成的图像理解计划是数字视觉早期研究比较成功的项目。在美国，只有拥有先进设备、技术和理论领先的高等院校、研究所或教授、科学家，才可望获取国防部的资助，类似于我国的 863、973 等高科技计划一样。

40 多年来信号信息处理理论的进展，特别是处理器实践的雪崩式发展，激发出人们开发并利用它的热情和企望。尽管 20 世纪 20 年代人类已经第一次实现了纽约与伦敦之间的跨洋新闻图片传输，但当时没有产生出高性能/价格比的处理器，人类的视频通信愿望基本上仍然是“南柯一梦”。然而研究者们并不气馁，从 20 世纪 60 年代至今近半个世纪以来，数字化处理和互联网络、处理器技术的结合与进展使许多涉及图像类的美梦成真，同时也吸引大批的科技精英为之奋斗。

尽管国家、工业界投资于图像处理、数字视觉视频技术的研究经费非常有限，但我国许多高校及科研机构中，从事这方面研究的教授、博士生有增无减。国内许多知名大学的计算机系、自动控制和电子工程系可能同时有多个课题组在从事视觉研究，其中也不乏一些数学专业出身的研究人才。大致说来，其人员组成主要来自具有两种背景的科技队伍：

(1) 电路与系统、信号与信息处理、自动控制学科、计算机工程(国外都归为电气工程——EE)的电子技术工程类师生。

(2) 基础和基础科学，包括计算机科学及应用数学专业在内的理论研究人员。

由此引发出这两种人之间的路线争论。马里兰大学的大师级人物 Rosenfeld 教授等人所持的观点有一定的代表性：“许多批评图像处理缺乏理论高度的人们，自己也未能依据某种理论示范性地导出高质量的图像”。这确实是一种挑战，它使人想起 20 世纪 40 年代初大批物理学家改行研究雷达。其直接的刺激因素是军事，正是第二次世界大战需求的牵引，结果使得雷达理论水平飞速提高，上升到一个新台阶。现在国外的科研现状是受高科技发展计划和项目的引导，促使许多不同领域的人才改行从事数字视觉研究。坦白地说，上述两方面的人才都有自己的优势特长和发展空间。

人们早已仰慕飞鸟翱翔天空的自由，经历过数次失败的尝试后便发明了飞机，以及宇宙飞船(太空梭, space shuttle)。它们的原理不同于飞鸟，但他们的功能，包括高度、速度、

距离等都远在飞鸟之上。那么能否指望类似的情况在数字视觉的进程中重演？这恐怕也是 D. Marr 所创立视觉计算(又称计算视觉)理论的初衷,而 D. Marr 的早逝使理论的进展严重受挫。因此数字视觉技术,正如有些科研工作者所言:它是一个先天不足的早产婴儿。

虽然属于早产,但是数字视觉这一“婴儿”正在后天的需求和实践的环境中迅速长大。目前,以处理器为工具,用一套不同于人脑的表示法和推理过程,利用数字视觉技术可以在个别、特殊、局部、专用的情况下,实现部分人类视觉功能,这已是不争的事实。但是如果将该成果推而广之,是否仍然适用?显然,至今还没有一个普遍适用、全面完整的视觉理论形成。

再者,处理器以其高速计算、信息处理和某种固定的逻辑推理见长。IBM 的“深蓝(deep blue)”存储了多种棋谱,每秒搜索计算可达两亿步,采用“蛮力”,才能与国际象棋世界冠军卡斯帕罗夫下成平局。而以色列的“小深(deep junior)”,虽然每秒只能搜索 300 万步,但它靠评价函数、理性推理等“智力”技术,也能与卡斯帕罗夫下成 3:3 的平局。这无疑已显露出人工视觉具有别开生面的可喜曙光。

在美国学术界位于视觉研究前沿的先驱人物有: D. Marr、K. S. Fu(傅京孙)、D. H. Ballard、A. Rosenfeld、T. S. Huang(黄煦涛)、T. Pavlidis、K. Fukunaga、R. C. Gonzalez、R. M. Haralick、T. O. Binford 和 L. G. Shapiro 等。

在视觉研究领域一直处于领先地位的大学有:麻省理工学院(MIT)、斯坦福大学(Stanford)、加州理工学院(California Institute of Technology)、卡内基梅隆大学(Carnegie-Mellon)、南加州大学(Southern California)、马里兰大学(Maryland)、伊利诺大学(Illinois)和普度大学(Purdue)等。

0.1.3 D. Marr 与视觉计算理论

计算视觉研究体系由 D. Marr 首创(见科学出版社 1988 年版“视觉计算理论”),他指出视觉是一种计算过程,其目的是从图像中得知外部场景中有什么物体以及它们的空间关系。他将视觉研究划分为三个创新层次:计算理论、表征与算法和系统实现(含硬软件设计)。

D. Marr 提出的视觉计算理论仍在发展。考虑到计算与处理的差异,是否称之为视觉处理理论更为合适?在理论研究成果的基础上,算法是实现进一步创新的中间环节。最后,根据算法进行系统硬软件结构和电路设计,同样是一项高难度的创造性工作。

在 D. Marr 所描述的计算视觉整体框架中,他又把从图像中推导出形状信息算法的过程分为三个表征阶段:

- (1) 二维图像性质的表征。
- (2) 以观察者为中心的坐标系中的物体可见表面等性质的 2.5 维表征。
- (3) 以物体为中心的坐标系中被观察物形状的三维结构和组织的表征。

也许人们长期保存的是第三种形式的表征,它是按照(1)→(2)→(3)的顺序进程得来的。

数字视觉本质上是根据不完整信息来恢复完整信息的,其解当然不是惟一的,因此它是一个“病态”问题。总之,实现视觉功能总体就是一个估计和优化问题。

0.1.4 黑白与彩色图像

到目前为止,黑白图像是数字视觉的主要研究对象。黑白图像有灰度图像和二值图像两种。目前有多种标准格式,我们主要指的是灰度裸图像,例如 $256 \times 256 \times 8$ bit 图像。人眼分辨最多为 6 bit;高分辨率的采集系统可以达到 9~12 bit,这对于某些场合是有益的,例如边缘检测时的子像素精度和边缘方向的求解。但是,一般数字存储还是以无符号的字节(byte, 8 bit)为单位比较方便,用灰度 255 表示最亮,用 0 表示最暗。图像的像素坐标系 (i, j) 一般以左上角为坐标原点 $(0, 0)$ 。

彩色视频图像已经越来越普及,这也预示着学科将来的研究方向。它提供了更多的信息,分辨率较高,也使得某些视觉任务更容易实现。例如检测绿叶中的红花时,人眼识别起来非常方便。如果是黑白图像的红花绿叶对人眼来说将是很难检测的,只有在彩色图像中才能更容易完成。但是,现在有些人类视觉中的彩色存储及视觉功能还没解释清楚,因此人工机器视觉暂时对于彩色也就没有什么太多的新技术问世。目前,人工彩色视觉需要更复杂的算法和更多的存储空间(例如 $8 \text{ bit} \times 3/\text{pixel}$),为此需要付出较高的代价。所以当前的数字视觉还是以黑白图像为主,而彩色视觉处理也必是将来的一个发展趋势。

0.1.5 数字视频技术

视像(video),俗称视频,主要指不断变化的电视图像,即动态图像,或称活动图像、运动图像,研究对象以彩色为主。最初是指电视中相对于音频(audio)而言的视频信号,原本是一个模拟信号的概念。但随着数字电视的大趋势,它的主要技术内容和难点与数字视觉已经相互交迭。

狭义的数字视频技术,或者计算机视频技术研究,涉及的内容有如下几方面:视频信号的描述、采集、数字化、编码、压缩、传输、存储、显示等。基本上是对视频数据自身的加工处理,它不包括太多对信息内容的理解。

发展中的视频图像处理技术领域不断扩大,包括对视频流中预期目标的实时检测;或者对数字视频录像带中特定感兴趣特征的检索;将场景视频流中的自然人与人工制作的猫、太空人合成;视频数据库管理中的摘要生成——将视频流分割成有代表性的镜头等。

视频技术的一个传统核心领域是压缩编码。国际上已经制定了静态/动态图像的压缩标准,包括 JPEG、H.26X、MPEG 等系列。以压缩解压缩为核心,视频研究的内容还包括:图像格式、彩色基与彩色编码、特征提取、图像分割、光流(optical flow)与运动估计、传送、存储、显示等技术。

这些主要的视频技术与图像处理、数字视觉中普遍采用的技术可以共享。视频是视觉的基础和对象,视频技术与视觉技术是相通的,所以统称为视觉视频处理技术更为合适。将视频技术与视觉技术打通研究,有助于形成交叉融合,实现新的突破。

近年来,许多从事数字视觉视频、视觉生理心理学、模式识别与人工智能、计算机图形学等方面研究的学者开始合作探索视频处理与压缩方案。比较活跃的有:整数 DCT 变换编码、小波变换编码、分形图像编码、模型基图像编码等。目前,许多涉及视觉视频技术相互融合的问题还有待于继续探索。例如对日渐庞大的视频数据库如何管理,包括摘要编写、内容修改和目标检索技术;对于场景的剧烈变化,包括相机和物体各自高速运动或者

场景迅速切换时的有效编码技术；对非刚性运动的估计和补偿问题等。

0.2 视觉系统与处理技术

0.2.1 数字视觉系统基本组成

在人类的视感知系统中，包括的器官有：作为光学传感器、光/电转换器和信号预处理器部分的眼睛，特别是视网膜；充当处理器、推理机和控制决策机构的大脑，特别是大脑皮层的视区。

数字视觉系统的主要部分可以分为两类，即感觉部件和知觉部件。

(1) 感觉部件：可见光或非可见光的传感器；转换成模拟或数字电信号的转换器。

(2) 知觉部件：用信号信息处理器实现决策用视觉信息的加工提取部件；通过逻辑思维与推理指挥输出与控制的智能推理部件。

此外，还可能有一些辅助部件，例如距离测量等。

在通用图像传感器部件中，其核心是 CCD 或 CMOS 成像芯片。目前已经可以将模/数信号转换器部分集成在同一个成像芯片内。成像芯片将模拟信号转换成二维数字矩阵输出，同时还可以进行亮度、对比度、饱和度、白平衡的自动设置并完成必需的前置预处理。完整的摄像头还要有快门、光圈、变焦、云台等伺服控制，甚至是距离测定功能等。

随着传感器的不同，人工视觉所处理的不仅仅是可见光成像。人工视觉的光源也可以有所选择，包括激光、超声波、X 光等都不能排除。在光源配置方面，还可以有结构光源、平面光源、特殊光源等多种选择。研究对象可以是雷达 SAR 成像、CT(或 CAT, Computerized Axial Tomography, 计算机轴向 X 光断层扫描成像)三维成像、核磁共振成像(MRI, Magnetic Resonance Imaging)、超声(声纳)成像、红外成像、柔性光纤传像束成像等。在各种工业应用中，人们最关心的是：不管用什么方法，只要能解决问题就行。通常，建造专用数字视觉系统的原则是：“不管白猫黑猫，逮住老鼠就是好猫”。目前，国内外都已经将非可见光数字视觉作为专题开展研究。因此可以预期，有朝一日数字视觉系统很有可能超过人眼。

广义的处理器部件，指的是数字处理系统，其中可以将计算机、处理机或 VLSI 专用机等都包括在内。除了作为硬件核心的 CPU/DSP/VLSI 芯片之外，也包括强大的底层支撑和高层处理软件在内。在外围的输出与控制部分，包括各种接口、控制器、监视器(执行装置)等。

0.2.2 视觉处理及系统设计技术

1. 视觉处理技术

视觉处理技术主要是研究算法层的视觉信号信息处理技术，这是本书内容的主体。包括以下几个方面：

(1) 边缘检测、滤波等低层预处理。进一步，低层视觉还包括边界提取、区域生成等。D. Marr 的视觉计算理论研究亦是自此开始的。