

计算摄影学基础

Computational Photography

张茂军 刘煜 王炜 张政◎著



科学出版社

(TP-6794.0101)

www.sciencep.com

ISBN 978-7-03-042094-7



9 787030 420947 >

定价：48.00 元

销售分类建议：图像处理

计算摄影学基础

张茂军 刘煜 王炜 张政 著

科学出版社

北京

内 容 简 介

本书将计算摄影学所涵盖的主要技术囊括其中。本书全面地介绍了计算摄影学中各种技术的基础知识,清晰地阐明了各基础知识的原理,以及目前最先进的技术成果,并对计算摄影学的未来进行了分析。本书特别关注计算摄影技术在工程中的实际应用,将各种技术应用后的效果作为实例呈现给读者,在加强理论和实践关联的同时,也让读者更直观地感受计算摄影技术给我们生活带来的改变。

本书可作为计算机视觉、图像处理、机器学习等领域的研究生教材,也可供计算机视觉领域科研人员、摄像机研制与生产相关工程技术人员以及数字摄影爱好者阅读。

图书在版编目(CIP)数据

计算摄影学基础/张茂军等著. —北京: 科学出版社, 2014

ISBN 978-7-03-042094-7

I. ①计… II. ①张… III. ①计算机应用—摄影学 IV. ①TB81-39

中国版本图书馆CIP数据核字(2014)第230142号

责任编辑: 陈晓萍 / 责任校对: 马英菊
责任印制: 吕春珉 / 封面设计: 北大彩印

科学出版社 出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

*

2014年9月第一版 开本: 787×1092 1/16

2014年9月第一次印刷 印张: 15 3/4 插页: 12

字数: 350 000

定价: 48.00元

(如有印装质量问题, 我社负责调换 < >)

销售部电话 010-62142126 编辑部电话 010-62138978-2009

版权所有, 侵权必究

举报电话: 010-64030229; 010-64034315; 13501151303

前 言

计算摄影学 (Computational Photography) 是由美国哥伦比亚大学、麻省理工学院、斯坦福大学等几个国际知名学府的教授发起创立的一门新兴学科, 它是在数字摄影学 (Digital Photography) 的基础上, 增加对成像过程与设备的精确控制、对影像的更复杂处理与计算, 并使用 3D 成像、全景处理等新颖的数字媒体处理技术, 从而在摄像过程中彻底引入计算单元的一套理论、方法与技术。当前, 计算摄影技术已经成为国际上的研究热点, 斯坦福大学、哥伦比亚大学、卡内基梅隆大学、麻省理工学院、北卡罗来纳大学等均成立了相关研究机构, 开设了相关研究生课程。在国内, 已有清华大学、浙江大学、国防科学技术大学、西北工业大学、西安电子科技大学等学术单位做了大量前沿性学术研究。

我们从事计算摄影学的相关研究 20 多年, 特别是近 10 年来, 从镜头设计、编码孔径设计、CMOS 的控制成像、颜色恒常性到后端的智能分析算法等都进行过较为深入的研究, 对计算摄影学相关研究领域有一定的理解, 并负责建设国防科学技术大学慕课课程: “计算摄影: 艺术与技术的交融” (课程视频可以从 “http://www.visionsplend.com/service/?catalog_332_1.html” 下载观看)。本书是对我们多年研究工作的总结, 同时尽量考虑覆盖计算摄影学涉及的研究领域, 所以在我们的研究工作的基础上做了适当的扩展, 虽然如此, 还是有很多计算摄影学相关研究领域在书中没有涉猎。

本书共 14 章。第 1 章讲述了计算摄影学的基本概念、主要研究领域, 并对计算摄影学未来发展方向发表一点看法。第 2 章总结了传统摄影学与数字摄影学的一些基础知识。第 3 章讲述了颜色与颜色空间。第 4 章概述了数字感光器件的基础知识。第 5 章论述了自动聚焦技术。第 6 章介绍了自动曝光和自动白平衡技术。第 5 章与第 6 章共同构成了摄像机的基础性技术——3A 技术。第 7 章介绍了高动态范围成像技术。第 8 章~第 10 章是与全景图像处理相关的技术介绍, 其中, 第 8 章总述了全景成像相关理论与方法, 第 9 章介绍了图像拼接技术, 第 10 章介绍了图像融合技术。第 11 章对前沿性的光场成像技术做了概括性介绍。第 12 章介绍了图像去噪技术。第 13 章对前沿性的压缩成像技术做了概括性介绍。第 14 章介绍了模糊图像复原技术。

第 1 章、第 2 章和第 4 章由张茂军教授撰写; 第 7 章、第 11 章、第 12 章和第 14 章由刘煜博士撰写; 第 3 章由王伟教授撰写; 第 5 章由张政博士撰写; 第 6 章由谭鑫撰写; 第 8 章由左承林撰写; 第 9 章由彭杨撰写; 第 10 章由李靖撰写; 第 13 章由娄静涛博士撰写。全书由张茂军统稿。

实验室熊志辉副教授、张军教授、包卫东教授、徐玮副教授、谭树人副教授、翟永平博士以及博士生赖世铭、陈旺、高辉、陈立栋、李乐、周韬、石崇林、王斌、李永乐、肖文华等, 硕士生尹晓晴、石皓、吕济民、王媛媛、夏青等, 他们的研究成果直接或间接为本书提供了素材, 也正是由于他/她们的多年共同努力, 为实验室积累了丰富的研究成果, 才有了本书的撰写成稿。计算摄影技术工程研究中心的詹万政、杨健群、韩金军、

刘剑平、陈捷、胡志、彭继湘、龙猜、刘镖、王博、段稳、黄锋、邓建辉、穆长富、刘电芝、刘述杰、周博将、车阳、彭素波、罗文峰、杨佳、朱可瑀、贺云飞、张弘强、李飞等，他们每天加班加点开展研发工作，其研发成果为我们进一步探索计算摄影学理论提供了技术基础。感谢西南财经大学母睿同学在高动态成像技术方面分享的研究成果。感谢微软亚洲研究院袁路博士在图像去模糊技术研究方面对我们做出的指导性工作。感谢欧阳合博士、周经伦教授为我们在光场成像技术方面的研究贡献了大量有价值的见解。

感谢我的恩师胡晓峰少将，是他在 20 世纪 80 年代末带领我们进入科学研究殿堂。感谢浙江大学的潘志庚教授，北京理工大学的王涌天教授，西北工业大学的王庆教授，西南财经大学的段江教授，国防科学技术大学的查亚兵教授、张维明教授，他们在学术上对我们的帮助与鼓励始终是我们不断前行的动力。

感谢国家自然科学基金委的支持，国家自然科学基金项目（No. 61403403, No. 61402491, No. 61175006, No. 61175015, No. 61271438, No. 61275016, No. 60803101, No. 60872150）的资助是我们实验室早期主要的经费来源。

由于计算摄影学属新兴学科，许多概念与理论尚处于讨论与探索阶段，本书的许多观点基于我们个人的认识与实践，难免有诸多疏漏、错讹与片面性，恳请读者不吝赐教。

目 录

前言

第 1 章 概述	1
1.1 计算摄影学的起源	1
1.2 计算摄影学的相关概念	2
1.3 计算摄影学涉及的研究领域	4
1.4 计算摄影学的未来	8
参考文献	11
第 2 章 摄影学基础知识	12
2.1 传统摄影学基础	12
2.1.1 摄影是用光的艺术	12
2.1.2 镜头与焦距	14
2.1.3 快门与曝光	17
2.2 数字摄影学基础	20
2.2.1 数码相机的基本原理	20
2.2.2 图像分辨率	22
2.2.3 图像信号处理器	22
2.2.4 压缩与存储	23
2.3 摄影学的技术性与艺术性	23
参考文献	24
第 3 章 颜色与颜色空间	25
3.1 颜色知觉与可见光	25
3.2 颜色视觉理论	28
3.3 颜色感知特性	30
3.4 颜色空间	32
3.5 基于彩色滤镜阵列的彩色感知	37
3.5.1 Bayer 格式的 CFA 彩色滤镜阵列设计	37
3.5.2 其他创新的 CFA 彩色滤镜阵列设计	38
参考文献	39
第 4 章 数字感光器件	41
4.1 数字感光器件的发展历程	41
4.2 数字感光器件的结构	42
4.2.1 CCD 的结构	42
4.2.2 CMOS 的结构	46
4.3 数字感光器件的指标	49

4.3.1	感光度	49
4.3.2	成像质量的衡量指标	50
	参考文献	51
第 5 章	自动聚焦技术	53
5.1	概述	53
5.2	对比度检测自动聚焦	54
5.2.1	聚焦值计算	55
5.2.2	聚焦峰值搜索	64
5.3	图像模糊度评估	67
5.3.1	模糊度与锐度的关系	67
5.3.2	基于视觉模糊阈的模糊度量	68
5.4	总结	71
	参考文献	71
第 6 章	自动曝光与自动白平衡技术	72
6.1	概述	72
6.1.1	自动曝光与自动白平衡的必要性	72
6.1.2	曝光的基本概念	72
6.1.3	白平衡的基本概念	75
6.2	自动曝光技术	77
6.2.1	常用测光技术	79
6.2.2	更先进的自动曝光控制	80
6.3	自动白平衡技术	87
6.3.1	基于图像统计特征的方法	88
6.3.2	基于学习训练的方法	93
6.4	总结	98
	参考文献	99
第 7 章	高动态范围成像技术	102
7.1	概述	102
7.2	HDR 技术的处理机制和步骤	103
7.2.1	HDR 图像的获取简介	103
7.2.2	全亮度图的合成简介	103
7.2.3	色调映射技术简介	104
7.3	HDR 图像的获取	104
7.4	全亮度图合成	106
7.4.1	照相机响应曲线的估计	107
7.4.2	HDR 合成图像存储格式	108
7.4.3	照度图的合成方法	109
7.5	色调映射技术	111
7.5.1	全局色调映射算法	111

7.5.2 局部色调映射算法.....	115
7.6 总结与展望	119
参考文献.....	120
第 8 章 全景成像技术.....	122
8.1 概述.....	122
8.1.1 全景图像.....	122
8.1.2 全景成像方法.....	123
8.2 拼接式全景成像	124
8.2.1 基于单相机旋转拍摄的全景成像.....	124
8.2.2 基于多相机同时拍摄的全景成像.....	125
8.2.3 投影算法.....	126
8.3 鱼眼全景成像.....	134
8.3.1 鱼眼镜头.....	135
8.3.2 成像投影模型.....	136
8.3.3 鱼眼图像的校正算法.....	138
8.4 折反射全景成像技术	141
8.4.1 折反射全景成像原理.....	141
8.4.2 单视点折反射全景成像.....	143
8.4.3 全向图像的展开.....	145
8.4.4 互补结构折反射全景成像系统.....	146
参考文献.....	149
第 9 章 图像拼接技术.....	150
9.1 概述.....	150
9.2 图像配准.....	151
9.2.1 基于特征的图像配准.....	152
9.2.2 基于区域的图像配准.....	154
9.3 最佳缝合线查找	154
9.3.1 静态图像的最佳缝合线查找.....	155
9.3.2 视频拼接的最佳缝合线查找.....	156
参考文献.....	164
第 10 章 图像融合技术.....	165
10.1 概述.....	165
10.2 加权平均融合法.....	165
10.3 金字塔融合法.....	167
10.3.1 算法思想	167
10.3.2 算法流程.....	168
10.3.3 融合结果.....	169
10.4 梯度域融合法.....	170
10.4.1 算法思想.....	170

10.4.2	算法流程	170
10.4.3	融合结果	171
10.5	结构变形	172
10.5.1	算法思想	173
10.5.2	算法流程	173
10.5.3	融合结果	175
	参考文献	176
第 11 章	光场成像技术	177
11.1	概述	177
11.2	光场的数学定义	177
11.3	照相机阵列光场成像技术	179
11.3.1	照相机阵列的结构和图像采集	179
11.3.2	照相机阵列的数字重聚焦原理	179
11.4	微透镜光场成像技术	184
11.4.1	微透镜光场照相机计算成像原理	184
11.4.2	微透镜光场照相机的数字重聚焦原理	185
11.5	总结与展望	186
	参考文献	187
第 12 章	图像去噪技术	188
12.1	概述	188
12.2	图像去噪的基本概念	189
12.2.1	图像噪声的分类	189
12.2.2	去噪效果评价指标	191
12.3	传统去噪算法	192
12.3.1	基于空间域的中值滤波	192
12.3.2	基于小波域的小波阈值去噪	192
12.3.3	基于 PDE 的图像去噪	193
12.3.4	全变分图像去噪	194
12.4	非局域均值去噪算法	195
12.4.1	NLM	195
12.4.2	BM3D	198
12.5	基于稀疏模型的去噪算法	201
12.5.1	稀疏表示简介	201
12.5.2	稀疏去噪原理及模型	202
12.5.3	字典构建算法	203
12.5.4	稀疏分解算法	204
12.5.5	稀疏表示去噪效果	205
12.6	总结	206
	参考文献	207

第 13 章 压缩成像技术	209
13.1 概述	209
13.2 压缩感知基本理论	209
13.2.1 稀疏表示	209
13.2.2 测量矩阵	210
13.2.3 重构算法	211
13.3 压缩成像技术	211
13.3.1 单像素照相机	212
13.3.2 基于编码孔径的压缩成像	213
13.3.3 CMOS 压缩成像	214
13.4 总结	216
参考文献	216
第 14 章 模糊图像复原技术	218
14.1 概述	218
14.2 图像模糊的类型	219
14.2.1 大气模糊	219
14.2.2 散焦模糊	219
14.2.3 运动模糊	220
14.3 图像复原的数学模型	222
14.3.1 盲卷积	222
14.3.2 模糊核的估计	224
14.3.3 非盲卷积	228
14.4 计算摄影中的技术应用	231
14.4.1 编码曝光技术	231
14.4.2 编码孔径技术	235
14.5 总结与展望	239
参考文献	240

第 1 章 概 述

又一个百年老店败落了！柯达，这家曾经拥有近 15 万员工、几乎是“摄影”的代名词的商业帝国风雨飘零。这次不是由于金融风暴，也不是由于经济危机，更不是由于天灾，而是被一项技术打败，一项他自己发明的技术！这使人们自然而然地想到美国好莱坞大片中的情景，科学家被自己研制出的机器人或“怪兽”吞噬！这个“怪兽”就是数字摄影技术。

1975 年，柯达发明了世界上第一台数码相机，从而开创了以硅代替胶片的数字摄影时代。如今，很多人的生活已离不开它，最近发生的许多热点问题，都与手机等数字摄影设备拍摄下来的照片或视频密切相关。所以说，数字摄影技术已经深入应用在人类社会活动的许多方面，而在数字摄影技术基础上发展起来的计算摄影学将为这场变革掀起更大的波澜。毫无疑问，变革还只是刚刚开始！

1.1 计算摄影学的起源

计算摄影学是 Computational Photography 的英译。这个词起源于学术界对光场技术的研究。所谓光场，即 Light Field，是指描述光在空间中传播的每一个方向和经过的每一个点的函数。光场的概念最初是由英国著名物理学家、化学家迈克尔·法拉第（见图 1-1）于 1846 年提出的，法拉第认为光应该像电一样，也可以用场来描述。

1936 年，Arun Gershun 在一篇论文中创造了“Light Filed（光场）”一词，用于描述光在三维空间的辐射特性。1996 年，斯坦福大学的 Marc Levoy 和 Pat Hanrahan（见图 1-2）把光场一词引入计算机图形学，他们基于光场提出的应用是基于图像的绘制（Image Based Rendering, IBR），即在不需要场景结构信息的情况下，可以从已知的视点位置观察到的场景，合成出新的视点位置上观察到的场景。

1996 年之后，研究人员从理论与实践两方面开展了光场的研究。在理论方面，研究人员对光场抽样进行了空间域和频率域分析，提出了几种新的光场参数法方法，包括表面光场和非结构化 Lumigraph。在实践方面，研究人员实验了很多种采集光场的方法，包括摄像机阵列、微透镜、万花筒等。斯坦福大学计算机图形实验室在光场研究过程中，主要集中在如何设计或引入各类高性能成像设备为光场研究服务。经过一段时间研究，他们发现光场这个词已经不能够涵盖他们的研究内容，于是 Computational Photography 的概念就被提出。2005 年 5 月，麻省理工学院、斯坦福大学、微软研究院的研究人员在麻省理工学院召开了首届计算摄影学研讨会。由于数字摄影技术在 2005 年后的飞速发展，特别是基于手机的数字摄影技术的井喷式普及，计算摄影学获得了学术界和工业界

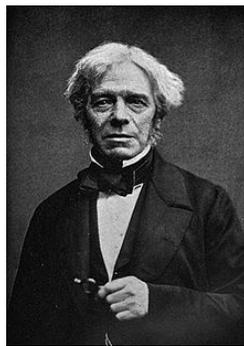


图 1-1 提出光场概念的
迈克尔·法拉第



(a) Marc Levoy 教授



(b) Pat Hanrahan 教授

图 1-2 首次把光场引入计算机图形学的两位教授

的极大重视,得到了快速发展,研究范畴已远超越光场,成为一个热点研究领域。自 2009 年,每年举办一届 IEEE 计算摄影学国际学术会议。在国际上,计算摄影学领域代表性的学术研究机构包括麻省理工学院、斯坦福大学、哥伦比亚大学、微软研究院、北卡罗来纳大学和卡内基梅隆大学等,并开设了相关研究生课程。

1.2 计算摄影学的相关概念

前面讲过,计算摄影学的概念早已突破光场的研究范畴,广义上说,它是在数字摄影学 (Digital Photography) 的基础上,增加对成像过程与设备的精确控制,对影像更加复杂的处理与计算,并使用 3D 成像、全景处理等新颖的数字媒体处理技术,从而在摄像过程中彻底引入计算单元的一套理论、方法与技术。

按照这种观点理解,大多数人对计算摄影学就一点都不陌生了,今天大多数人用的数码相机和手机都含有计算摄影的技术,如光学防抖、人脸识别、宽动态、自动白平衡、全景拼接等。

对计算摄影学有如下不同的定义。

1. Wikipedia 的定义

计算摄影学是指基于计算设备的图像采集、处理、操作技术,以增强或拓展数字摄影的能力。^[1]

2. Ramesh Raskar 的定义

麻省理工学院媒体实验室的 Ramesh Raskar 副教授在 *Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors* 一书中对计算摄影学是这样定义的:“计算摄影学组合了大量计算、数字传感器、现代光学、传动装置和智能照明设备,以突破传统胶片式摄像机的局限,创造出一些新颖的摄影应用。”^[2]

3. Marc Levoy 等人的定义

斯坦福大学 Marc Levoy 教授等在其开设的计算摄影学课程中对计算摄影学是这样定义的：“计算摄影学广义上是指用于增强或扩展数字摄影能力的感知策略和算法技术，这些技术的输出是一张正常的照片，但可能不是采用传统的照相机拍摄的，代表性的技术包括高动态成像、编码光圈与编码曝光成像、结构光摄影、多透视与全景拼接、数字蒙太奇、全聚焦成像和光场成像等。”^[3]

4. Alexei (Alyosha) Efros 的定义

卡内基梅隆大学的 Alexei (Alyosha) Efros 教授在其开设的计算摄影学课程中对计算摄影学是这样定义的：“计算摄影学是融合计算机图形、计算机视觉和摄影学而创造出的一种新的研究领域。它的作用是通过使用计算技术，突破传统照相机的诸多限制，产生出对我们现实视觉世界一种更丰富、更生动、认知上也许更有意义的表达。”^[4]

5. Peter N. Belhumeur 的定义

美国哥伦比亚大学的 Peter N. Belhumeur 教授在其开设的计算摄影学课程中对计算摄影学是这样定义的：“最近几年，计算机图形、计算机视觉和摄影学的领域交叉融合产生了一种新的、很活跃的研究领域——计算摄影学。计算摄影学的目标是使用计算技术重新定义照相机，以产生一种新形式的图像和视觉表达。”^[5]

6. James Hays 的定义

美国布朗大学的 James Hays 在其开设的计算摄影学课程中是这样定义计算摄影学的：“计算摄影学描述了计算机图形、计算机视觉、因特网与摄影学的交叉融合。它的目的是使用计算技术来克服传统的摄影学局限性，以增强我们采集、操纵和交互视觉媒体的方式。”^[6]

从上述定义不难总结出，计算摄影学是对数字摄影能力的增强和扩展，集成了现代光学、计算机图形学、计算机视觉、因特网与摄影学等多学科知识，把计算技术引入到传统摄影学中，以创造出各种新颖应用的新的学科。

狭义上讲，计算摄影技术就是计算摄影学中涉及的大量关键技术。广义上讲，计算摄影、计算摄影技术、计算摄影学三个词的含义是一样的，是同一个英文词组 Computational Photography 在不同语境中的不同表达。

那么，计算摄影技术与数字摄影技术这两个概念如何区分呢？数字摄像技术是以 CCD/CMOS 数字感光器件代替胶卷为起点而对传统照相机进行改造所涉及的一系列技术，包括自动曝光技术、自动白平衡技术、图像压缩技术、3D 降噪技术等。光学镜头与数字感光器件是数码相机的两大主要部件，而光学镜头与胶卷是胶片相机的两大主要部件。数字摄像技术与传统摄像技术相比，整体结构没有发生很大变化，只是由胶卷换成了数字感光器件。而计算摄影技术则是在数字摄影技术的基础上，对摄影技术做出的很大变革，光学镜头、曝光方式、聚焦方式乃至相机所成的图像都完全不同。

1.3 计算摄影学涉及的研究领域

计算摄影学的研究领域包括以下几个方面。

1. 摄像机结构与图像形式化

对摄像机结构与图像形式化 (Camera and Image Formation) 理论的研究属于计算摄影学基础理论研究范畴。相关理论的突破会导致计算摄影技术上的飞跃。

2. 人类视觉认知

人类视觉认知 (Human Visual Perception) 的研究涉及生理学、心理学以及信息科学, 也是计算摄影学的基础研究课题。

3. 图像金字塔、梯度场与双边滤波

图像金字塔、梯度场与双边滤波 (Image Pyramid, Gradient and Bilateral Filter) 是与计算摄影学相关的图像处理技术, 是视频压缩、降噪、运动估计等相关算法的基础。

4. 颜色、白平衡、Bayer 模式与去马赛克

颜色 (Color) 模型是图像与视频的基础, 也是白平衡 (White Balance) 处理的基础, Bayer 模式是 CMOS 感光器件输出的原始信号, 用于显示、压缩与存储的视频或图像一般是 RGB 或 YUV 格式, 这就需要利用去马赛克处理 (Bayer Pattern and Demosaicing) 把 Bayer 模式的信号转换为 RGB 或 YUV 信号, 并尽可能保持画面的层次感和亮丽程度。

5. 计算光照

计算光照 (Computational Illumination) 是指以结构化的模式控制光照, 以得到特殊效果的图像, 通过处理图像, 实现图像增强、图像去模糊以及被摄场景结构恢复等操作。

6. 编码光圈与编码曝光

编码光圈 (Coded Aperture) 是将特定模式的掩膜放置在照相机的光圈前, 照相机可拍摄到一幅全模糊的图像, 通过特定的算法可以把图像恢复成全清晰的图像。编码曝光 (Coded Exposure) 是指按特定的模式控制摄像机的曝光时间长短, 摄像机曝光时间长, 在低照度环境中依然能够拍摄到明亮的画面, 但容易产生运动模糊, 曝光时间短, 不会产生运动模糊, 但画面较暗, 通过摄像机曝光时间长短的不同搭配, 有可能得到没有运动模糊的明亮画面。

7. 自动聚焦

自动聚焦 (Auto Focus) 是照相机和摄像机中十分重要的一项功能。如果自动聚焦不好, 拍摄出的画面就会产生模糊, 再高分辨率的画面或再亮丽的画面也无法看清楚。虽然自动聚焦算法已经被研究很多年, 在照相机和摄像机中也应用了很多年, 但相关的

技术还需要研究，特别是在低照度环境或弱纹理环境下，如何准确自动聚焦依然是需要解决的问题。

8. 去运动模糊与去散焦模糊

当曝光时间较长时，拍摄运动会产生运动模糊。当照相机聚焦不好时，画面会产生散焦模糊。有运动模糊或散焦模糊的图像或视频需要采取特殊的算法——去运动模糊与去散焦模糊算法（Motion Deblurring and Defocus Deblurring）才能把它们变得清晰。

9. 高动态成像

动态范围表征图像中最亮的亮度值与最暗的亮度值之间的变化范围。比如，人们对窗户拍摄，动态范围高的照相机拍摄出的画面既可以看到室外的场景，也能看清室内的场景。如何得到高动态成像（High Dynamic Range Imaging），或叫宽动态成像（Wide Dynamic Range Imaging），是计算摄影学领域当前研究的热点之一。

色调映射技术（Tone Mapping）是高动态成像经常采用的技术（见图 1-3），在左边图中，中间区域太亮了，而上边区域太暗了，如果采用色调映射技术，可以对光线进行处理，使得既能看到亮的区域，又能看到暗的区域，得到右边图所示的效果。



图 1-3 色调映射技术的应用^[7]

10. 图像修复

当照片年久失修有残缺时，或数字图像在网络传输中有误码而导致部分像素错误时，图像修复（Image Complete）技术可以把残缺或错误的像素恢复出来。传统图像修复的基本思路是从残缺照片中截取与残缺部分最相近的像素，贴补到残缺部位。如今，这项技术已扩展到通过互联网收集大量的照片集，采用统计的方法从照片集中找到与残缺部分最相近的像素贴补到残缺部位。

11. 图像合成

图像合成（Image Compositing）或叫数字蒙太奇技术（数字 Photomontage），比如家庭聚会的照相（见图 1-4），图中上面四张照片中每张都有一些人的面部表情不太合适，在没有计算摄影之前，要反复去拍很多张，挑出合适的一张。在有了计算摄影之后，就可以把四张照片中面部表情比较好的区域运用蒙太奇处理，做出一个合成图像，在这个合成图像中，每个人的表情都达到满意的效果。

12. 全聚焦成像

人们知道，普通照相机只能聚焦在场景中的某个距离，超出这个距离一定范围的物体在照相机中不能清晰成像。而全聚焦成像（All-focus Imaging）可以对场景中任何距

离的物体清晰成像。这是一项全新的技术，彻底改变现有照相机的聚焦方式，摄影师不用在拍摄现场聚焦，可以在拍摄完成后任意调整聚焦点。



图 1-4 蒙太奇处理^[8]

13. 多透视全景拼接

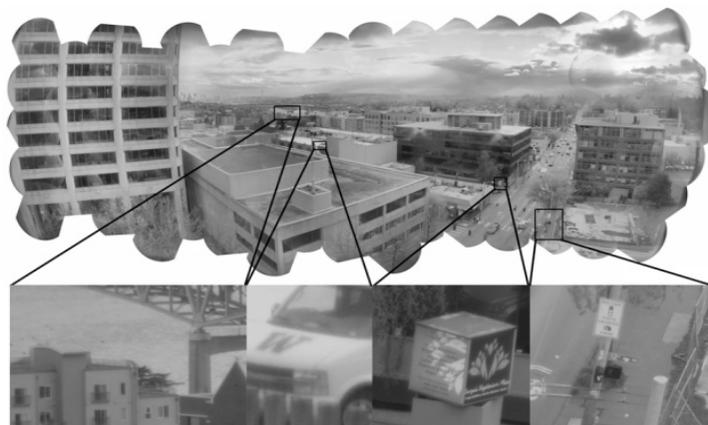
一般照相机或摄像机的成像视角在 60°左右，如果需要拍摄到 180°或 360°的高分辨率全景画面，往往会把多幅图像拼接成全景画面，由于每幅图像都是透视图，这种拼接又被称为多透视全景拼接（Multi-Perspective Panoramic Stitching）。

比如要照一个宽景时，可以分别沿着画面左右摇摆拍很多张照片，拼接出一个分辨率高、画面完整、连续的整体性的图像，如图 1-5 所示。



图 1-5 全景图拼接

图像拼接技术可用于构建宽视角超高分辨率的成像系统。在美国国防部先进研究项目局 DARPA 的支持下，杜克大学的研究人员采用图像拼接技术研制出总分辨率达到数十亿像素的实时成像系统 AWARE2，该系统由 98 个 1400 万像素的 CMOS 拼接而成，图 1-6 显示的是该系统拍摄的一幅画面，上图显示的是全貌，下图显示的是某些局部区域的细节场景。该项目被美国媒体评为“美国国防部 20 项有望改变人们生活的军事项目”之一。^[9]

图 1-6 AWARE2 拍摄出的一幅画面^[9]

14. 视频总结

随着计算摄影技术的发展，摄像机或带摄像功能的手机成为人们生活的必需品。于是，视频内容越来越庞大，为了对视频内容进行更好的管理，视频总结（Video Summarization）功能就十分必要。所谓视频总结，是指通过分析视频内容，从视频中提取出典型的、有代表性的画面或文字。

15. 光场成像

光场成像（Light Field Imaging）是指能获取光场数据的成像过程。世界上第一台消费级光场照相机是 Lytro 公司于 2011 年 10 月 20 日推出的，如图 1-7 所示。通过光场技术的应用，无论抓拍的照片模糊与否，只要在照相机的焦距范围内，对焦点可以在拍完之后随意选择，因为照相机在拍照时就把焦距范围内所有光学信息都记录在内了。另外，这种照相机拍出来的照片是直接支持 3D 的。Lytro 选用 1100 万像素的 CMOS，采用微透镜技术构建光场。由于一块 CMOS 需要记录光场中的所有光线信息，因而处理后的输出图像像素不会超过 1200×900 像素，其分辨率过低。

德国 Raytrix 公司面向工业与科学研究推出的光场照相机 R29 如图 1-8 所示。R29 选用 2900 万像素的 CCD，也是基于微透镜技术构建光场，可得到超过 700 万像素的输出图像。但价格比较昂贵。

图 1-7 Lytro 公司的光场照相机^[10]图 1-8 Raytrix 公司的光场照相机^[11]

16. 稀疏编码

如图 1-9 所示，人们可以看到左边的图像有非常多的噪声，通过稀疏编码（Sparse Coding）技术降噪以后可以恢复出来没有噪声的图片，即右边的图像。从这么大噪声的图片中恢复出完美的右边的图像，采用普通的去噪技术达不到这样的效果，必须采用最先进的稀疏编码技术。当然，这项技术发展成熟进入到人们生活中还有待时日。



图 1-9 稀疏编码技术^[12]

1.4 计算摄影学的未来

对于计算摄影学的未来，大概会有以下几种观点。

一种观点认为这项技术已经非常成熟了，不需要太大的发展了。人们随身携带的手机已经能够拍摄高清的照片和视频，而且分辨率越来越高。比如 Nokia 推出的 PureView 808 可拍摄 1080p 全高清视频、4100 万像素静态图像。再发展，也只是在分辨率数值上的更进一步提高，比如从 4100 万像素发展到 5000 万像素，再到 6000 万像素等，不会有实质上的进步了。

另一种观点认为计算摄影学再怎么发展，起主导作用的还是镜头的设计。就如同现在的数码相机，虽然机身计算性能在不同提升，图像分辨率在不同提高，但摄影爱好者花钱最多的却依然是镜头，这与胶片摄影时代没有太大的改变。也就是说，持这种观点的人认为，计算摄影学发展与否无关紧要，因为它在整个摄影领域不占主导地位。

还有一种观点认为在未来计算摄影学的发展中起主导作用的不是成像技术的发展，而是拍摄得到的图像或视频的各类应用。由于带高清摄像头的智能手机越来越普及，硬件方面的条件已经完全具备，而依托移动互联网技术发展而兴起的视频内容服务与各类应用程序将会逐步渗透到每一个人的日常生活中，并在计算摄影学未来的发展中占据主导地位。

对于上述观点，本书作者没有持否定立场。只是需要强调的是，计算摄影学不是发展到了很高的阶段，而是刚好相反，计算摄影学当前达到的技术水平还远远没有满足用户的需求，计算摄影学的发展还只是刚刚开始。

1975 年，柯达工程师 Steven J. Sasson 研制了一台名叫“手持电子照相机”的原型机（见图 1-10）。把原型机技术报告原文摘录如下：“此报告中所描述的照相机是指首次

试验成功的照相机，此次试验旨在证明，随着技术的进步，摄影系统必将对未来的拍照方式造成实质性的影响。未来的照相机可以想象成是一种能在光照条件极差的情况下拍摄出彩色照片的小型设备。那时的照片将存储在一种磁介质内，一种非易失性、稳定性极佳的存储器，可从照相机内取下以进行播放。这种照片的分辨率将至少相当于现在的110胶卷。声音也可同影像一并录下，以增加照片的诠释性。电子形式的照片经稍作修改或不作修改便可通过现有的通信信道发送出去。红外摄影和可视摄影可通过同一部摄像机实现，只需更换光学滤镜即可。照片将保存在胶卷、磁带或视频光盘上，并且相机存储介质将可重复使用。”^[13]



图 1-10 第一台数码相机及其发明人

从这段文字不难看出，Steven J.Sasson 当时展望的“未来照相机”的主要功能在今天的数码相机中都能实现。但是也应看到，有些功能还是没有实现，比如今天的数码相机还是不能“在光照条件极差的情况下拍摄出彩色照片”，还是不能达到“红外摄影和可视摄影可通过同一部摄像机实现，只需更换光学滤镜即可。”

也许有人会说，要实现这两点也不难，数字感光器件技术发展了就可以了。是的，感光器件的性能提升很重要，但计算摄影系统的结构设计 with 处理算法的发展更加重要，以 Pelican Imaging 公司的阵列式摄像机 PiCam 为例^[14]。

如图 1-11 所示，Pelican Imaging 公司在 2014 年推出的这款阵列式摄像机的感光器件是由 16 块（4 行×4 列）小尺寸、低分辨率的 CMOS 裸片封装而成[见图 1-11 (c)]，每个 CMOS 的分辨率是 1000×750 像素。在感光器件上覆盖了 1.5mm 左右厚度的镜头模块，镜头模块也分为 16 个独立部分，每个部分对应一个 CMOS 单元。也就是说，在 PiCam 中，每个 CMOS 是单独成像的，所成的像如图 1-12 所示。整个摄像机模块的厚度只有 2.75mm。



(a) 摄像机模块

(b) 镜头模块

(c) 感光器件模块

图 1-11 PiCam^[14]

从图 1-12 不难看出，PiCam 得到的原始图是 16 幅灰度图，后端的算法可以把 16 幅 1000×750 像素的低分辨率灰度图像合成出 800 万像素（ 3264×2448 像素）的高分辨率彩色图像（见图 1-13），并且可以得到所拍场景的深度图（见图 1-14）。通过软件处理，摄影者还可在所拍照片的屏幕上随意对焦。例如，在聚会上的照片，可根据不同参加者定制出不同的照片，让每个人都有机会成为主角。使用这种相机，摄影者不必花时间先对焦再按快门，也不必担心匆忙之间对错焦，照片一片模糊。低光线，无闪光灯，影像快速移动时，仍能正常拍摄，而且可以显示 3D 照片。



图 1-12 PiCam 所成的原始图像^[14]



图 1-13 PiCam 合成后图像^[14]

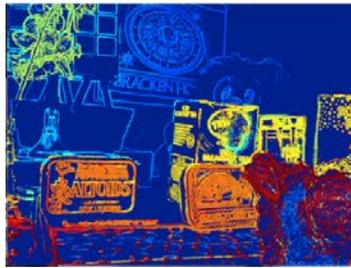


图 1-14 PiCam 计算得到的深度图（不同颜色表示不同的深度）^[14]

Pelican Imaging 公司的阵列式摄像机技术给出三点启示。

(1) 未来计算摄影技术的发展不完全依赖于感光器件的发展。多个低性能的 CMOS 组成阵列，辅以高水平处理算法，可以得到很高水平的成像效果。

(2) 未来的照相机/摄像机输出的将不仅仅是二维的图像，可以输出光场数据。有了光场数据，人们可以直接获取被摄物体的三维信息，这些信息送给 3D 打印机即可直接输出实物。也就是说，未来的照相机加上 3D 打印机后，拍摄一次便可以得到被摄场景的类似雕塑的实物，而不仅仅是一纸照片。

(3) 随着纳米技术、超高精度加工技术的发展，也许未来的摄像机将不会有光学镜头，科学家直接在芯片上附加一层透明薄膜就可以成像，也就是说，一块芯片就是一台摄像机。这台摄像机体积小，而且不需要瞄准镜、不需要聚焦、不需要测光，更不需要再配置专门的光学镜头，因为摄像机已经能够记录全光场数据，哪怕光线很弱，只要有光线，它就能记录。这将会彻底改变人们对照相机/摄像机的基本认知，即

照相机 = 光学镜头 + 机身

需要说明的是，做这样的预测时，不是降低光学设计人员在未来计算摄影学发展中的地位，恰恰相反，这种芯片式摄像机的设计首先就离不开光学设计人员。

以上三点启示就是对计算摄影学未来发展的看法。毫无疑问，计算摄影学的每一项

重大创新，都将会对摄影技术产生颠覆性影响，同时也将直接改变人们的日常生活。甚至可以说，计算摄影学将会是 200 年摄影史上的又一次革命。

参 考 文 献

- [1] Wikipedia. Computational photography[EB/OL]. 2013 [2013-12-31]. http://en.wikipedia.org/wiki/Computational_photography.
- [2] Raskar R, Tumblin J. Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors[M]. AK Peters, Ltd., 2009.
- [3] CS 448A. Computational photography[S/OL]. 2010 [2013-12-31]. <http://graphics.stanford.edu/courses/cs448a-10/>.
- [4] Carnegie Mellon University. Computational Photography[EB/OL]. 2007[2013-12-31]. http://graphics.cs.cmu.edu/courses/15-463/2007_fall/.
- [5] Peter N. Belhumeur. Computational photography[EB/OL]. 2010[2013-12-31]. <http://www1.cs.columbia.edu/~belhumeur/courses/ompPhoto/>.
- [6] CSCI 129. Computational photography[EB/OL]. 2012[2013-12-31]. <http://cs.brown.edu/courses/csci1290/>.
- [7] Marc Pollefeys, Kevin Koester, Luca Ballan. Computational Photography & Video[EB/OL]. 2010[2013-12-31]. <http://www.inf.ethz.ch/personal/pomarc/courses/CompPhoto/>.
- [8] Agarwala, Aseem, et al. Interactive digital photomontage[J]. ACM Transactions on Graphics 2004,23(3).
- [9] Brady D J, Gehm M E, Stack R A, et al. Multiscale gigapixel photography[J]. Nature, 2012, 486(7403): 386-389.
- [10] Lytro: Lytro User Manual[EB/OL]. 2013[2013-12-31]. <https://www.lytro.com>.
- [11] Raytrix: Ratrix 3D light field camera technology[EB/OL]. 2013[2013-12-31]. <http://www.raytrix.de>.
- [12] <http://www.cvchina.info/2010/06/01/sparse-representation-vector-matrix-tensor-1/>.
- [13] ZOL. 数码相机回顾[EB/OL]. 2006 [2013-12-31]. http://dcbbs.zol.com.cn/4/139_34723.html.
- [14] Venkataraman K, Lelescu D, Duparré J, et al. PiCam: an ultra-thin high performance monolithic camera array[J]. ACM Transactions on Graphics, 2013, 32(6): 166.

第 2 章 摄影学基础知识

摄影技术的发明是人类文明史上继文字发明之后颇为重大的事件之一，它彻底改变了人们认知客观世界的方式。在摄影技术发明之前，人们是通过口口相传或文字，再辅以绘图来认知时间和空间上距离较远的未知世界。俗话说，“百闻不如一见”，这种认知方式无疑存在很大的局限。自摄影技术发明之后，人们就可以较为轻易地捕捉美好的瞬间，给远地的家人朋友欣赏，或留待以后慢慢追忆。

随着计算机、互联网与数码相机的发展，摄影技术的功效得到极大的扩展。沿用了 100 多年的胶卷加相纸已被数字化，转变为存储在各类存储卡、硬盘或网络上的图片文件。这些文件可以进行压缩、编辑、合成、网络传输等处理。拍摄器材也由专业性设备发展为人们的生活用品（比如手机）的必备功能部件，拍摄出来的图片以光速在人们之间传播，从而彻底改变了人们的生活与工作方式。至此，人类彻底步入计算摄影^[1~6]时代。

2.1 传统摄影学基础

什么是摄影？按《百度百科》的解释^[7]，“摄影”的英文单词是 Photography，这个词是源于希腊语 φωσφος（光线）和 γραφι graphis（绘画、绘图）或 γραφη graphê，两字组合在一起的意思是“以光线绘图”。摄影是指使用某种专门设备记录物体反射光线的过程，所以说，没有光就没有摄影，摄影本质上就是一门用光的艺术。

2.1.1 摄影是用光的艺术

摄影是光线进入胶片或者数字成像器件所产生出来的效果（见图 2-1），所以说，摄影是用光的艺术。另外以开玩笑的方式说，所谓“用光的艺术”（见图 2-2）就是摄影爱好者需要采购不同种类的镜头，这些镜头都非常昂贵，摄影爱好者的钱就“用光”了。



图 2-1 摄影是用光的艺术

既然光在摄影领域如此重要，下面就先来认识光。

光是由光子组成的，具有粒子性与波动性，称为波粒二象性。在这里，主要考察光的波动性。从科学意义上说，光是指所有的电磁波谱。如图 2-3 所示，电磁波的波长短的小于 10^{-16}m ，长的超过 100km，包括无线电波、红外光、可见光、紫外光、X 光、 γ 射线等。



图 2-2 昂贵的镜头——“用光的艺术”

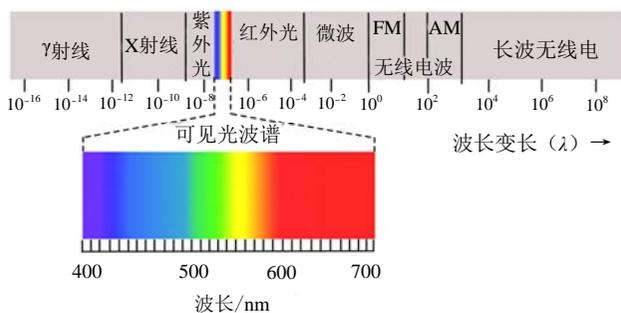


图 2-3 光谱图

摄影所感知的光主要是可见光，特殊情况下包括部分红外光或紫外光。对于可见光的光谱范围也没有一个明确的界定，一般人的眼睛所能接受的光的波长为 380~760nm，见表 2-1。

表 2-1 可见光的波长

(单位: nm)

紫	蓝	绿蓝	蓝绿	绿	黄绿	黄	橙	红
380~450	450~480	480~490	490~500	500~560	560~580	560~610	610~650	650~760

红外光在军事侦察等领域应用广泛。它的波长比可见光要长，红外光是由弗里德里希·威廉·赫歇尔 (Friedrich Wilhelm Herschel, 见图 2-4) 于 1800 年发现的，他将太阳光用三棱镜分解，并在不同颜色的色带位置上放置温度计，以此测量各种颜色光的加热效应。结果发现，位于红光外侧的那支温度计升温最快。由此得出结论：太阳光谱中，红光的外侧必定存在看不见的光线，这就是红外线。值得一提的是，弗里德里希·威廉·赫歇尔不仅仅是英国著名的天文学家，他发现了天王星及其两颗卫星、土星的两颗卫星、太阳光中的红外辐射等，而且还是有名的古典作曲家、音乐家。



图 2-4 弗里德里希·威廉·赫歇尔

所有高于热力学温度零度 (-273.15°C) 的物质都可以产生红外光。一般认为，红外光的波长为 $0.76\sim 1000\mu\text{m}$ ，科学家将不同波长范围的红外光分为近红外光、中红外光和远红外光。远红外光是红外光中波段最宽的，一般定义远红外光波长范围为 $4\sim 1000\mu\text{m}$ (日本远红外协会定义为 $3\sim 1000\mu\text{m}$)，远红外光占据太阳辐射能量的 72%。波长为 $0.76\sim 1.5\mu\text{m}$ 的是近红外线，波长为 $1.5\sim 4\mu\text{m}$ 的是中红外线，如图 2-5 所示。

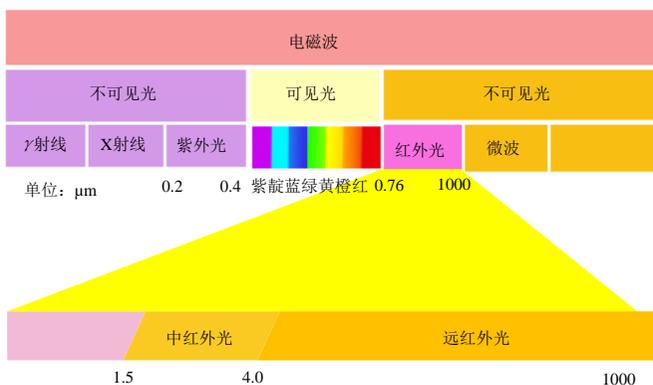


图 2-5 红外光谱图

2.1.2 镜头与焦距

1. 透镜组

镜头是照相机或摄像机两大核心部件之一，它由多块透镜组成，其作用是尽可能地吸收一定视场范围内的物体反射的光线并改变光线方向，使它们会集在感光器件表面形成实像。

人们知道，按透镜成像原理：物体放在焦点之外，在凸透镜另一侧成倒立的实像，如图 2-6 所示， F 表示焦点的位置，实像有缩小、等大、放大三种。物距越小，像距越大，实像越大。

既然单个透镜也能成像，而镜头一般都是由多块透镜组成，这是为什么呢？

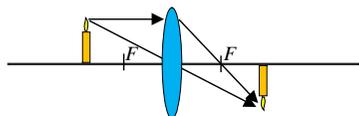


图 2-6 凸透镜成像原理

这是因为根据单透镜理论确定的理想像与实际存在偏差，叫像差。像差一般包括色像差和单色像差。

色像差简称色差。不同波长的光在通过透镜时的折射率不同，物体上的一个点，在像平面上则可能形成一个色斑。色差一般有位置色差、放大率色差。位置色差使像在任何位置观察，都带有色斑或晕环，使像模糊不清；而放大率色差使像带有彩色边缘^[8]。采用多块透镜组合的光学系统，最主要的功效就是消除色差。

单色像差是与色无关的像差，单色像差又分为五种，分别是球面像差、彗形像差、像散、像场弯曲和畸变。

球面像差是为了降低成本，物镜的制作是球面磨制，而能使光线汇聚一点的凸透镜，不是球面的，由此造成的像差。

彗形像差是指由于物镜中心与边缘的像差程度不同，造成带尾巴的分散圈。

像散是指光线每经过一片镜片，都要产生损耗，大多数直线通过，其余的部分被打散了。

像场弯曲是指在一个平坦的像平面上，像的清晰度从中央向外发生变化，聚焦形成弧形。如图 2-7 所示，存在像场弯曲的光学系统中，焦点不在像平面上，而是在像平面前成弧形。

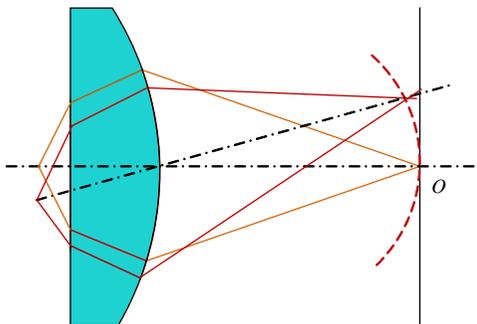
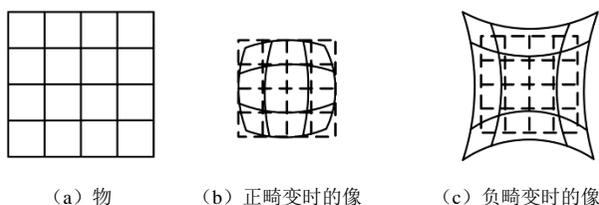


图 2-7 像场弯曲

畸变是指画面中心与边缘的折射角度不同，造成拍摄物体变形。如图 2-8 所示，图 2-8(a) 显示的是方格形的物体，图 2-8(b) 和图 2-8(c) 所示则是不同畸变导致变形的像。



(a) 物

(b) 正畸变时的像

(c) 负畸变时的像

图 2-8 畸变后的像

除了多块镜片组合消除像差之外，在光学设计中还经常采取的手段是镀膜。镀膜是指在镜片表面镀上非常薄的透明薄膜。镀膜的目的是减少光的反射，增加透光率。

2. 焦距

在变焦镜头中，不同的镜头元件可以前后移动。通过改变特定透镜之间的距离，就能调节整个镜头的放大率——焦距。焦距是镜头的关键参数之一。当平行于主光轴的光线穿过透镜时，会汇聚到一点上，这个点叫做焦点。

焦点到透镜中心（即光心）的距离，被称为焦距。焦距的单位一般用 mm 来表示。焦距越长，相机所看到的场景越窄，视场角越小，看得越远；焦距越短，相机看到的场景越宽，视场角越大，看得没有焦距长的那么远，如图 2-9 所示。



图 2-9 镜头焦距的意义

镜头的视场角不仅仅与焦距相关，还与成像靶面相关。人们经常说的 135 胶片相机所用的 135 胶卷是指胶卷的高度是 35mm，由于上下两端有齿孔，所以有效高度为 24mm。由于成像的画幅是 3:2 的比例，所以水平方向是 36mm，高是 24mm，有效的成像面积是 36mm×24mm，如图 2-10 所示。



图 2-10 135 胶卷及 135 胶片相机

50mm 焦距的镜头用在 135 胶片相机上，其视角大约是 46°，而用在 APS-C 画幅（成像靶面是 22.5mm×15mm）的相机上，其视角就是大约 30°。而这 50mm 镜头在 APS-C 的机器上的拍摄视角大概与 75mm 焦距的镜头在 135 胶片相机的底片上成像的视角相当，都是 30°左右。

3. 镜头接口

这里所说的镜头接口是指可更换镜头与摄像机/照相机机身的连接方式。部分摄像机/照相机的镜头与机身在出厂时就固定在一起，镜头是不可更换的。但更多的摄像机/照相机镜头被设计成可更换模式，从而方便用户根据拍摄需要选配不同的镜头。

早期摄像机/照相机的机身与镜头的连接是通过机身与镜头上的螺纹连接在一起的，这种方式虽然比较麻烦，但安装后比较牢固。目前安防类的摄像机仍然沿用这一连接方式。这种连接方式有 CS、C、M12、M9 等不同的标准。

CS 接口与 C 接口螺纹部分相同，但两者从镜头到焦平面的距离不同。C 接口从镜头安装基准面到焦平面的距离是 17.526mm。CS 接口从镜头安装基准面到焦点的距离是 12.5mm。所以，如果把 C 接口的镜头安装在带有 CS 安装座的摄像机上，则需要使用镜头转接器，垫高摄像机的镜头安装座。如果把 CS 接口的镜头安装在带有 C 安装座的摄像机上，则需要将摄像机机前部的垫圈取下再安装镜头。

M12、M9 的镜头都是采用长螺纹的方式连接，对镜头到焦平面之间的距离没有做特定的限制。M12 接口表示镜头连接部分的直径是 12mm。M9 接口的镜头连接部分直径是 9mm。

照相机的镜头与机身之间的连接方式已经抛弃了螺纹模式，基本上都是采用卡口。采用卡口后，只需要镜头和机身上的卡槽对准，旋转很小的一个角度，镜头就安装到机身上，并且光圈的连杆也很方便地与机身配合到位。卡口因为有技术专利，各个照相机厂家各自为政，没有国际统一的标准卡口。

卡口的参数除包含镜头安装基准面到焦平面的距离、卡口环直径外，还包含镜头与机身间的控制信号参数。1959 年美能达推出第一款 135 单反相机 SR-2，采用了卡口的方式固定镜头，MD 卡口就被固定下来，如图 2-11 所示是采用 MD 卡口的 SR-2。

从 1959 年尼康第一台 135 单反相机采用 F 卡口以来，F 卡口以其稳固的姿态一直

被尼康相机采用。F卡口的内径是44mm，近似于35mm胶卷对角线的长度(43.27mm)，其意义是可以将从镜头射入的光线完全投射到35mm胶片上。

佳能镜头的卡口经历了四次比较大的变迁，从最初的R卡口到FL，再到FD，再到EF。如图2-12所示，左边显示的是佳能FD卡口，右边是EF卡口。



图 2-11 SR-2 照相机



图 2-12 佳能 FD 卡口和 EF 卡口

2.1.3 快门与曝光

曝光也是摄像机/照相机的一个重要特性。曝光使得合适的光量进到感光器件。决定曝光的因素有两个，一个是快门速度(Shutter Speed)，另一个是光圈的大小(Aperture)。

曝光=时间(允许进入光线的长短)×光圈(允许进入光线的口径大小)

1. 快门速度

快门速度决定了感光器件的进光时间，基本上和曝光是线性关系。一般快门速度用 $1/ns$ 表示，即在1s的时间内曝光 n 次。显然， n 越大，快门速度越高。典型的快门速度： $1/30s$ 、 $1/60s$ 、 $1/125s$ 、 $1/250s$ 、 $1/500s$ 。在手持相机的情况下，人手可以在 $1/60s$ 的快门时间内不抖动。

快门速度过低对高速运动的物体会产生拖影。如图2-13所示，快门速度低时[见图2-13(a)]，拿开盒子会有拖影；快门速度高时，即使盒子在高速运动也能看清盒子[见图2-13(b)]。



图 2-13 高速运动物体产生拖影

快门速度直接决定成像的清晰度。为了防止高速运动物体拍摄出来变得模糊，应控制好快门速度，即安全的快门速度。一般地，对于走路的行人，安全的快门速度是 $1/125s$ ；跑步者为 $1/250s$ ；行进的汽车为 $1/500s$ ；行进火车为 $1/1000s$ 。

2. 光圈

光圈位于镜头的透镜组中间，通过控制光圈孔径的开合来控制镜头通光量的大小。最初的光圈只是一系列大小不同的圆孔排列在一个有中心轴的圆盘的周围，转动圆盘可将适当大小的圆孔移到光轴上，达到控制孔径的效果。猫眼式光圈由两片有半椭圆形孔的金属薄片对排组成，相对移动便可形成大小不同的光圈。现在一般采用虹膜式光圈，这种光圈由多个相互重叠的弧形薄金属叶片组成，叶片的离合能够改变中心圆形孔径的大小，如图 2-14 所示。弧形薄金属叶片可多达 18 片。弧形薄金属叶片越多，孔径越接近圆形。只用 7 片薄金属叶也可以得到接近圆形的孔径。

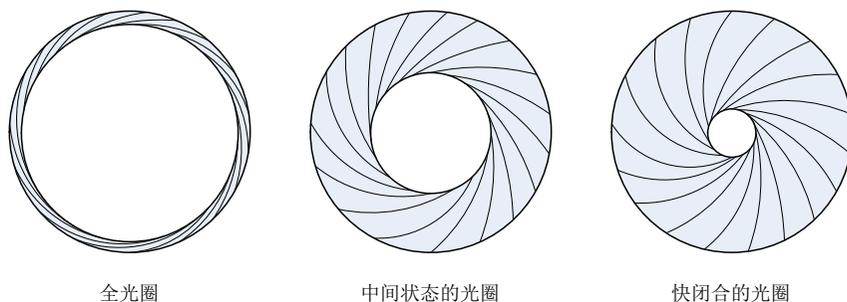


图 2-14 虹膜式光圈

光圈的大小一般用 f 值表示。 $f2.0$ 光圈是指当镜头焦距为 50mm 时其光圈直径大小是 25mm。即 f 值=镜头焦距/镜头口径的直径。 f 值是一个比值而不是一个有具体单位的数字。 f 值越小，光圈越大。典型的光圈大小： $f/1.4$ 、 $f/2.0$ 、 $f/2.8$ 、 $f/4$ 、 $f/5.6$ 、 $f/8$ 、 $f/11$ 、 $f/16$ 、 $f/22$ 、 $f/32$ ，如图 2-15 所示。

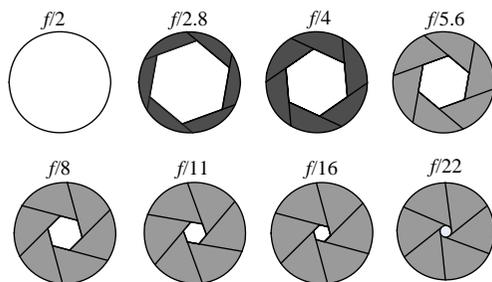


图 2-15 光圈大小与 f 值的关系

值得一提的是，由于光圈值与镜头口径的直径成比例关系，而通光量应该与镜头口径的面积成正比，所以，通常 $f/2.0$ 的通光量是 $f/1.4$ 的 2 倍 ($2.0^2 \div 1.4^2 = 2$)。

3. 景深

人们在选择光圈大小时，还有一项因素必须要考虑，那就是景深。在聚焦完成后，在焦点前后的范围内都能形成清晰的像，这一前一后的距离范围，便叫做景深^[9]。在进一步讲述景深概念前，先来看看什么叫弥散圆？

理想情况下，与光轴平行的光线射入凸透镜时，所有光线将会聚在焦点，再以锥状的形态扩散开来，如图 2-16 所示，在焦点前后，光线开始聚集和扩散，点的影像变成模糊的，形成一个扩大的圆，这个圆就叫做弥散圆。当弥散圆的直径小到人眼不能分辨时，这时的弥散圆被称为允许弥散圆。不同的应用环境下，允许弥散圆直径的定义不同。以 135 单反相机为例，如果考虑洗出的相片尺寸为 5 英寸×7 英寸（1 英寸=2.54cm），观察距离不小于 25cm，允许弥散圆的直径大约是底片对角线长度（43.2mm）的 1/1500~1/1000，通常按经验值取 0.035mm。

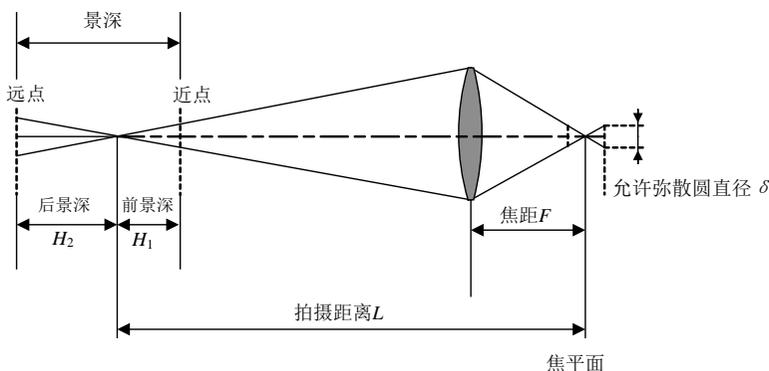


图 2-16 弥散圆与景深

在焦点前后各有一个允许弥散圆，这两个弥散圆之间的距离是景深，对应拍摄景物那一边就是景深。景深分为前景深和背景深，如图 2-16 所示。前景深 H_1 与背景深 H_2 的计算公式分别为

$$H_1 = \frac{f \cdot \delta \cdot L^2}{F^2 + f \cdot \delta \cdot L} \quad (2.1)$$

$$H_2 = \frac{f \cdot \delta \cdot L^2}{F^2 - f \cdot \delta \cdot L} \quad (2.2)$$

式中， F 为镜头焦距； f 为镜头光圈值； L 为拍摄距离； δ 为允许弥散圆直径。

因此，景深 $H = H_1 + H_2$ 。

从上式不难看出，背景深大于前景深。影响景深的因素主要有如下三个方面。

- (1) 光圈大小：光圈越大景深越小，光圈越小景深越大，如图 2-17 所示。
- (2) 镜头焦距：镜头焦距越长景深越小，镜头焦距越短景深越大。
- (3) 拍摄距离：镜头离被摄景物越近景深越小，镜头离被摄景物越远景深越大。



图 2-17 光圈与景深（左图比右图有更浅的景深）

4. 曝光

前面讲过，曝光主要是由光圈和快门两个因素决定的。合适的选择光圈和快门的配对组合，就可以防止运动模糊，防止手抖动模糊。

当需要控制景深时，必须设置为光圈优先模式。当需要景深比较大时，把光圈调小；需要景深比较小时，把光圈调大。在光圈优先模式下，照相机自动配合快门的速度，以对画面进行清晰成像。

当需要拍摄运动的景物，比如运动员跑步动作或鸽子起飞动作时，必须设置为快门速度优先模式。类似的，照相机自动为用户选择的快门速度匹配合适大小的光圈。比如拍摄一张鸽子从地上起飞的场景，根据鸽子的起飞速度选择合适的快门速度，并选择快门优先模式，以拍摄到不会模糊的鸽子起飞的图片。

如果采用傻瓜模式，曝光时间和快门速度都是适中的；手动模式下，拍摄者决定一切设置。有时候为了取得艺术效果，故意拉长曝光时间。图 2-18 显示了故意延长曝光时间后，可以拍摄出的特殊艺术效果。

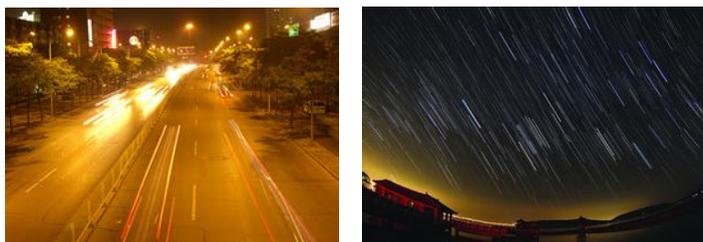


图 2-18 延长曝光时间

2.2 数字摄影学基础

前面讲过，1975 年，柯达公司的工程师 Steven J.Sasson 以盒式磁带作为存储介质，开发出了世界上一款能记录 1 万像素黑白影像的“手持式电子照相机”。这款原型机重约 4kg，记录一张影像的时间约 23s，当时所用的感光器件是 Fairchild 201100 型 CCD 阵列。这台原型机花费了 Steven J. Sasson 一年左右的时间，就是它的出现颠覆了摄影的物理含义，也颠覆了柯达公司自己。

2.2.1 数码相机的基本原理

如同第一台数码相机，今天的数码相机依然主要由光学镜头和 CCD 或 CMOS 感光器件组成。光学镜头与胶片照相机类似，这里不另做介绍。感光器件又分 CCD 和 CMOS。CCD 和 CMOS 的差别，在后面会讲到。感光器件的作用就是完成从光到电的光电转换，光的强弱影响着传感器上电荷的分布，电荷的分布变化影响着电流的强弱，光越强，电流越大，按电流强弱转换图形信号的过程，就是将光信号转换成电子信号的过程，即光/电转换过程。这从根本上动摇了传统摄影的地位，从而改变了人们熟知的银盐感光成像原理，电子芯片取代了传统胶片，计算机加软件取代了传统的暗房工艺。

前面讲过，早期的感光器件——135 胶卷的画幅是 $36\text{mm} \times 24\text{mm}$ ，在数码相机中，这种画幅被称为全画幅。受制作工艺及成本的影响，数字感光器件的画幅往往比全画幅小，在一些相机（如 Canon 20D-60D）中，感光器件是 $22.5\text{mm} \times 15.32\text{mm}$ ，在很多卡片式数码摄像机上，感光器件是 $7.18\text{mm} \times 5.32\text{mm}$ 。比如像图 2-19 这幅图，感光器件变小，所拍摄的场景会发生什么变化呢？最左边的图像是全画幅拍摄的效果，如果采用小一些的画幅拍摄的画面就如右边的两幅图像的效果。如果镜头不变，也就是焦距不变，当画幅越小时，呈现出来的画面也会越小。

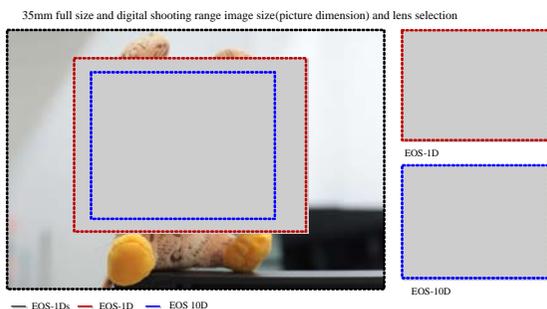


图 2-19 画幅变化效果

如何表述感光器件的画幅呢，或者叫成像靶面？在 CCD 出现之前，摄像机是利用一种叫做“光导摄像管（Vidicon Tube）”（俗称电视摄像管）的成像器件感光成像的。这是一种特殊设计的电子管，其直径的大小，决定了其成像面积的大小。因此人们就用光导摄像管的直径尺寸来表示不同感光面积的产品型号。比如型号为“1/1.8 英寸”的 CCD 或 CMOS，就表示其成像面积与一根直径为 1/1.8 英寸的光导摄像管的成像靶面面积近似。光导摄像管主要由镜头、光电靶、电子枪及偏转线圈等组成，如图 2-20 所示。

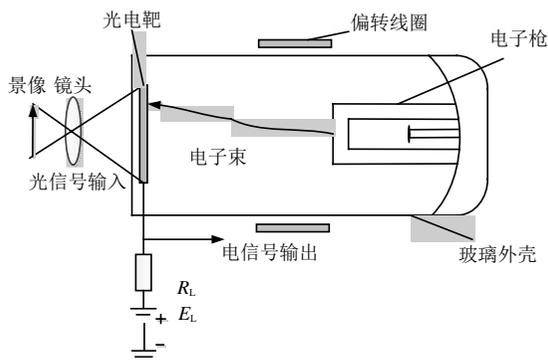


图 2-20 光导摄像管构成

按照这样一个表征特征，来看感光器件的画幅或者叫成像靶面是如下定义的（如图 2-21）。

135 全画幅： $36\text{mm} \times 24\text{mm}$ 。

APS-C 画幅： $22.5\text{mm} \times 15\text{mm}$ 。

4/3 英寸： $17.3\text{mm} \times 13\text{mm}$ 。

2/3 英寸：13.546mm×10.159mm。

1/1.8 英寸：11.289mm×8.487mm。

1/2.5 英寸：8.127mm×6.095mm。

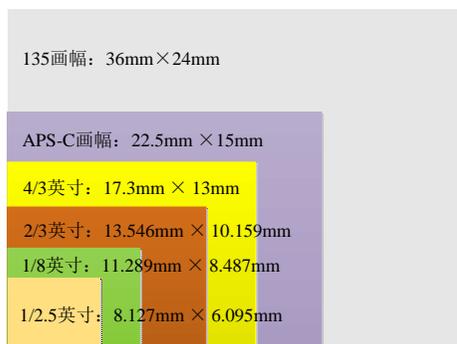


图 2-21 画幅定义

2.2.2 图像分辨率

图像分辨率是指数码相机或数码摄像机所能拍摄到的离散光点数，其基本单位是像素，一般表达方式为“水平像素数×垂直像素数”，比如 2048×1024，意思是数码相机所拍摄的照片水平方向有 2048 列像素，垂直方向有 1024 行像素，总像素是 200 万像素，所以这样的图像分辨率也可表达为“200 万像素”。图像分辨率指标的高低反映了图像清晰度的好坏。

2.2.3 图像信号处理器

在数码相机或数码摄像机中，直接连接 CCD/CMOS 的是图像信号处理器（Image Signal Processor, ISP），ISP 的主要作用是对前端图像传感器输出的信号做后期处理，主要功能包括自动曝光、自动白平衡、坏点去除、去马赛克、透镜阴影校正、颜色校正、伽马校正、边缘增强与 3D 降噪处理等（见图 2-22），ISP 处理水平在很大程度上决定了

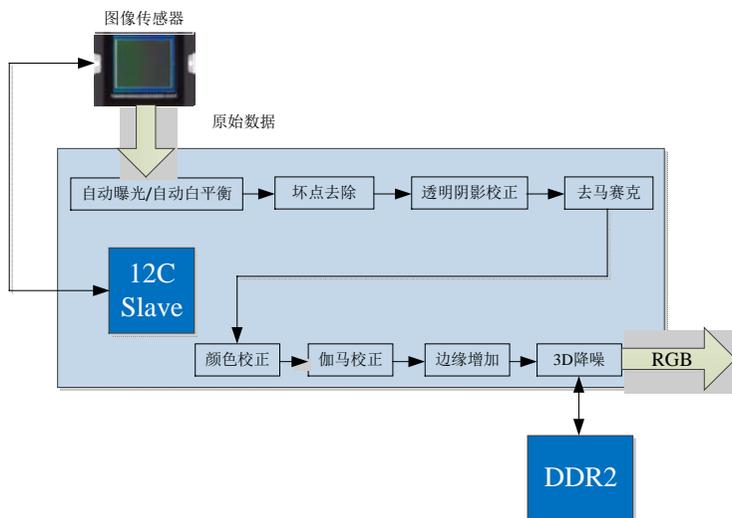


图 2-22 图像信号处理器工作流程

照相机/摄像机的成像质量。在数码相机中，ISP 不仅仅负责上述处理，还负责图像数据的压缩与存储。

2.2.4 压缩与存储

由于图像数据量大，为了方便存储与传输，一般采用 JPEG 标准压缩算法将数码相机拍摄出的图像进行压缩成 JPEG 文件。JPEG 全称为 Joint Photographic Experts Group，是一种有损的压缩算法，一般情况下，只要不追求图像过于精细的品质，是完全满足应用需求的，JPEG 通常压缩比率在 10:1~40:1，这样 JPEG 可以节省很大一部分存储卡的空间，从而大大增加了图片拍摄的数量，并加快了照片存储的速度，从而也加快了连续拍摄的速度，所以广泛用于数码相机中。

一些高端的数码相机可拍 TIFF 格式和 RAW 格式的照片。

TIFF 格式：扩展名是.tif，全名是 Tagged Image File Format。TIFF 是一种非失真的压缩格式（一般是 2~3 倍的压缩比）。这种压缩的文件在事后可完全还原，能保持原有图像颜色和层次，图像质量好，但占用空间大。

RAW 格式：很多的数码相机开始使用 RAW 格式保存图像。前面讲到，数码相机中都会有一个 ISP，负责对 CCD/CMOS 输出的原始数据进行图像处理，变成处理后的 RGB 图像数据，再进行图像压缩与存储。由于作为嵌入式芯片的 ISP 处理能力毕竟有限，而且有些用户往往希望后期能够按自己的意愿进行白平衡或其他色彩调节，于是就出现了 RAW 格式。RAW 文件不需要 ISP 进行处理，直接保存从 CCD/CMOS 出来的原始图像数据，加上光圈、快门、焦距等数据。摄影师拿到这些数据后，可以在 PC 上使用照相机厂家提供的图像处理软件任意调整色彩空间、锐化值、白平衡、对比度、降噪等，进行类似“暗房”的创造性制作，而且不会造成图像质量的损失，保持了图像的品质。由于 RAW 格式能够保存每个像素点更深的数字深度，为摄影师的创作保留了很大的空间，摄影师通过后期对图像色彩的调节，丰富了图片的层次，使其色彩更加细腻，存储文件大小也只有相对应的 TIFF 文件的一半左右，从存储空间节省上讲要比 TIFF 有明显的优势。

2.3 摄影学的技术性与艺术性

摄影不仅仅是一项技术，更是一门艺术。比如，人们不断地提高数码相机的分辨率，从最初的 100 万像素，逐步提高到数千万像素，每一次提高都涉及感光器件、图像信号处理器与存储器等众多部件的技术更新。不断提高数码相机分辨率的目的是为了更加真实地记录被摄场景的细节，而摄影师在拍摄时可能会选择虚化场景。这是为什么呢？

图 2-23 是应用小光圈拍摄的照片。几个模型娃娃都被拍得很清晰，整个画面的几个模型都是主角，画面内容的焦点有些分散。再看图 2-24，同样还是这几个模型娃娃，大光圈的应用导致背景被虚化，位置处于最前面的模型娃娃一下子凸显了出来，俏皮感顿时跃然纸上。这就是说，追求艺术性的摄影与追求技术先进性的摄影往往不一样，但另一方面，技术的先进性能够为艺术上的完美提供支持，而艺术上的更高追求又让人们更好地认识到技术先进性的价值所在。



图 2-23 一幅背景与前景同样清晰的图像



图 2-24 一幅背景被虚化的图像

参 考 文 献

- [1] CS 448A. Computational photography[S/OL]. 2010 [2013-12-31]. <http://graphics.stanford.edu/courses/cs448a-10/>.
- [2] Wikipedia. Computational_photography[EB/OL]. 2013 [2013-12-31]. http://en.wikipedia.org/wiki/Computational_photography.
- [3] Carnegie Mellon University. Computational Photography[EB/OL]. 2007[2013-12-31]. http://graphics.cs.cmu.edu/courses/15-463/2007_fall/.
- [4] COMP 790-096. Computational photography[EB/OL]. 2008[2013-12-31]. <http://www.cs.unc.edu/~lazechnik/research/fall08/>.
- [5] CSCI 129. Computational photography[EB/OL]. 2012[2013-12-31]. <http://cs.brown.edu/courses/csci1290/>.
- [6] Peter N. Belhumeur. Computational photography[EB/OL]. 2010[2013-12-31]. <http://www1.cs.columbia.edu/~belhumeur/courses/compPhoto/>.
- [7] 百度百科. 摄影[EB/OL]. 2013 [2013-12-31]. <http://baike.baidu.com/subview/7378/>.
- [8] 百度百科. 像差[EB/OL]. 2013 [2013-12-31]. <http://baike.baidu.com/view/23487.htm>.
- [9] 百度百科. 景深[EB/OL]. 2013 [2013-12-31]. <http://baike.baidu.com/view/7278.htm>.

第3章 颜色与颜色空间

3.1 颜色知觉与可见光

彩色摄影与颜色密不可分，对颜色的成像更是其主要研究的问题之一。颜色可以说是大自然的一个有机组成部分，人们每天都身处其中。自然界五彩缤纷的景物，如蓝天、白云、红花、绿叶、青山、碧水等，对这些景物影像的感受，都离不开颜色。那么，什么是颜色？颜色又是如何被感知的？

如图 3-1 所示，光线照射到被观察物体上，经过被观察物体的反射或者透射，入射到观察者肉眼中，光刺激形成神经生物电流信号，到达大脑，大脑对信号携带的信息进行处理，形成了对颜色的感觉，人们就分辨出物体的颜色类别以及明亮程度。

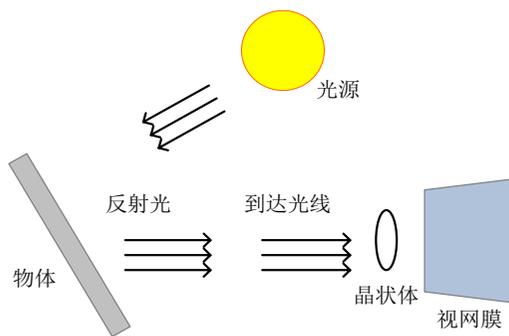


图 3-1 形成颜色感觉的三个要素

光源、被观察物体、观察者是形成颜色感觉的三个不可缺少的要素。颜色是光作用于观察者的视觉系统后产生的一系列复杂的生理和心理活动的综合效果。颜色感觉与听觉、味觉、嗅觉一样，都是外界刺激作用到观察者的感觉器官，在大脑中产生的认知。

光线的存在是人们感觉到颜色的前提。当光线很弱时，人们能感受的只是景物的灰度轮廓，难以分辨颜色；而当完全黑暗没有光线时，也就什么都看不到了。所以，颜色是光线刺激的结果，所有的颜色都离不开光线。但光线本身并不是颜色。根据《颜色术语》(GB/T 5698—2001) 标准，颜色概念的准确定义是光作用于人眼引起空间属性以外的视觉特性。即眼睛对光线的刺激产生了生物电流响应，响应信号传递到大脑，大脑对信号加工和处理后，形成了包括颜色感觉在内的视觉。客观上，光线仅仅只是自然界中的一种电磁波，本身是没有颜色属性特征的，是人类包括眼睛和大脑在内的视觉系统的感受赋予它以颜色特性。这是颜色与波长、频率等其他的电磁波属性截然不同的地方。

自然界电磁波的波谱范围很广，人眼并不是对所有波长的电磁波都能产生颜色的感受。如第 2 章所述，颜色是人的视觉系统对波长范围在 380~780nm 之间的电磁波，也

就是人们常说的可见光的感知结果。事实上，要想严格地界定可见光的波长范围是不太可能的。每个人都有一定的光辐射颜色感知波长范围的个体差异；即使同一个人，不同环境下，光辐射颜色感知波长范围也不尽相同。例如，当光线很强，而人眼又处于暗适应情况下，可感受颜色的电磁波长范围一般可扩大到 350~900nm。一般统计意义上，把波长在 380~780nm 的电磁波称为可见光。

不同波长的可见光辐射刺激视觉系统引起的颜色感受不同。单一波长的光辐射表现为一种颜色，即单色光或者光谱色。如图 3-2 所示，人的视觉系统对波长连续变化的单色光的颜色感受也是连续渐变的，可以把可见光的光谱划分为多个不同色名的连续区域。625~740nm 可见光给人以红色的颜色感受，520~565nm 可见光给人以绿色的颜色感受，435~500nm 可见光则给人以蓝色的颜色感受。同样，这也只是一个大致的范围划分，不存在严格的边界划分；此外，实验也发现，可见光波谱中，除了 572nm、503nm，478nm 这三个光谱点光辐射的颜色感受不受光强变化的影响，其他单色光的颜色感受随着光强的变化而略有变化。

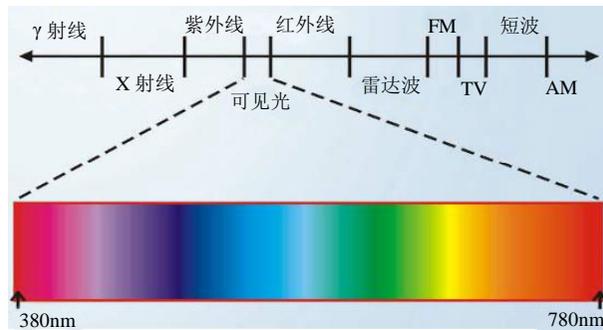


图 3-2 电磁波/可见光的波谱划分^[1]

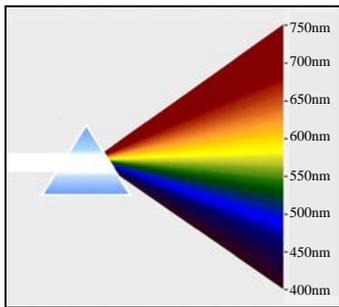


图 3-3 基于棱镜的白光色散^[2]

牛顿（Newton）通过色散实验揭示，白色阳光可分解为红、橙、黄、绿、蓝、靛、紫等光谱色组成的可见光谱带（见图 3-3）。同样，这些光谱色也可以合成出复色光，即原来的白色太阳光。

人们在自然世界中很难直接接触单色光，接触到的都是不同波长的单色光混合而成的复色光，如太阳光、月光、烛光等。鉴于自然世界中很难直接接触单色光，一般用光谱来描述可见光。不同成分单色光构成特定的复色光，对这种复色光的定义就是在 400~750nm 波段的辐射能量分布。

图 3-4 和 3-5 分别显示了水银灯和氙弧灯等几种典型光源发出的复色光的光谱。

光源发出的光照射到物体表面，对物体的颜色感受是通过物体对照射光线的反射或投射来认识的。不同的物体具有对照射光的不同程度的选择性吸收、透射和反射性质，称为物体的光谱特性。当光照射到物体上时，入射的光谱能量部分被反射，部分被吸收和散射，部分穿透物体。图 3-6 所示是光线照射到不同物体表面后，经吸收和反射，最后进入肉眼的光线的光谱能量分布，这就是不同物体给人以不同颜色感受的主要原因。

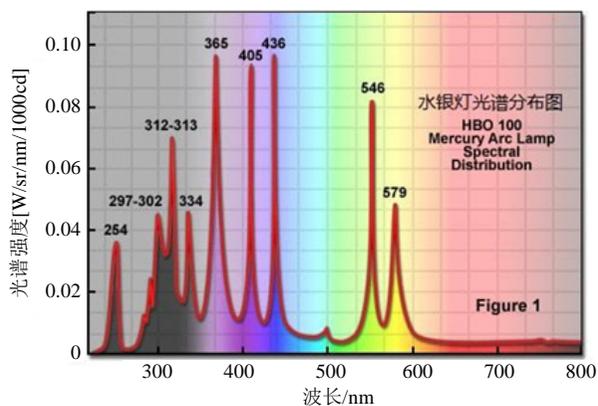


图 3-4 水银灯的典型发射光谱分布^[3]

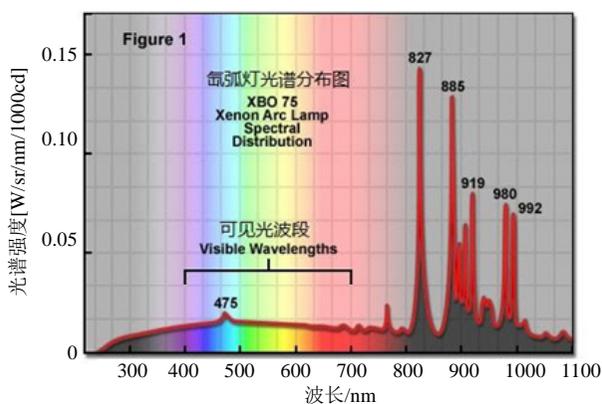


图 3-5 氙弧灯的典型发射光谱分布^[4]

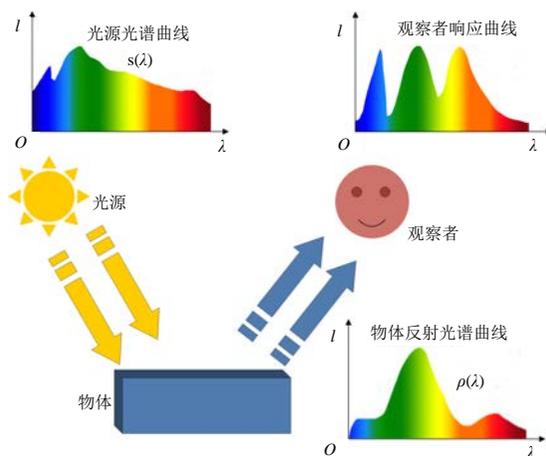


图 3-6 不同物体光谱特性示意^[5]

复色光的不同波长能量辐射的相对功率分布决定了人们对复色光的颜色感受。一定成分的复色光相应地会有一种确定的颜色感受。但是反过来，一种确定的颜色感受却并不是只对应着一种光谱组合。两种不同成分的复色光完全可能引起相同的颜色感受，这

种现象称为同色异谱。

需要注意的是：不同成分复色光的同色异谱现象是在一定的条件下成立的。如果改变观察环境的照明条件，可能颜色感受也会变得不一样。例如，在阳光下观察和荧光灯下观察，同样物体反射出的复色光给同一个观察者的颜色感受可能就会有明显的不同；或者，改变观察者也可能引起颜色感受的不同。对一个观察者来说是同色异谱，可能换个观察者却能感受到两者的颜色差异。

3.2 颜色视觉理论

人眼除了能够分辨可见光的明暗亮度以外，还能分辨颜色。一直以来，人眼如何分辨颜色的机理是最吸引人的研究领域之一。颜色是通过肉眼、大脑和人们的生活经验所产生的一种对光的视觉效应。对颜色的感觉不仅仅由光的物理性质决定，还包含心理等许多因素。例如，同色异谱理论指出，人类对颜色的感知结果往往也受到周围环境的影响。因此，这是一个相当复杂的科学问题，涉及了物理学、生理学、心理学等多种学科。

自从牛顿通过棱镜色散实验揭示单色光与复色光的关系后，人们一直在探讨颜色视觉的机理。不同的学者提出了不同的理论。最有影响力的主要是两种：杨-赫姆霍尔兹（Young-Helmholtz）提出的三色学说和赫林（Ewald Hering）提出的对立色学说（也称四色学说）。

1802年，杨（Thomas Young）提出三色学说，约半个世纪后，赫姆霍尔兹（Hermann von Helmholtz）对其进行了定量发展和完善。三色学说的核心观点是假定人眼视网膜上存在三种能够感受红、绿、蓝色（三原色）的光接收器，每一种受到刺激后引起的兴奋响应都能产生一种原色的感觉。一切颜色感受都由兴奋响应量的比例确定。三色学说不是理论推导的学说，而是建立在颜色混合实验规律的基础上，即不同比例的红、绿、蓝三原色可以混合出几乎所有的颜色。但三色学说不能很好解释色盲现象。

1878年，赫林通过研究发现，红和绿、黄和蓝、黑和白不能同时存在于任何一种颜色感受中。例如，有偏黄的红色，但是没有偏绿的红色。基于此，赫林认为，红和绿、黄和蓝、黑和白分别是三组对立色，假定人眼视网膜中存在能够响应对立色的白—黑、红—绿、蓝—黄三种视素。这三种视素中包含建设和破坏两种对立过程。当有光刺激，白—黑视素被破坏，产生白色感受，无光刺激时，白—黑视素被建设，产生黑色感受；类似地，红光对红—绿视素，蓝光对蓝—黄视素起破坏作用，绿光对红—绿视素，黄光对蓝—黄视素起建设作用。因为每种颜色都有一定的明亮度，因此不仅影响其对应视素的破坏与建设，也影响白—黑视素的破坏与建设。三种视素对立活动的组合就产生了各种颜色感受。四色学说由视觉现象总结而来，能够很好地解释色盲现象，但对三原色能够混合出任意颜色这一规律却没有给予说明，而这恰恰是近代色度学的重要基础。

三色学说和四色学说一时难以判断哪种正确，各自的支持者们长期争论不休。1965年前后，显微光谱—电生理学实验证实，人眼视网膜上存在两种感光细胞：锥体细胞（Cone）和杆体细胞（Rod），如图 3-7 所示。杆体细胞感光灵敏度高，但细节分辨能力低，不能分辨颜色。杆体细胞含有视紫红色素。视紫红色素物质的吸收光谱特性与人的暗视觉光谱光视效果极其一致。在暗视环境下含视紫红色素的杆体细胞主要起作用，产

生物体灰度轮廓感受。锥体细胞感光灵敏度低，但细节分辨能力高。锥体细胞有三种，即亲蓝锥体、亲绿锥体和亲红锥体，分别含有亲蓝、亲绿、亲红视色素，分别对蓝、绿、红光所在波段敏感。人眼对不同颜色的感受，就是光分别作用于三种锥体细胞的综合结果。图 3-8 所示是实验测定后绘制的不同感光细胞的感光—响应曲线。这在一定程度上验证了颜色视觉机理的三色学说。

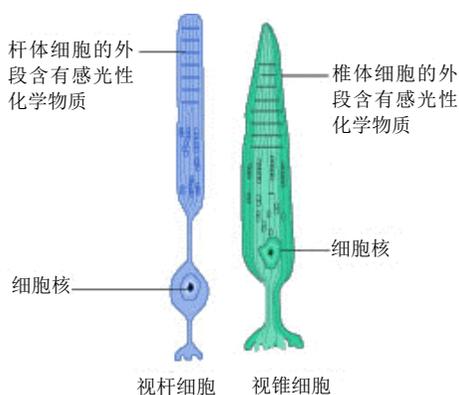


图 3-7 视网膜上的锥体细胞和杆体细胞^[6]

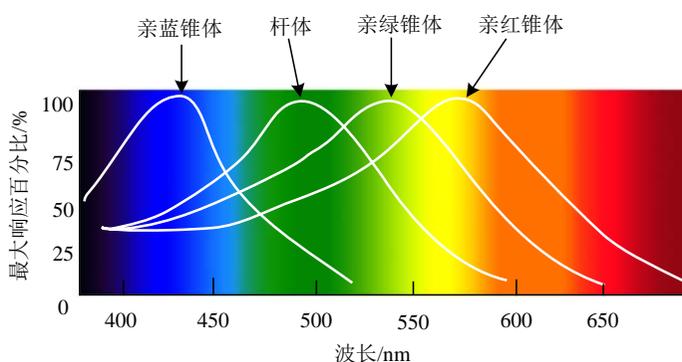


图 3-8 视网膜锥体和杆体感光细胞的感光特性^[2]

但后来更深入的显微光谱—电生理学实验又证实，离开锥体细胞几十微米位置处测得的 S 电位频谱响应呈现的亮度响应和对抗颜色响应结果又部分支持了四色学说。

实验证实了锥体细胞中确实存在蓝、绿、红三原色响应，但其产生的生物电信号在邻接细胞层进行着如同四色学说描述的对立色破坏—建设的信号处理过程。人们意识到两种学说都不全面，也并不是完全对立的。每种学说都一定程度上反映了颜色视觉的一个方面，把它们结合起来才形成更能反映颜色视觉客观机理的学说。

1971 年 Vos 和 Walraren 提出了相应的阶段学说。如图 3-9 所示，第一阶段是视网膜上三种锥体细胞对蓝、绿、红的颜色响应和明亮度响应，以及杆体细胞对明亮度的响应。杆体细胞对明亮度的响应直接成为暗视觉的明亮度响应 V' ，锥体细胞对明亮度的响应是由 R、G、B 输出的组合合成，直接构成亮视觉的明亮度响应 V ；第二阶段中，三种锥体细胞响应中，亲红锥体和亲绿锥体的一部分红色、绿色输出合成为黄色信号，之后进行信号混合运算，得到两种对立色响应 R-G 和 Y-B。亮视觉的明亮度响应 V 和暗视觉的

明亮度响应 V' 组合则构成了第三种白色与黑色对立色响应 $W-B$ 。视神经将这些经过处理后的信号传送给大脑视觉中枢，就形成了包含明亮度在内的颜色视觉。

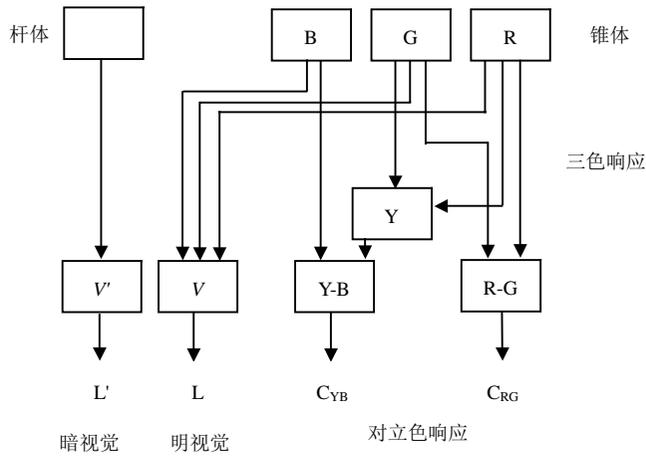


图 3-9 阶段颜色视觉学说

阶段学说较好地统一了三色学说和四色学说，并能以更合理的方式解释颜色视觉的机理和过程。但关于颜色视觉理论的探索研究仍未结束。

3.3 颜色感知特性

颜色可分为彩色和非彩色两大类。非彩色是指白色、黑色和介于黑白之间的各种深浅不同的灰色。非彩色只有描述白光亮度变化的明亮度的差异。彩色是除了黑白灰以外的各种颜色，不同的彩色通过三种特征属性描述，例如，色调、饱和度、明度。

1. 色调

色调 (Hue) 是视觉系统对物体呈现出的颜色的感觉，用红 (Red)、橙 (Orange)、黄 (Yellow)、绿 (Green)、青 (Cyan)、蓝 (Blue)、靛 (Indigo)、紫 (Violet) 等术语来刻画，黑、灰、白为无色彩。不同波长的单色光具有不同的色调。发光物体的色调取决于它辐射光的光谱组成，非发光物体的色调则取决于照明光源的光谱组成和物体的反射 (透射) 光谱特性。即色调取决于可见光的光谱构成，它是颜色彼此区分的外观特性，与颜色明暗无关。色调数目多于 1000 万种，颜色方面的专业人士可辨识的色调为 300~400 种。

2. 饱和度

饱和度 (Saturation) 是对颜色的纯洁程度的刻画。可见光光谱中的各种单色光是完全饱和的颜色。完全饱和的颜色就是指没有渗入白光所呈现的颜色，当一种颜色掺入其他光成分越多时，光谱带越宽，就说该颜色越不饱和。反过来，光的光谱带越窄，颜色的饱和度就越高。

3. 明度

明度 (Brightness) 是人眼感受物体明暗程度的描述, 明度是视觉系统对可见物体辐射光或发射光多少的感知属性。发光物体的亮度越高, 明度就越高, 非发光物体的反射比越高, 明度也就越高。黑色 (没有光) 的明度最低, 白色的明度最高, 在这两个极端之间的是不同明度的灰色地带。

明度是一个主观感觉值, 无法用物理设备测量, 但可以用亮度, 即辐射的能量来度量。根据国际照明委员会 (International Commission on Illumination, CIE) 的定义, 亮度是用反映视觉特性的光谱敏感函数加权之后得到的辐射功率, 并在 555nm 处达到峰值, 它的幅度与物理功率成正比。需要指出, 亮度虽然更易度量, 但它与明度并不是同一概念, 它可以认为是光的强度, 与明度的关系并不是简单的线性关系。

图 3-10 所示这个三维空间的纺锤体, 可以把色调、饱和度、明度这三个颜色的基本特征属性表示出来。垂直轴代表从白到黑的颜色明度变化; 水平圆盘剖面的径向表示颜色饱和度变化, 从圆心向圆周画一条直线段, 其上的色调和明度保持不变, 但从内向外, 饱和度越来越高; 水平圆盘剖面上的每个圆周用于表示具有相同饱和度和明度的连续变化的色调, 图中最大半径的圆周表示的是可见光谱中完全饱和的单色光的各种不同色调。

实际上, 图 3-10 就是用色调、饱和度和明度构造的 HSB 颜色空间表示, 色调用角度标定, 红色标为 0° , 青色标为 180° , 饱和度的深浅用半径大小表示, 明度用垂直轴表示。也可以使用类似的其他三个颜色特征属性来表示颜色, 这就构成了其他的颜色空间。更详细的介绍见颜色空间一节。

有光才有颜色。不同谱段光的叠加和分解就形成了颜色的相互混合, 从而造成人们的不同颜色感受。颜色混合可以是色光的混合, 也可以是颜料的混合。这两种混合方法得到的结果是不同的。在色光混合中, 混合后的色光的光谱分布是每个组成色光光谱分布的简单线性叠加, 因此称为颜色相加混合。在颜料混合中, 混合颜料对背景照明光的吸收光谱近似等于几种颜料各自的吸收光谱总和, 未被吸收的剩余光谱反射回来, 决定了对混合颜料的颜色感受, 因此称为颜色相减混合。

实验证实, 人的视觉系统只能感受颜色, 而不能感知和区分组成特定颜色刺激的光谱成分。只要是色度、饱和度和明度属性给人的感觉相同, 视觉系统就认为这是同一种颜色, 这就形成了前面指出的颜色同色异谱现象。1854 年, 格拉斯曼 (H.Grassmann) 在大量混色实验的基础上, 总结得到定性描述颜色混合规律的格拉斯曼定律, 格拉斯曼定律是现代色度学的重要基础, 共包括以下四条定律。

- (1) 人的视觉只能分辨颜色的三种变化, 即明度、色调、饱和度。
- (2) 在由两种成分组成的混合色中, 如果一种成分连续地变化, 混合色的外貌也连

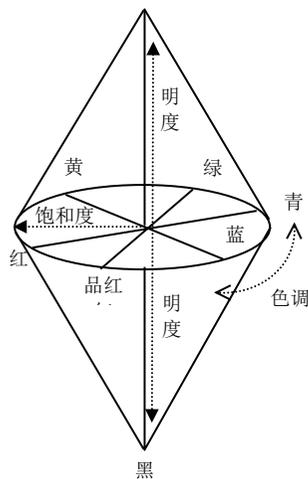


图 3-10 颜色的色调、饱和度、明度表示

续地变化。

(3) 颜色外貌相同的光, 不管它们的光谱组成是否相同, 在颜色混合中具有相同的效果。

(4) 混合色光的总亮度等于组成混合色的各种色光的亮度总和。

第一条定律的内容已在前面做了介绍。正是因为格拉斯曼定律第一条所指出的, 人的视觉只能分辨出颜色的明度、色调、饱和度属性变化, 国际照明委员会用颜色的明度、色调、饱和度这三个固有的并且是截然不同的特性来定义和区分颜色。

从第二条定律可以推想, 如果在由两种成分组成的混合色中, 固定一种成分, 使另一种成分连续地变化, 总可以使混合色变成白色或灰色。此刻的两种成分互为补色。显然, 每一种颜色都有一种对应的补色。如果某种颜色与它的补色以特定比例混合, 就会生成白色或灰色, 如果按其他比例混合, 就会生成与混合比例大的颜色接近的非饱和色, 以上推论称为补色律; 如果任意两种非补色混合, 则会产生中间色, 中间色的色调和饱和度随这两种颜色的色调及相对数量变化而变化, 以上推论称为中间色律。

第三条定律又称为等效律, 它指出凡是在视觉上相同的颜色都是等效的, 由此可进一步推出代替律。凡是视觉感觉上相同的颜色, 就可以相互代替, 不影响视觉感受效果。两种相同的颜色, 不论它们的光谱组成如何, 当与其他两种相同的颜色分别混合后, 两种混合色仍然相同; 同样, 两种相同的颜色, 分别减去其他两种相同的颜色后, 剩下的两种颜色仍然相同。

第四条定律又称亮度相加定律, 这与光作为电磁波的能量叠加是对应的, 不再解释。

3.4 颜色空间

为了便于对颜色认知的交流, 需要有一些规定的方法来量化精确表征各种颜色。颜色的定量表征涉及观察者的视觉生理和心理、照明与观察物理条件等诸多复杂因素。如何实现与人的颜色视觉特性一致的颜色量化表示技术, 以满足对颜色特性定量化和标准化的需要?

经过长期的探索实验, 国际照明委员会 1931 年提出了 CIE 1931 标准色度系统, 形成了 RGB 颜色空间。此后, 在此基础上, 科学家经过努力, 陆续提出了许多满足不同需求的颜色空间或颜色模型。那么, 什么是颜色空间或颜色模型呢?

颜色空间是表示颜色的一种数学方法, 用表示颜色的基本参数来描述和记录颜色, 通常用三维模型来定义颜色空间, 空间中的颜色用代表三个颜色属性参数的三个维度的坐标指定, 这些三维参数描述颜色在颜色空间中的位置, 但并没有告诉人们是什么颜色, 其颜色要取决于使用的坐标。

定义颜色空间使用的属性参数根据颜色空间针对不同对象和应用目的的不同而变化多样。例如, 对人, 可以通过色调、饱和度和明度来定义颜色; 对 CRT 显示设备, 可以用红、绿和蓝磷光体的发光量来定义颜色; 对打印或印刷设备, 可以使用青色、品红色、黄色和黑色的反射和吸收来指定颜色。

理论上, 可定义表示颜色的颜色空间的数目是无穷的。颜色空间有设备相关与设备无关颜色空间之分。设备相关的颜色空间是指生成的颜色与生成颜色的设备有关。下面

将介绍的 RGB 颜色空间就是指与显示设备相关的颜色空间。计算机显示器用 RGB 显示颜色，用像素值 $R=250$ ， $G=123$ ， $B=23$ 生成的颜色将随具体显示器的亮度和对比度的改变而改变。设备无关的颜色空间是指该颜色空间中与生成的颜色与生成颜色的设备无关。例如，下面将介绍 CIE $L^*a^*b^*$ 颜色空间就是设备无关的颜色空间，用该空间指定的颜色无论在什么设备上生成的颜色都相同。下面，具体地介绍几种与本书内容密切相关的颜色空间。

1) RGB 颜色空间

几乎所有的颜色空间都是从 RGB (Red, Green, Blue) 颜色空间导出的，因此首先介绍 RGB 颜色空间。RGB 颜色空间利用红、绿和蓝三基色的不同比例相加合成产生各种不同的颜色。R、G、B 取值范围都是 $0.0\sim 1.0$ 。在大部分 RGB 颜色空间应用中，R、G、B 的取值常用 8 比特表示，因此代表三基色的红、绿、蓝通道分别用 $0\sim 255$ 的整数表示强度。0 表示无光，最暗；255 表示强度最大，最亮。因为三种颜色都有 256 个亮度水平级，所以叠加就形成 1670 万种颜色。对于人眼分辨颜色能力而言，通过 RGB 颜色空间足以再现绚丽的世界。

RGB 颜色空间[见图 3-11 (a)]是混合加色型颜色空间，在电视机和计算机的图像颜色显示系统中广泛应用。例如，主要用在显示器、扫描仪等设备上，对于 CRT 显示器，R、G、B 分别表示显示器红、绿、蓝三种荧光粉受到电子枪轰击后激励产生的颜色；对于扫描仪，R、G、B 分别表示光电转换接收的通过红、绿、蓝滤光片的色彩。

RGB 颜色空间的优点是原理简单，使用 RGB 生成颜色容易实现。但 RGB 颜色空间也有不足，是与设备相关的颜色空间，在 RGB 空间中，对颜色指定不够直观，在 RGB 颜色空间中度量颜色时，视觉对颜色的感知是非线性的，如图 3-11 (b) 所示，R、G、B 通道的相关性也比较强。

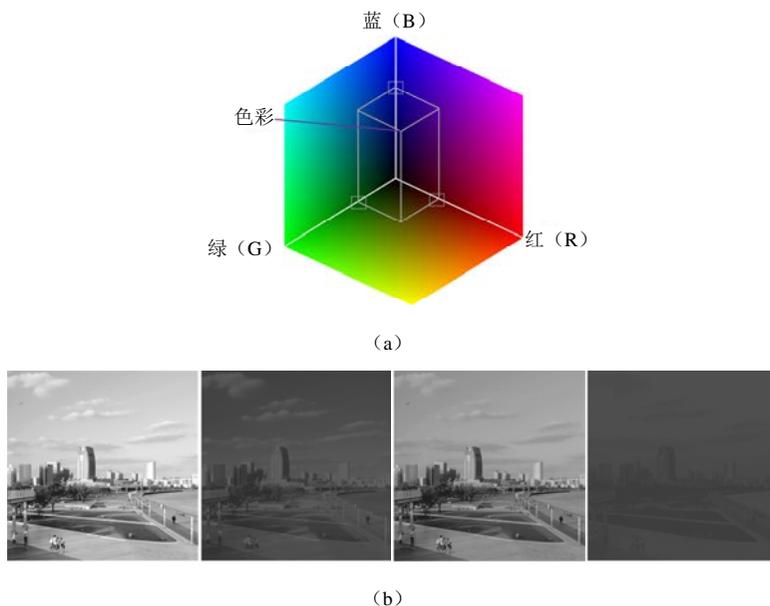
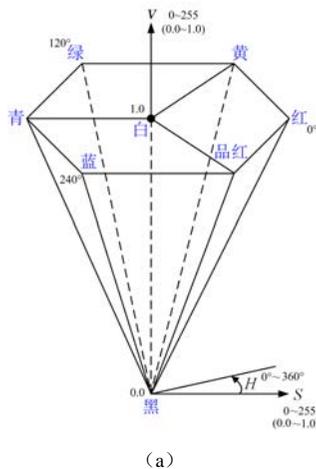


图 3-11 RGB 颜色空间，RGB 图像及 R、G、B 颜色通道分量^[7]

2) HSV 颜色空间

HSV (Hue, Saturation, Value) 颜色空间是根据颜色的直观特性定义的。如图 3-12 (a) 所示, 在 HSV 中, 色调 H 用水平圆周上的角度度量, 取值范围在 $0^{\circ}\sim 360^{\circ}$ 。按逆时针方向计算, 红色色调 H 定义为 0° , 绿色为 120° , 蓝色为 240° , 它们对应的补色调 H 取值分别为青色 180° , 品红色 300° , 黄色 60° 。饱和度 S 用水平圆周的从圆心到边缘的径向表示, 取值范围为 $0.0\sim 1.0$ 。亮度值 V 自底向上, 取值范围为 $0.0\sim 1.0$, 对应从黑色到白色。同样, 在用 8 比特表示 H、S、V 通道的取值时, 也把它们转换成 $0\sim 255$ 之间的整数。

HSV 颜色空间属于强度、饱和度、色调型颜色空间, 是面向计算机绘图时颜色编辑使用的较直观的颜色空间。HSV 是设备相关的颜色空间, 但优点在于用色调和饱和度描述对色彩的感知, 指定颜色的方式和对颜色的解释非常直观, 而且对消除光亮度的影响很有用。如图 3-12 (b) 所示, 4 幅小图分别是 HSV 方式表示的彩色图像以及 H、S、V 颜色通道的内容, 在把光亮度从颜色信息中分离出来后, 很容易在颜色编辑时选择出所需要的色调。



(a)



(b)

图 3-12 HSV 颜色空间, HSV 图像及 H、S、V 颜色通道分量^[8]

类似的强度/饱和度/色调型颜色空间还包括 HSB (Hue, Saturation, Brightness) 颜色空间, HSL/HLS (Hue, Saturation, Lightness) 颜色空间, HIS (Hue, Intensity, Saturation) 颜色空间, HCI (Hue, Chroma/Colourfulness, Intensity) 颜色空间, HVC (Hue, Value, Chroma) 颜色空间等。这些颜色空间都是与设备相关的类似颜色空间, 大同小异, 本书不再赘述。

3) YCbCr 颜色空间

YCbCr 颜色空间是在 ITU-R BT.601 和 BT.709 等标准中明确定义的,用于数字电视颜色空间。主要用于优化彩色视频信号的传输。如图 3-13 所示, Y 是亮度,通过把 RGB 输入信号的特定部分叠加得到。Cr、Cb 则表示色度,描述颜色的色调与饱和度。其中, Cr 反映 RGB 中红色部分与 Y 亮度值之间的差异。而 Cb 反映 RGB 中蓝色部分与 Y 亮度值之间的差异。YCrCb 颜色空间的重要体现在亮度信号 Y 和色度信号 Cb、Cr 是分离的,如果只有 Y 分量而没有 Cb、Cr 分量,那么图像就是黑白灰度图像。如果 YCbCr 取值用 8 比特表示, Y 的数值范围从 16~235, 16 表示黑电平, 235 表示白电平, Cb、Cr 的取值范围为 16~240, 使用 128 的偏移时范围为-112~112。



图 3-13 YCbCr 图像及 Y、Cb、Cr 颜色通道分量^[8]

因为人眼对区域内色差的变化没有对灰度的变化敏感。在图像质量损失忽略不计的前提下,可以降低 Cb、Cr 通道的采样率,从而有效压缩 Cb、Cr 通道的数据量。常见的 YCbCr 采样频率格式包括 4:4:4, 4:2:2, 4:1:1 以及 4:2:0。

4:4:4 是指 YCbCr 三个通道的采样率相同,因此在采用 YCbCr 颜色空间 4:4:4 格式生成的图像里,每个像素的三个分量信息完整,经 8 比特量化后,未经压缩的每个像素占用 3B。

4:2:2 是指图像中水平方向的色度采样率是 4:4:4 中的一半,即每个色差通道 Cb 和 Cr 水平方向采样率为亮度通道 Y 的一半。对非压缩的 8 比特量化图像,每两个水平方向相邻的像素占用 4B。

4:1:1 是指和亮度 Y 相比,在水平方向上对色度 Cb 和 Cr 只进行 4:1 的采样,对非压缩的 8 比特量化图像,每四个水平方向相邻的像素占用 6B,其视觉效果对低端用户和消费类产品仍可接受。

4:2:0 并不是说只有 Y、Cb,而没有 Cr 分量。而是对图像每行而言,只有一种色度分量以相对于 Y 分量 2:1 的采样率采集或存储;而上下相邻的两行存储不同的色度分量,即如果上一行 YCbCr 是 4:2:0,下一行 YCbCr 就是 4:0:2,再下一行又是 4:2:0,依此类推。对非压缩的 8 比特量化图像,每个 2 行 2 列相邻的 4 像素块占用 6B。可见, YCbCr 颜色空间最大的优点是表示图像数据时只需占用较少空间。

类似的还包括 YUV、YIQ、YPbPr 等颜色空间,都属于亮度/色度型电视系统颜色空间。YUV 是用在 PAL 和 SECAM 模拟彩色电视制式中的颜色空间, Y 表示亮度,也就是灰阶值, U 和 V 表示色度或色差分量; YIQ 是用在 NTSC 模拟彩色电视制式中的颜色空间; YPbPr 是用于高清晰度电视的颜色空间。无论是面向数字电视还是模拟电视的亮度、色度型颜色空间,主要用途都是电视信号传输,都是把 RGB 颜色空间转换成亮度和色度,将亮度和色度分离,用一个分量表示非色彩的亮度感知,用两个独立的分量

表示色彩的感知，目的是通过压缩色度数据，在有效播送彩色电视图像的前提下，显著压缩信号传播过程中图像的数据量。当需要黑白图像时也非常方便。这些颜色空间都是与设备相关的，而且在闭环系统中的使用条件也相当严格。

4) CMY 颜色空间

当背景光照射到物体表面时，物体将吸收一部分光线，并将剩下的光线反射，反射的光线就是人们看到的物体的颜色，这是一种减色色彩模式。CMY (Cyan, Magenta, Yellow) 颜色空间就是描述减色混合模式的适合印刷用的一种颜色空间，在印刷机和打印机在纸张或其他印刷材料上再现色彩时使用。有时也称为 CMYK (Cyan, Magenta, Yellow, black) 颜色空间，其中，C 代表青色，M 代表品红色，Y 代表黄色，K 代表黑色。它在印刷和打印系统中得到广泛应用，按照青色、品红色、黄色、黑色油墨/颜料的比例混合，白光中因颜料吸收的原因，减去不同数量的青色、品红色和黄色三种颜色，合成产生出各种彩色。理论上，青色、品红色和黄色颜料混合后应该得到黑色，但黄色颜料的透明度相对比较高，混合后变成深浊灰褐色，为了有效控制印刷时的色彩表现，改善印刷质量，在 CMY 基础上特意增加了黑色颜料分量，构成了 CMY(K) 颜色空间。

CMY(K) 颜色空间的优点是基于三基色，理论简单，在印刷领域使用时容易生成颜色。但 CMYK 颜色空间是和设备或者是印刷过程相关的，如工艺方法、油墨特性、纸张特性等不同条件有不同的印刷结果。因此 CMY(K) 也是与设备相关的颜色空间，颜色指定也不够直观，同样，视觉对颜色的感知是非线性的。此外，CMY(K) 空间具有多值性，对同一种具有相同绝对色度的颜色，在相同的印刷过程前提下，可以用多种 CMY(K) 数字组合来表示和印刷出来。

5) Lab 颜色空间

在颜色空间数学模型中，用三维坐标值描述颜色，解决了判断颜色相同与否的问题，但如何判断颜色相差多少仍是个问题。实验证明，在颜色空间坐标系中，两种不同颜色之间的距离值并不能正确地反映人们感受色彩感觉的差别大小，即上述颜色空间是非线性、不均匀的。

希望定义一类颜色空间，在这个均匀空间中每个点代表一种颜色，两点之间距离代表两种颜色的颜色差异，而对差异的度量和人眼辨色感受是相同的。

Lab 颜色空间就是 CIE 国际照明委员会 1976 年制定的这样一种均匀度量的颜色空间。它用数字化的方法来描述人的视觉感应，是一种基于生理特征的颜色空间。如图 3-14 (a) 所示，Lab 颜色空间由三个通道分量组成，L 通道是亮度，L 通道的颜色是从黑色（低亮度值）到灰色（中亮度值）再到白色（高亮度值）。a 和 b 是色彩通道。a 通道的颜色是从深绿色（低亮度值）到灰色（中亮度值）再到亮粉红色（高亮度值）；b 通道则是从亮蓝色（低亮度值）到灰色（中亮度值）再到黄色（高亮度值）。L 分量的取值范围是 0~100，表示从纯黑到纯白；a 分量从绿色到红色对应取值范围是 -128~127；b 分量从蓝色到黄色对应取值范围是 -128~127。

Lab 作为基于生理特征的一种颜色空间，用数字化方法描述人眼视觉感应，这就意味着 Lab 是一种与设备无关的颜色空间。此外，从坐标轴不难看出，Lab 使用黑—白、蓝—黄、绿—红对立色坐标轴，意味着 Lab 建立在赫林对立色视觉理论的基础上。

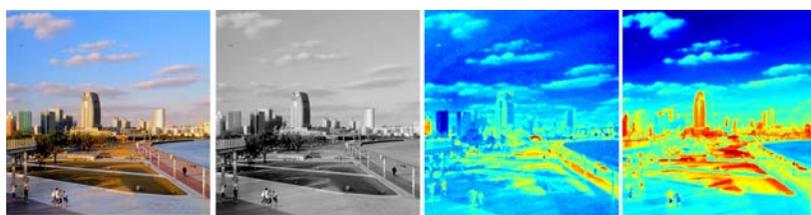
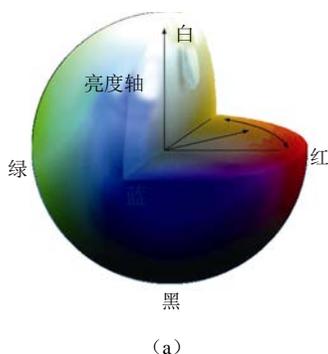


图 3-14 Lab 颜色空间, Lab 图像及 Lab 颜色通道分量^[9]

与 YCrCb 颜色空间类似, Lab 颜色空间也是一种亮度/色度型颜色空间, 但 Lab 颜色空间最大的特点是, 它是一种均匀的颜色空间, 色差与人眼感知的差异更加一致。同样是均匀颜色空间的还包括 CIE 提出的 CIE 1976 Luv 颜色空间和 CIE DE 2000 等。

需要指出, 现有的颜色空间还没有一个能完全符合人的视觉感知特性、颜色本身的物理特性或发光物体和光反射物体的特性。因此, 随着科技的进步, 人们还在不断地继续开发形形色色的颜色空间。

3.5 基于彩色滤镜阵列的彩色感知

在彩色光电摄影中, 可以通过分光, 使用三个传感器分别捕获红、绿、蓝三个通道的颜色分量。但为了降低成本, 绝大部分成像系统中采用单芯片 CMOS、CCD 传感器结合彩色滤镜阵列 (Color Filter Array, CFA) 的方式捕获彩色图像, 虽然牺牲了物理分辨率, 但显著简化了结构, 降低了价格。做法是在每个感光像素的上面覆盖一个与物理像素相同面积大小, 但分别只让红、绿、蓝光线透过的滤镜。

3.5.1 Bayer 格式的 CFA 彩色滤镜阵列设计

图 3-15 所示是 Bryce Bayer 于 1976 年发明并注册专利的 Bayer 格式彩色滤镜阵列, 被广泛运用于现代数码相机、摄像机和手机摄像头中, 是实现单片 CCD 或 CMOS 传感器拍摄彩色图像的主要技术之一。Bayer 阵列模拟人眼对色彩的敏感程度, 采用 1 红 2 绿 1 蓝的排列方式将灰度信息转换成彩色信息。

如图 3-16 所示, 彩色光线经过 Bayer 格式彩色滤镜阵列后, 传感器实际每个物理像素仅有一种颜色信息, 形成的扩展名为 .raw 格式图像有红绿蓝间隔的明显马赛克感觉。

如何得到每个物理像素位置处的 RGB 通道分量并去除马赛克？

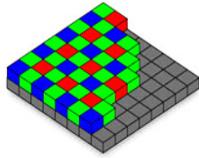


图 3-15 Bayer 格式单芯片 CMOS/CCD 的彩色滤镜阵列^[10]

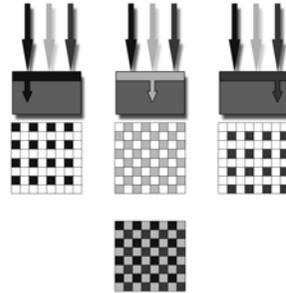


图 3-16 Bayer 格式图像获取示意^[10]

每一个像素仅仅包括了光谱的一部分，必须通过插值来实现每个像素的 RGB 值。为了从 Bayer 格式得到每个像素的 RGB 格式，需要通过插值填补缺失的两个色彩通道分量。如图 3-17 所示，反马赛克算法需要利用特定插值计算，根据每个像素旁边像素的红、绿、蓝通道值，通过加权平均，插值算出每个像素点缺失的颜色通道取值，获得最终图像。插值的方法有很多，典型的如邻域、线性、 3×3 插值等，在具体方案中，通过速度与质量的权衡决定最终采用哪种。不同的插值补偿算法差异取决于参与插值的邻近像素的数目和各自权重。

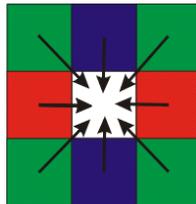
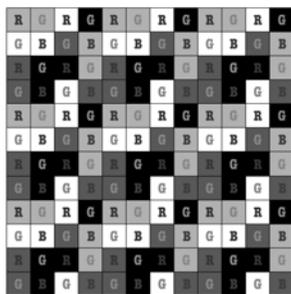


图 3-17 邻域插值方法计算每个像素缺失的颜色通道^[10]

3.5.2 其他创新的 CFA 彩色滤镜阵列设计

Bayer 格式是采用最为广泛的彩色滤镜阵列。除此之外，索尼公司提出了如图 3-18 所示的 Assorted Pixels 彩色滤镜阵列设计^[11]，主要思想是不仅对颜色，对亮度也进行马



(a) Assorted Pixels 设计



(b) 传统 Bayer 结果



(c) Assorted Pixels 结果

图 3-18 Assorted Pixels 彩色滤镜阵列及插值后的效果对比^[11]

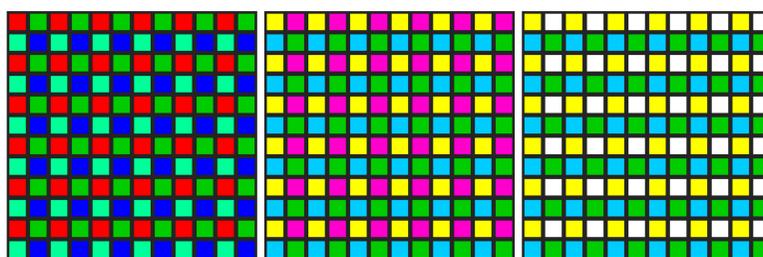
赛克设计，以拍摄出更好的图像颜色效果。

富士公司提出了如图 3-19 所示的 X-Trans 传感器彩色滤镜阵列技术^[12]，通过彩色滤镜排列中红、绿、蓝的无周期性（无序性）排列，显著抑制了彩色图像中摩尔纹和杂色的产生。



图 3-19 X-Trans 传感器彩色滤镜阵列及效果对比^[12]

如图 3-20 所示，其他创新的传感器彩色滤镜阵列设计还包括 Sony DCS F828 芯片中使用的红/绿/蓝/翠绿四色滤波以及减色法三原色+绿色构成的四色滤波^[13]，四色滤波可以得到更好的光灵敏度。还有 JVC 公司提出的黄/蓝绿/绿/白四色滤波，它具有更宽的动态范围。



(a) 红/绿/蓝/翠绿 (b) 减色法三原色+绿色 (c) 黄/蓝绿/绿/白

图 3-20 其他创新的传感器彩色滤镜阵列设计^[13]

作为本章的结束，最后指出，到目前为止，虽然已经有了不少关于颜色感知的理论、技术和标准，但似乎目前还没有哪种颜色理论普遍被人们接受。关于颜色、颜色视觉感知、颜色空间理论的研究仍然在不断深入。

参 考 文 献

- [1] 雷士照明. 产品与应用解决方案[EB/OL]. (2009-03-03) [2014-05-25]. <http://www.nvc-lighting.com.cn>.
- [2] 中国科学院物理研究所. 科普文章[EB/OL]. (2010-03-03) [2014-05-25]. http://www.iop.cas.cn/kxcb/kpwz/shzwlzl/201009/t20100919_2965961.html.
- [3] ZEISS. Education in Microscopy and Digital Imaging[EB/OL]. (2012-03-03) [2014-05-25]. <http://zeiss-campus.magnet.fsu.edu/articles/lightsources/mercuryarc.html>.
- [4] ZEISS. Education in Microscopy and Digital Imaging[EB/OL]. (2013-03-03) [2014-05-25]. <http://zeiss-campus.magnet.fsu.edu/articles/lightsources/xenonarc.html>.
- [5] 威福光电科技有限公司. 产品展示. (2008-05-03) [2014-05-25]. <http://www.wavegd.com/newsshow.asp?id=155&sid=14>.

- [6] 互动百科. IPAD[EB/OL]. (2010-03-03) [2014-05-25]. <http://www.baik.com/ipadwike>.
- [7] 互动百科. 图片[EB/OL]. (2011-03-03) [2014-05-25]. <http://www.baik.com/ipadwike>.
- [8] CSDN 博客. RGB 和 HSV 颜色空间[EB/OL]. (2005-03-03) [2014-05-25]. <http://blog.csdn.net/carson2005/article/details/6243892>.
- [9] CSDN 博客. Lab 颜色空间[EB/OL]. (2012-03-03) [2014-05-25]. <http://blog.csdn.net/carson2005/article/details/7200440>.
- [10] 石鑫华视觉网. 摄像机[EB/OL]. (2010-03-03) [2014-05-25]. <http://shixinhua.com/camera/201302/314.html>.
- [11] SONY. 相机[EB/OL]. (2009-03-03) [2014-05-25]. <http://www.sony-krc.com/jpg>.
- [12] 富士. 相机[EB/OL]. (2012-03-03) [2014-05-25]. <http://www.fujifilm.com.tw/information/2012/0906/index.html>.
- [13] Columbia University. projects[EB/OL]. (2009-03-03) [2014-05-25]. http://www1.cs.columbia.edu/CAVE/projects/hdr_ap/images/assortment_types.jpg.

第 4 章 数字感光器件

4.1 数字感光器件的发展历程

首先，来回顾一下数字感光器件的发展历程。第一张永久性的照片拍摄于 1826 年，是由法国人尼埃普斯拍摄的，叫做《窗外的景色》，也叫《鸽子窝》，记录的是从他家楼上看到的窗户外场景，如图 4-1 所示。



图 4-1 第一张永久性照片^[1]

当时，尼埃普斯拍摄所用的感光物质是沥青加上薰衣草油。它的片基是铅锡合金板。1826 年之后，感光技术得到充分发展，到 1888 年，美国柯达公司生产出了一种新型感光材料——柔软且可卷绕的“胶卷”（见图 4-2），这是感光材料的一个飞跃。它是将卤化银涂抹在聚乙酸酯片基上面，当有光线照射到卤化银上时，卤化银转变为黑色的银，经显影工艺后固定于片基，成为人们常见的黑白负片。彩色负片则是涂抹了三层卤化银以表现出三原色。

第一个数字感光器件 1969 年在贝尔实验室诞生，是由物理学家维拉·博伊尔（Willard S. Boyle）以及乔治·史密斯（George E. Smith）发明的。当时，博伊尔和史密斯发明一种被命名为“电荷‘气泡’元件”的装置，在光电效应的激发下，该器件表面产生电荷，从而组成数位影像。1974 年，他们用 CCD 搭建了一个分辨率为 100×100 像素的摄像机，图 4-3 显示了博伊尔和史密斯正在调试他们研制的 CCD 摄像机。作为 CCD 的发明者，他们于 2009 年获得诺贝尔物理学奖。



图 4-2 柯达公司生产的胶卷^[2]



图 4-3 贝尔实验室 1974 年研制的 CCD 摄像机^[3]

CCD 的英文全拼是 Charge-Coupled Device，中文全称为“电荷耦合器件”，如图 4-4 所示，它的特性是利用交互变化的电压在半导体表面传递电荷。大多数重要科技发明都需要经过一段时间的酝酿才能形成真正能用的装置，CCD 的发明是个意外，从概念的提出到得到可用的系统只用了几个月的时间，到大规模使用只用了几年的时间。

另外一款数字感光器件是 CMOS，英文全拼为 Complementary Metal Oxide Semiconductor，中文全称是“互补式金属氧化物半导体”，是一种制备芯片的工艺。CCD 一般是采用 TTL 工艺制备的，采用 CMOS 工艺制备的感光器件俗称为 CMOS，如图 4-5 所示。它功耗低，可以把图像采集单元与信号处理单元集成到同一块芯片上，因而便于大规模批量生产，极大降低了生产成本，适用于要求小尺寸、低价格的应用。现在人们使用的手机、桌面式摄像机等都要求低成本，都采用了 CMOS 器件。随着 CMOS 技术的发展，在很多应用场合 CMOS 将会取代 CCD。

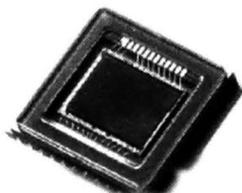


图 4-4 CCD^[4]

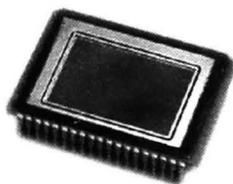


图 4-5 CMOS^[5]

4.2 数字感光器件的结构

4.2.1 CCD 的结构

下面介绍数字感光器件的结构。首先介绍 CCD 的结构，CCD 分为三层，最上面一层是微透镜，中间层是滤色层，最下面一层是感光层，如图 4-6 所示。

微透镜用于更好地收集光线。如图 4-7 所示，最上面是微透镜层，相当于在感光层上加了一副眼镜，能够扩展单个像素的受光面积，感光面积扩大的多少取决于微透镜的表面积。也就是说，微透镜在每一个 CCD 前面都有一个往外凸出的部分，很小，肉眼

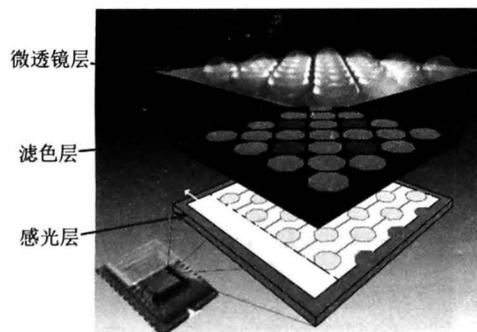


图 4-6 CCD 结构示意图^[6]

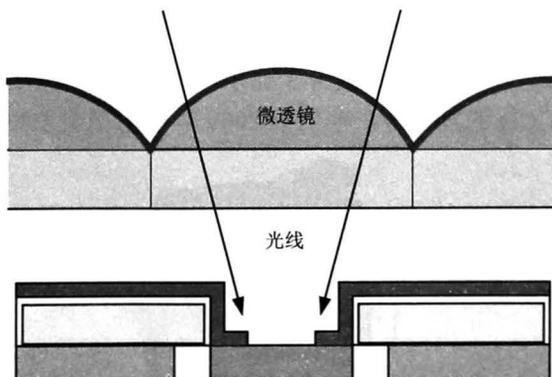


图 4-7 CCD 微透镜

几乎看不见。

微透镜下面就是滤光片或者叫滤色片，英文叫 Color Filter，这种滤光片是采用不同颜色的过滤纸放在感光器件的前面，能够滤掉不同颜色的光，如图 4-8 所示。比如红色滤光片滤掉蓝色和绿色，只剩下红色光进到 CCD 里面来；绿色滤光片就是让绿色光进来；蓝色就是让蓝色光进来。如图 4-9 所示，滤色层中左边是红色滤光片，中间是绿色滤光片，右边是蓝色滤光片，左边只有红色分量进入感光层，中间只有绿色分量进入感光层，右边只有蓝色分量进入感光层。人们从感光层得到同一个场景的三幅灰度图，分别对应场景的三个颜色通道，颜色还原方法可以把三幅灰度图还原成关于这个场景的彩色图。

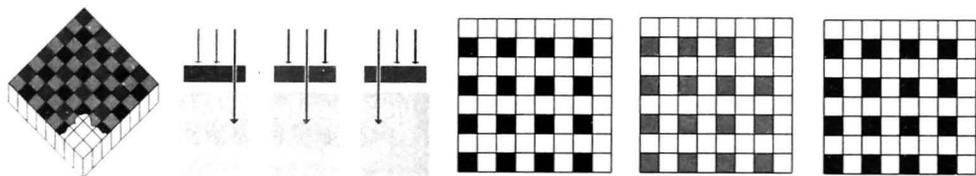


图 4-8 CCD 滤光片

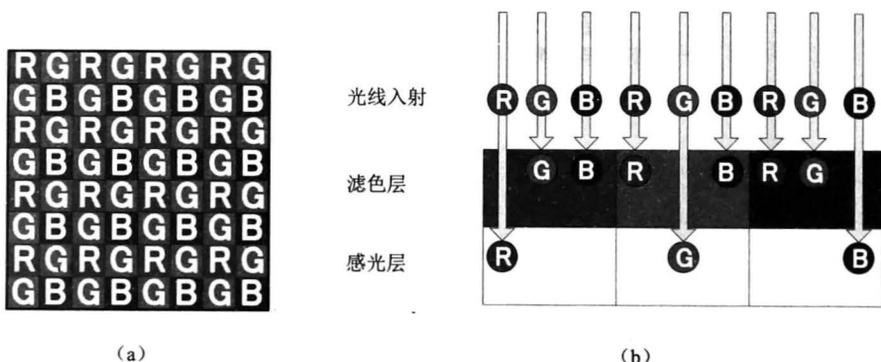


图 4-9 RGB 原色分色法

常用的滤光片有两种：一种是 RGB 原色分色法，就是滤光片里面是由 RGB 三色交叉排列所组成的滤光片。这种滤光片的特点是画质锐利，色彩真实，但是噪声大。

还有一种是索尼公司生产的 CCD 中常用的，叫做 CMYK 互补分色法。其中，C——青色，M——粉红色，Y——黄色，G——绿色，CMY 这三种颜色和 RGB 是互补的，所以叫互补分色法。由 CMYK 四种颜色组成的滤光片，如图 4-10 所示，这种滤光片对噪声的抑制能力强，而且多了一个黄色滤光片，在色彩的分辨上更加仔细，但却牺牲了部分影像的分辨率。

滤光片的下面就是感光层，感光层是 CCD 的核心。它是由很多个感光二极管组成的，也就是说，感光二极管是感光层的核心。感光二极管就是负责把光子转化为电荷，图 4-11 中的方块为感光二极管，感光二极管把光子转换成电荷，这些电荷就暂时储存在并行寄存器中。并行寄存器存储的电荷，在特定时序的控制下传递到下面的串行寄存器中，再经过电荷到电压的转换、放大、ADC 变换，变成计算机能接收的数字信号。

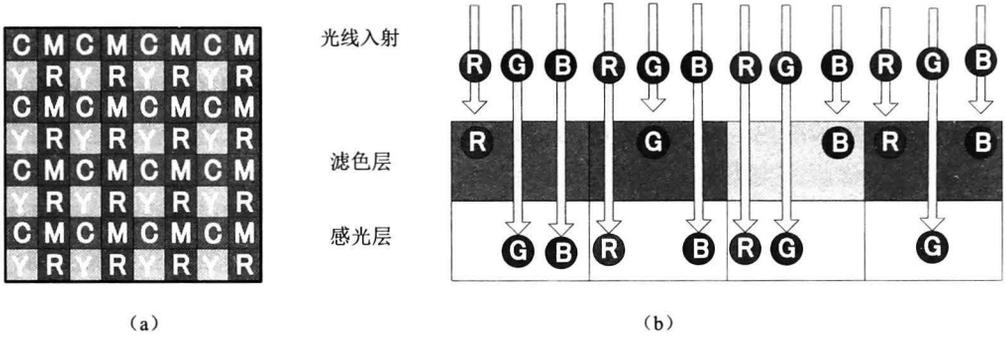


图 4-10 CMYG 互补分色法

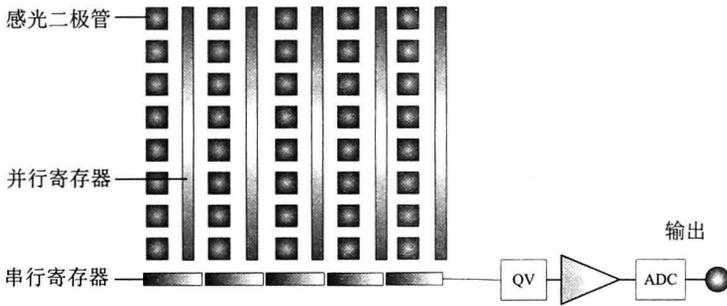


图 4-11 CCD 感光层示意图

图 4-12 描述了光转换成电荷，同时传输到计算机的过程。在时序的控制下：第一步，CCD 接收光线的照射产生电荷[见图 4-12 (a)]；第二步，电荷在外加的电压作用下移动到并行寄存器中[见图 4-12 (b)]；第三步，在时序的控制下，电荷被转换为电压，电压经 ADC 转换为数字信号[见图 4-12 (c) 所示]；第四步，在时序的控制下，依顺序信号从并行寄存器组合到串行寄存器，再传到外面的缓冲区中。这就是感光层中，光转换为电荷，电荷转换为电压，电压经放大及转换以后，转换成数字信号存储起来的全过程。

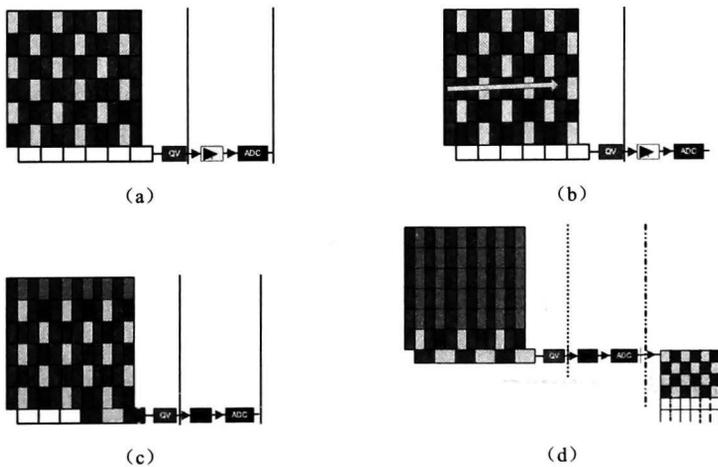


图 4-12 CCD 感光层工作原理

把 CCD 的感光过程可以形象地比喻成水滴滴入小桶的过程。CCD 的结构就像是一排排输送带上并排放满了的小桶，光线就像水滴洒入各个小桶，每个小桶就是一个像素。按下快门拍照的过程，就是按照一定的顺序测量某一短暂的时间间隔中，小桶中落进了多少“光滴”，并记在文件中，如图 4-13 所示。一般的 CCD 每种颜色通道的宽度用 8 位来记录，即其小桶上的刻度有 256 格，也有的是 10 位，甚至 12 位，10 位或者 12 位的 CCD 在记录色彩时可以更精确。



图 4-13 CCD 结构模拟图

图 4-14 描述的是 CCD 摄像机原理。右边是 CCD 芯片图，主要是由感光二极管组成。感光二极管把光转换为电荷，电荷传到电压转换器，放大后传输到 CCD 控制电路板。左边描述的是 CCD 控制电路板的功能模块，该控制电路板用于产生时钟，并把 CCD 芯片输出的信号施加适当的增益，转化为数字信号后输出到计算机中。CCD 控制电路板主要分为两部分，上部分为时钟产生器，为 CCD 芯片提供时序，下面部分就是 ADC 转换以及与计算机的接口驱动组成的。

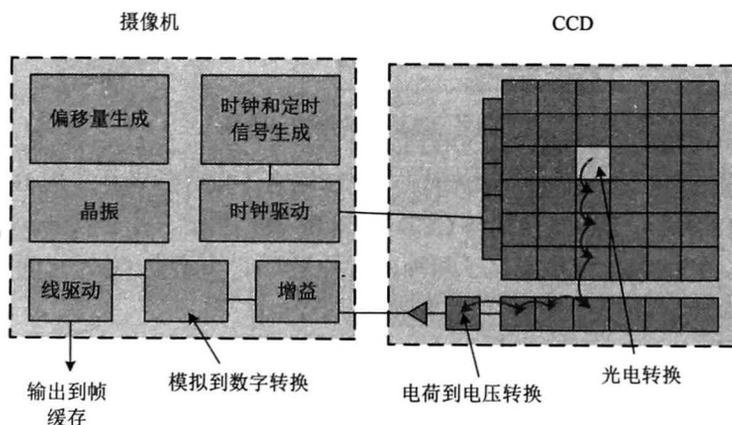


图 4-14 CCD 摄像机原理图

电荷在 CCD 中的传输有几种模式，最常见的模式是行间传输模式 (Interline Transfer)，这种传输模式就是前面讲到的模式。感光二极管把光转换为电以后，信号传递到并行寄存器，再传到串行寄存器，再传到计算机系统。这种方式的特点是快速，即曝光和数据读出可同时进行，如图 4-15 所示。另外可采用软件控制的电子快门进行工作。它的缺点是，在每个感光二极管边上都有一个并行寄存器，这个寄存器占据了感光单元的面积，使每一个像素感光面积减小。

相对应的还有一种模式是全帧传输模式，如图 4-16 所示。这种模式没有并行寄存器，感光二极管产生电荷，电荷转移到串行寄存器中，串行寄存器再经过放大、模数转换以后输出。在这种模式中，它的信噪比高，因为每一个像素的面积比行间传递模式感光面积大。它的缺点就是成像速度慢，必须借助机械快门来控制放光量。

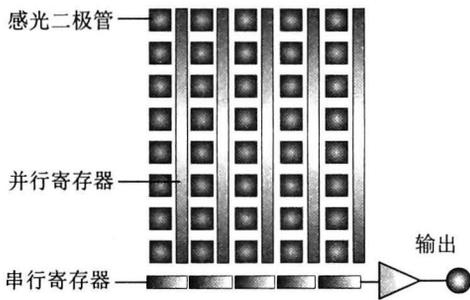


图 4-15 行间传输模式

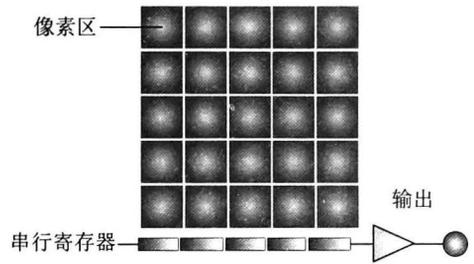


图 4-16 全帧传输模式

4.2.2 CMOS 的结构

CMOS 的结构总体上和 CCD 类似，包括三层：微透镜层、滤光片和感光层，如图 4-17 所示。

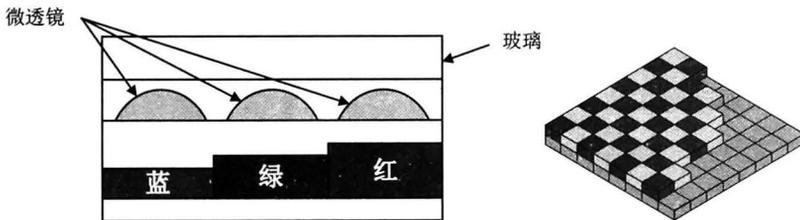


图 4-17 CMOS 的结构示意图

微透镜层和 CCD 是一样的，滤光片与 CCD 略有不同，后者是采用 Bayer 模式的滤光片，如图 4-18 所示。Bayer 模式与 CCD 的 RGB 模式相近，由于在 Bayer 模式中比较强调绿色分量，它的绿色分量定义了两种，一种是 Gr，一种是 Gb。它的整个画面的亮度和色彩的控制主要靠绿色分量。

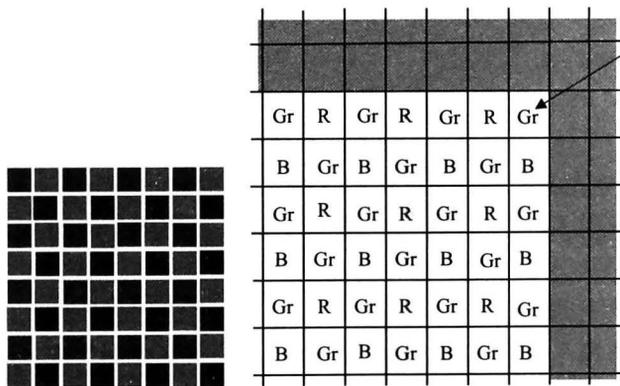


图 4-18 CMOS 采用 Bayer 模式的滤光片

CMOS 的感光层与 CCD 差别很大，如图 4-19 所示，当然它的核心器件还是感光二极管 (Photodiode)，但是在每一个感光二极管边上，也就是在每个像素里面，不仅仅是有感光二极管还包含了一个放大器和一个模数转换电路。所以，CMOS 中每一个感光器件里面都包含了一个感光二极管和三颗晶体管。每一个像素中感光面就更小了，一般每

个感光二极管的面积只有像素面积的 70%，有 30%没有感光。

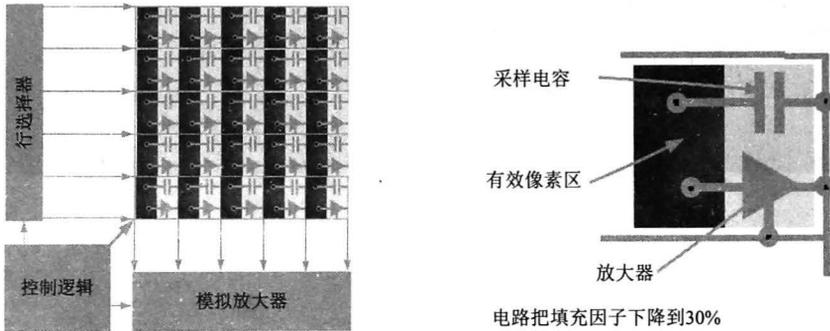


图 4-19 CMOS 的感光层结构

在工艺上 CCD 与 CMOS 也不同。CCD 是采用 TTL 工艺，CMOS 是采用 CMOS 工艺，CCD 采用总放大器，而每一个 CMOS 感光单元本身带放大器，如图 4-20 所示，这种方式的好处就是 CMOS 可以采用并行模式读出，读出速度比较快。但是由于每一个感光单元都有一个放大器，所以放大的效果难以达成一致。

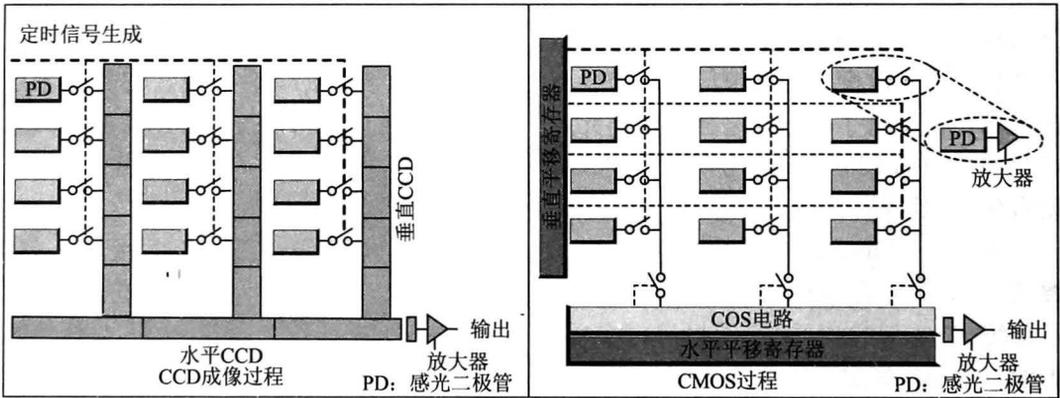


图 4-20 CCD 与 CMOS 的区别

CMOS 快门模式也比较特殊，一种是最常见的电子卷帘快门 (Electronic Rolling Shutter)。目前大多数 CMOS 传感器采用这种快门。对任一像素，在曝光开始时先将其清零，然后等待曝光时间过后，将信号值读出。因为数据的读出是串行的，所以清零、曝光、读出也只能逐行顺序进行，通常是从上至下。对高速运动的物体会产生明显的变形。例如，如果数据的读出速度是每秒 20 帧，那么图像顶部和底部的曝光先后差异将多达 50ms。

这样的话，对有些场景会产生不好的效果。比如图 4-21 左边的图，在某一时刻，上面偏暗，下面偏亮；到了下一时刻，上面又偏亮，下面偏暗，画面曝光的效果不一致。当然，这是指画面有非常明显的闪烁的情况下会导致的效果，一般情况下是没有这种效果的。图 4-22 显示的是高速运动的物体，整个场景可能发生偏移。

另外一种 CMOS 快门方式是全局快门 (Global Shutter/Snapshot Shutter)。与前一种方式最主要区别是在每个像素处增加了采样保持单元，在指定时间达到后对数据进



图 4-21 电子卷帘快门曝光时间差^[7]

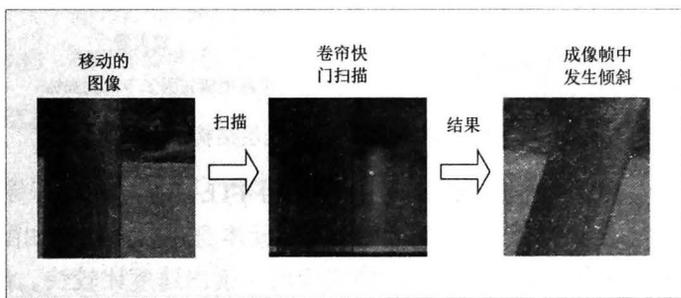


图 4-22 高速运动的物体发生偏移

行采样，然后顺序读出，虽然读出的像素仍然在进行曝光，但存储在采样保持单元中的数据却并未改变。这种快门方式的缺点是增加了每个像素的器件数目，使得 CMOS 感光面比例降低了。另外采样保持单元还引入了新的噪声源，使得噪声也增加了。但是全局快门的 CMOS 适合于拍摄高速运动的画面，比如说有快速闪烁的闪光灯，或者是有快速运动的物体。

传统的 CMOS 是前照式结构。2008 年，索尼公司推出一种新的 CMOS 接口——背照式 CMOS^[8]。背照式 CMOS 把感光层的器件调转方向，让光能从背面直射进去，避免了传统 CMOS 传感器结构中，光线会受到微透镜和感光二极管之间的电路和晶体管的影响，从而显著提高感光效率，大大改善低光照条件下的拍摄效果。背照式 CMOS 传感器的结构如图 4-23 所示，图 4-23 (a) 显示的是传统的 CMOS 结构，看到光线通过微透镜后还需要经过电路层才能到达受光面，中途光线必然会遭到部分损失（包括被阻挡或被减弱）。图 4-23 (b) 显示的是背照式 CMOS 结构，在改变了结构后，光线通过微

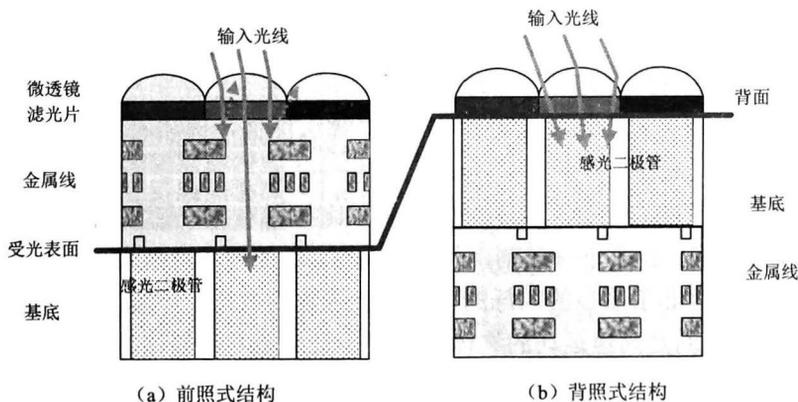


图 4-23 前照式与背照式 CMOS 结构图

透镜后就可以直接到达感光层的背面，完成光电转换，从进光量上改善了感光过程。而且由于感光二极管离微透镜更近了，光线在传输过程中的损耗也就更小了，因而背照式 CMOS 传感器比传统 CMOS 传感器在灵敏度上有质的飞跃，结果就是在低光照度下的画质有很大提升。

4.3 数字感光器件的指标

4.3.1 感光度

顾名思义，感光度就是指照相机对光线的敏感程度。无论是胶片相机时代还是数码相机时代，感光度都是用户衡量相机对光线感应程度的最直接的指标。

在胶片相机时代，感光度主要用于测量胶卷对光线的敏感程度^[9]。为了规范不同厂家胶卷对感光度的描述，国际标准化组织 ISO 在 1974 年开始陆续制定了相关国际标准，包括 ISO 5800:2001、ISO 6:1993 和 ISO 2240:2003，分别用于规范各类胶卷的感光度描述。常见的感光度数值有 ISO 50、100、200、400、800、1600、3200 等。ISO 值越高说明胶卷对光线越敏感，反之，胶卷对光线越不敏感。ISO 值与感光度呈线性关系。比如，ISO 100 的胶卷对光线的敏感程度是 ISO 50 胶卷的两倍。当在较暗的光线下拍摄需要使用 ISO 值大的胶卷。图 4-24 显示了不同 ISO 值下的拍摄效果。



图 4-24 不同 ISO 拍摄效果^[10]

随着数码相机时代的到来，ISO 在 2006 年正式发布了标准 ISO 12232:2006，用于定义数码相机的 ISO 值。在数码相机中，ISO 值定义了照相机采集光线的能力。数码相机把落在感光器件上的光转化为电信号来处理。当电信号被放大时，ISO 值就会增加。放大 ISO 值一倍，意味着电信号放大一类，进入感光器件的光线只需要一半，以实现最佳的曝光。也就是说，如果 ISO 值从 ISO 100 增加到 ISO 200，光圈保持不变，则只需要一半的曝光时间。

在暗的场景中拍照，为了提高照片亮度，摄影师往往会把曝光时间延长，这会导致运动模糊，如图 4-25 所示。如果增加 ISO 值，就不需要延长曝光时间，从而减少运动模糊。这就是为什么当光线较差时人们需要增加 ISO 值的原因。在数码相机中，摄影师可以手动增加 ISO 值。

但是 ISO 值调大之后，噪声也会加大，图像的质量会下降，如图 4-26 所示。不同的感光器件，噪声不同，感光器件的质量越好，噪声就越小。那么，如何衡量感光器件的好坏呢？

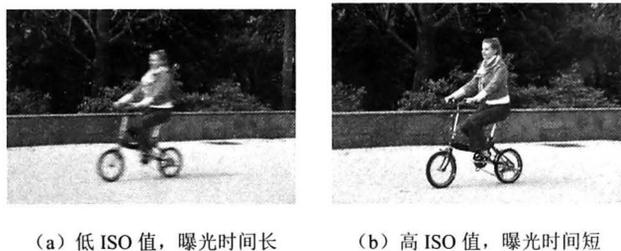


图 4-25 ISO 值与运动模糊^[11]

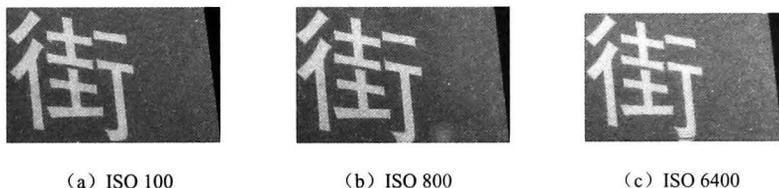


图 4-26 ISO 越大，噪声越大，图像越模糊^[10]

4.3.2 成像质量的衡量指标

衡量数字感光器件成像质量的指标有很多，其中第一个指标是量子效率 (Quantum Efficiency, QE)。所谓量子效率是指不同波长的光子撞击感光器件所产生的光电子与入射光子数量之比。QE 越高，代表感光性能越好。

CMOS 或者 CCD 感光器件的感光除了可见光谱之外，还可以覆盖部分近红外光。硅能够感知的光谱范围，CMOS 或 CCD 感知的光谱范围在 400~1000nm。一个典型的 CMOS 的量子效率如图 4-27 所示，它能够感知绿光的感知曲线如图 4-27 中的细实线和虚线所示，红光的感知曲线如图 4-27 中的粗实线所示，蓝光的感知曲线如图 4-27 中的点画线所示。不同的感光器件对不同光谱光的量子效率也不一样。

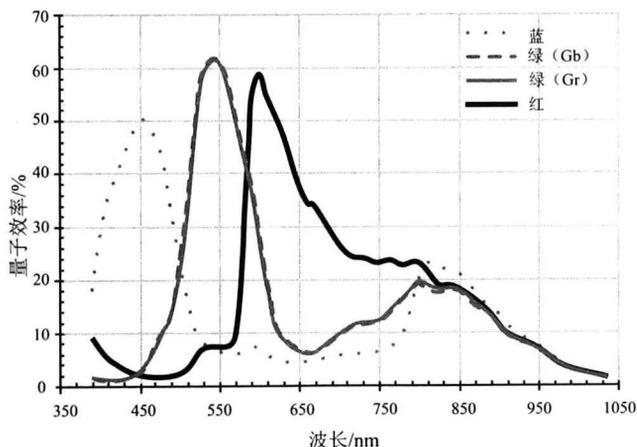


图 4-27 CMOS 量子效率图

第二个是采光方式，前面讲过，普通的 CMOS 或者 CCD 采光方式是前照式，也就是光从正面照射芯片形成电荷。光从正面经过透镜，经过滤色片再照到 CMOS 或者 CCD 的感光层上。以 CMOS 为例，从正面照射过来，CMOS 感光的部分只占到像元的 70%，

还有 30% 被晶体管占据了。所以它的感光面积减少。采用背照式，光从背面通过并直接进入感光层的光子效率可达 80%。

第三个是动态范围，所谓动态范围，有一个公式的定义：

$$DR = 20 \log_{10} (L_{\max} / L_{\min}) \quad (4.1)$$

画面中最亮的亮度除以最暗的亮度然后求对数，再乘以 20，就是它的动态范围。通常 CMOS 传感器动态范围是 40~60dB，CCD 为 60~70dB，人眼为 90dB，自然场景 >100dB。这也是衡量数字感光器件很重要的一个指标，后面有专门的章节来讲动态范围。

第四个是填充因子 (Fill Factor)，即实际感光面与像素面积的比值。理想情况下填充因子为 100%，然而实际值往往在 30%，这是因为感光面会被其他元器件所占据。虽然通过微透镜可以得到改善，但是微透镜的应用会影响到紫外光的检测。填充因子是影响灵敏度的一个很重要的因素，如图 4-28 所示。

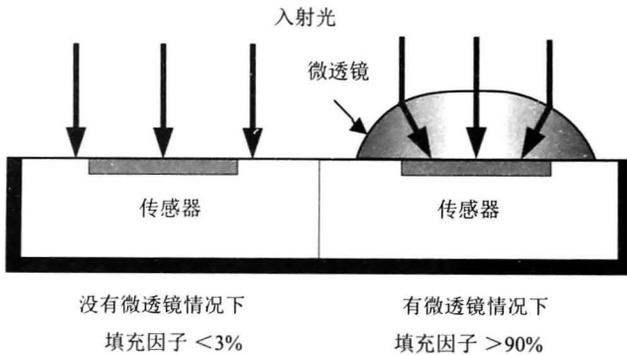


图 4-28 填充因子示意图

还有一个指标叫做井深 (Well Capacity)，是指每个像素可堆积电荷的多少。多数 CCD 可堆积 85 000 个电荷，高品质的 CCD 可堆积 350 000 个电荷。堆积的电荷越多，感光器的灵敏度越高，如图 4-29 所示。

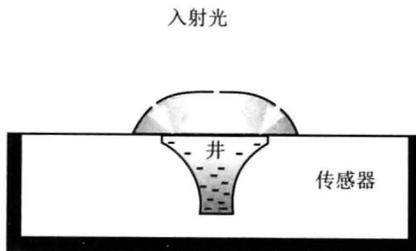


图 4-29 井深示意图

参考文献

- [1] SONY. CMOS image sensor[EB/OL]. (2008-06-11)[2014-05-25]. <http://www.sony.net/SonyInfo/News/Press/200806/08-069E/>.
- [2] Wikipedia. Film_speed[EB/OL]. (2013-06-14)[2014-05-25]. http://en.wikipedia.org/wiki/Film_speed.
- [3] University of Minnesota. VISUAL RESOURCES CENTER[EB/OL]. 2014-02-09[2014-05-25]. <http://www.arthist.umn.edu/vrcinfo/>.
- [4] 中金在线. 图片[EB/OL]. 2008[2014-05-25]. http://i6.cnfolimg.com/z/2_1151171.jpg.
- [5] Nobel Prizes and Laureates. The Nobel Prize in Physics 2009[EB/OL]. 2009[2014-05-25]. http://www.nobelprize.org/nobel_

prizes/physics/laureates/2009/smith-photo.html.

[6] 百度百科. CCD[EB/OL]. 2013[2014-05-25]. <http://wapbaike.baidu.com/view/18579.htm>.

[7] 百度百科. CMOS[EB/OL]. 2011[2014-05-25]. <http://baike.baidu.com/view/22318.htm>.

[8] 中关村在线. 富士[EB/OL]. 2008[2014-05-25]. <http://dcdv.zol.com.cn/115/1156658.html>.

[9] dvxuser. CMOS-CCD[EB/OL]. 2009[2014-05-25]. <http://dvxuser.com/jason/CMOS-CCD/>.

[10] 中关村在线. 从零开始玩单反——相机参数之 ISO/焦距篇 [EB/OL]. 2010[2014-05-25].http://dcdv.zol.com.cn/243/2430148_all.html.

[11] Nikon. Products & Support[EB/OL]. 2009[2014-05-25]. <http://chsvimg.nikon.com/history/basics/13/index.htm>.

第5章 自动聚焦技术

5.1 概 述

众所周知,人眼对事物的聚焦是本能的,几乎是不需要人脑有意识去控制的行为,其过程是瞬时而无法察觉的。但是,对于机器来说,准确而快速地聚焦不是一件容易的事情。早期的图像及视频采集系统中,多采取手动聚焦的方式得到清晰的图像。手动聚焦的主要缺点是速度慢,无法满足需要快速或实时聚焦的场合;另外,通过人眼主观判断的手动调节在一些情况下往往是难以准确聚焦的。自动聚焦(Auto Focus, AF, 也称为自动对焦)技术就是为了解决手动聚焦的问题而发展起来的。早在1963年,佳能公司就在原西德科隆博览会上展示出一架能自动聚焦的照相机样机;1974年,尼康公司同样也展示出一架自动聚焦原型相机。由于结构复杂、体积庞大的原因,这些照相机没能得到商业推广。直到1975年,美国霍尼韦尔(Honeywell)公司研制出一种具备实用价值的自动聚焦器件。在此自动聚焦器件的基础上,日本村尼卡公司在1977年成功推出世界上第一架广泛使用的能自动聚焦的35mm照相机。这之后,各种不同的自动聚焦技术先后出现并发展起来。近20年来,随着计算机以及数字信号处理等技术的发展,基于数字图像处理技术的自动聚焦技术得到越来越广泛的研究和应用,已成为研制具备先进功能的数字摄像系统及计算摄影系统。

简单地说,自动聚焦是能使目标场景在成像系统中准确清晰成像的某种自动调节的过程。自动聚焦方式主要可分为主动聚焦(Active AF)和被动聚焦(Passive AF)两类。主动聚焦依靠某种距离探测方式,如超声或红外测距等,以测量出目标场景与镜头之间的距离,然后据此调焦而获得对焦准确的位置。主动聚焦能适应各种光照情形,尤其是在低照度情况下也能正常工作。这类方法的缺点除了需要更多的成本及电源(或电池)等消耗外,还在于聚焦目标区域被局限于距离准确探测到的目标点,以及由于红外或超声的高反射性而无法透过玻璃准确对焦。

与主动聚焦不同的是,被动聚焦则不需要向对焦目标物发射任何能量或信息,仅仅通过利用透入的光线和形成的图像信息进行分析来调节聚焦。被动对焦主要包括相位检测聚焦(Phase Detection AF)和对比度检测聚焦(Contrast Detection AF)两类方法。相位检测聚焦是单反相机中用的最普遍的自动对焦方法,这种聚焦系统一般由反光镜、微透镜,以及多个成像传感器等硬件构成。其对焦原理如图5-1所示,简单地说,通过镜头的光束被分为两个部分,分别在不同的两组成像传感器上成像。如果对焦准确,则两幅图像相同;如果对焦不准,则两幅图像会出现偏移。通过比较两幅图像而检测这个偏移量,就能检测出聚焦的偏离状态。相位检测自动聚焦的优点在于对焦迅速且比较精确,缺点在于需要独特的硬件构造,造价高,且其复杂的构造不适合需要结构紧凑的情况,比如不适合在一般的数码相机及移动手机等平台上应用。

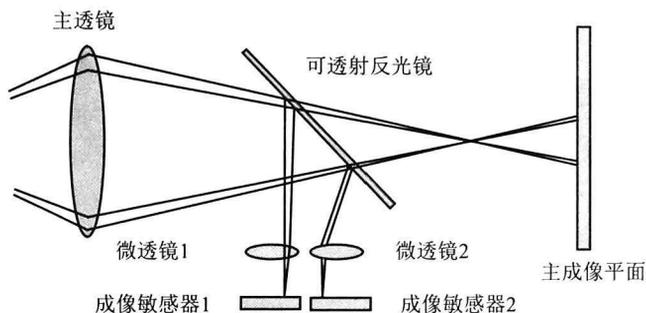


图 5-1 相位检测自动对焦的工作原理示意

对比度检测聚焦是另外一种被动聚焦方法，这种方法不需要额外的硬件构造，它仅通过对连续获得的实际图像进行分析处理，根据前后两帧或多帧图像的对比度计算结果来控制调焦的方向和大小。其基本原理是随着调焦向准确位置越靠近，图像越清晰，对应的图像对比度也越大。由于不需要额外的硬件支持，这种聚焦方法广泛应用于摄像机、数码相机等成像系统。但是，对比度检测聚焦方法往往存在以下缺点：①聚焦速度相对较慢；②对光照较为敏感，尤其是在低照度情况下容易失效；③当目标物缺乏明显的纹理细节（比如白色墙壁）时，难以准确聚焦。为了提高聚焦的速度、精度以及适应性，对比度检测自动聚焦方法受到越来越多的关注和研究，这是本章要重点介绍的内容。

本章结构如下，5.2 节首先将基于对比度检测自动聚焦的一些概念和基本框架进行介绍，然后将重点讲述聚焦值计算和聚焦峰值搜索方法。5.3 节将介绍模糊度计算方法，这是与自动聚焦紧密联系的一个问题。

5.2 对比度检测自动聚焦

图像对比度通常也称为锐度（Sharpness），是用来描述图像细节分辨程度和清晰度的度量值。对比度检测聚焦采用锐度来评价聚焦的准确程度，图像锐度越大也就是聚焦值越大，说明越接近准确聚焦的位置。在聚焦调节范围内，每一个调焦位置会对应一个锐度值，也就是聚焦值，这些聚焦值形成一曲线，通常称为聚焦曲线。锐度值通过锐度评价函数计算获得，锐度评价函数有时也称为聚焦函数。一个好的锐度评价函数对应的聚焦曲线应该是一个如图 5-2 所示的单峰曲线，并且曲线峰值对应的调焦位置即为准确聚焦的位置。

图 5-3 所示给出了对比度检测对焦的一般框架。当自动聚焦触发后，系统对当前的实际图像进行聚焦值计算，并结合前面的聚焦值数据综合判别当前聚焦值是否是聚焦曲线峰值。如果判定是聚焦峰值，则聚焦结束；否则，结合相关信息确定下一步调焦的移动方向和大小，驱动调焦步进电动机作相应的移动，对新的图像再次进行同样的步骤，以此形成一个反馈控制环路。由此可见，对比度检测聚焦的关键因素有两点：①锐度评价函数；②峰值判定和搜索策略。

对这两点，本章将在 5.2.1 节和 5.2.2 节分别阐述。在一些实际的成像系统中比如监控摄像机，往往需要对是否需要再次触发聚焦进行判定。比如，当镜头由聚焦清楚的近处场景转到远处场景时，图像可能会变模糊，此时是否需要再次触发自动聚焦是需要考虑的，这涉及图像模糊度评价问题。本章将在 5.3 节对图像模糊度评价方法作介绍。

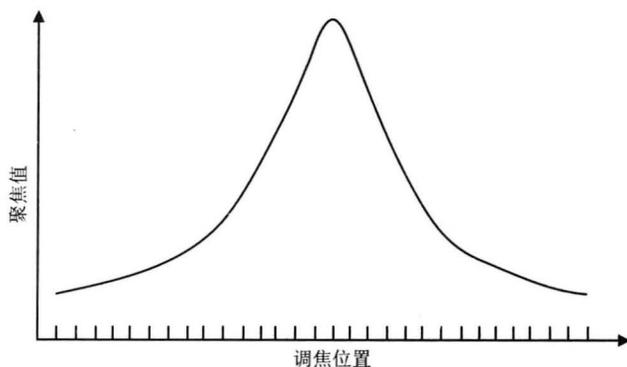


图 5-2 聚焦曲线

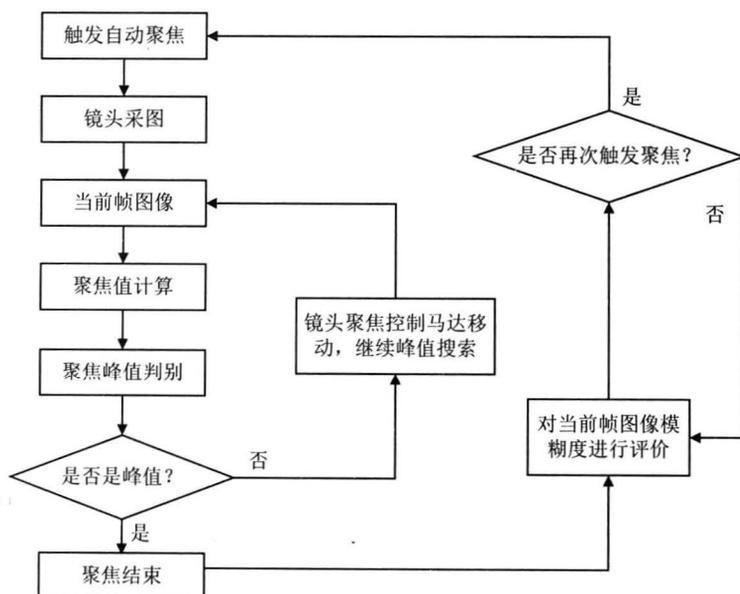


图 5-3 对比度检测自动聚焦系统的一般框架

5.2.1 聚焦值计算

1. 离焦与聚焦图像的差异

在介绍聚焦值计算方法之前,有必要弄清楚离焦与聚焦情况下图像有什么不同。一般的光学成像系统符合薄透镜成像规律,可以利用如图 5-4 所示薄透镜成像及聚焦的示意图对对焦调节的原理作简单说明。假定图示透镜焦距为 f ,物体离透镜距离为 u 时在成像平面上清晰成像,此时透镜光心与成像平面距离为 v ,则焦距 f 、物距 u 和像距 v 三者满足以下成像公式:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (5.1)$$

在对焦调节过程中,透镜焦距 f 和成像平面是保持不变的,可调节的是透镜在物体与成像平面的具体位置。假设物体在无穷远处清晰成像时透镜位于位置 b ,在最近距离

清晰成像时透镜位于位置 a ，则 ab 间距 F 即为聚焦的调节范围，一般成像系统常采用步进电动机控制调焦镜头在调节范围内移动。值得一提的是，实际的光学成像系统，如摄像机变焦镜头，往往是由多个透镜构成的，其聚焦调节过程则更为复杂。

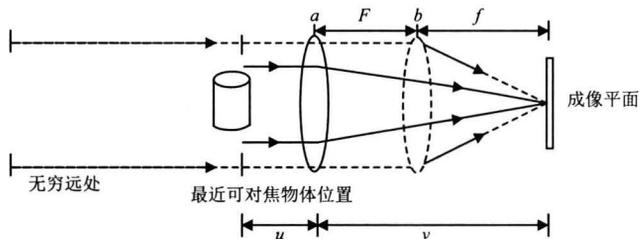


图 5-4 薄透镜成像及对焦示意

假设透镜焦距为 f (见图 5-5)，最大尺寸半径为 d 。实际图像敏感器成像平面位于点 D 处，能在成像平面上清晰成像的最佳物平面位于点 D' ，对应的距离分别为 u' 、 v' 。物点经透镜清晰成像于点，对应的物距和像距分别为 (u, v) 。由于物点 P 并不在最佳物平面的位置，其在成像平面上的实际成像不再是一个点，而是表现为半径为 R 的圆形散斑。由透镜成像公式 (5.1) 可得到：

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v'} \tag{5.2}$$

$$\frac{1}{f} = \frac{1}{u'} + \frac{1}{v} \tag{5.3}$$

则有

$$\Delta v = v - v' = \frac{f^2(u' - u)}{(u' - f)(u - f)} = \frac{-f^2 \Delta u}{(u' - f)(u - f)} \tag{5.4}$$

同时，由几何关系得到

$$\frac{R}{\Delta v} = \frac{d}{v} \tag{5.5}$$

从而，可以得到散斑半径

$$R = \frac{df|\Delta u|}{u|u^2 - f|} \tag{5.6}$$

可以观察到散斑半径 R 的大小反映了离焦的程度， R 越小则表示越接近准确聚焦。实际

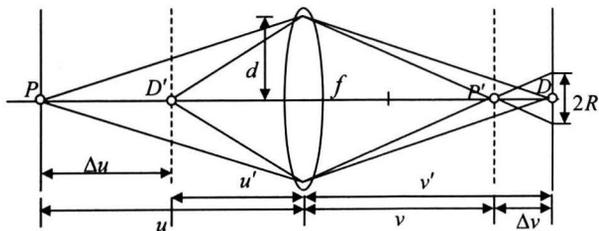


图 5-5 离焦示意

上, 模糊散斑就是成像系统对点光源的响应, 此响应用点扩散函数 (Point Spread Function, PSF) $h(x, y)$ 来表示。假设 P 为点光源, 在曝光时间内发出一个单位的光能量, 且在整个成像过程没有能量损失, 则这些光能量将全部分布在散斑之内, 可以得到

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) dx dy = 1 \quad (5.7)$$

在理想情况下, 圆形散斑内的亮度认为是均匀分布的, 而散斑外的亮度为零, 因此有

$$h(x, y) = \begin{cases} \frac{1}{\pi R^2} & x^2 + y^2 \leq R^2 \\ 0 & \text{其他} \end{cases} \quad (5.8)$$

对于实际的成像系统, 点扩散函数经常表示为

$$h(x, y) = \frac{1}{\sqrt{\pi R^2}} e^{-\frac{x^2+y^2}{R^2}} \quad (5.9)$$

由此可以看出, 点扩散函数相当于一个低通滤波器, 实际图像 $g(x, y)$ 可以看成是由理想聚焦成像 $f(x, y)$ 与成像系统的点扩散函数 $h(x, y)$ 卷积而形成的, 也就是

$$g(x, y) = f(x, y) * h(x, y) \quad (5.10)$$

聚焦越准确, 低通滤波的影响越小, 使得实际图像中的高频成分越少受到抑制, 在空域中表现为图像越锐利、局部特征 (比如边缘、灰度等) 变化越大; 离焦量越大, 低通滤波的影响越大, 使得图像中的高频成分抑制得越厉害, 甚至缺失, 图像就越模糊。因此, 聚焦程度即聚焦值的大小能通过图像的高频成分的多少来直接反映, 这是几乎所有有计算聚焦值的锐度评价函数的基本原理。

2. 常用锐度评价函数

由上述可知, 聚焦程度比较好时, 图像细节丰富, 反映在空域上表现为局部特征, 如边缘、梯度等变化越大, 反映在频域上则表现为高频成分的多少, 因此, 锐度评价函数方法主要可分为两类: 空间域的方法和频率域的方法。

3. 空间域锐度评价函数

在空间域上有许多方法来描述图像的细节丰富及特征变化程度, 从不同角度可以得到不同的锐度评价函数。下面选择介绍以下几种典型的空域方法^[1, 2]。

1) Sobel 梯度函数

Sobel 梯度函数也称 Tenengrad 函数, 是一种简单且最常用的锐度评价函数。此函数的计算思想是, 聚焦越准确, 图像的边缘越明显, 梯度也就越大。假定计算锐度的图像块大小为 $M \times N$, $g(i, j)$ 表示像素 (i, j) 的灰度值, 其公式为

$$F_{\text{sob}} = \sum_M \sum_N (G_x^2(i, j) + G_y^2(i, j)) \quad (5.11)$$

式中, $G_x(i, j)$ 和 $G_y(i, j)$ 分别为像素 $g(i, j)$ 对应的与 Sobel 算子卷积得到的水平和垂直方向梯度分量。

2) 基于阈值的绝对梯度函数

此函数计算思想与 Sobel 梯度函数类似, 其公式为

$$F_{\text{thg}} = \sum_M \sum_N |g(i, j+1) - g(i, j)| \quad \text{当 } |g(i, j+1) - g(i, j)| \geq t \quad (5.12)$$

式中, t 为预先定义的阈值。有时也将相邻像素差的平方作为计算依据, 即用 $(g(i, j+1) - g(i, j))^2$ 代替 $|g(i, j+1) - g(i, j)|$ 。

3) 方差函数

方差函数的计算思想是, 聚焦越准确, 图像就越清晰, 对应的像素灰度变化就越大; 反之, 图像越模糊, 局部范围内像素灰度变化就越小。其公式为

$$F_{\text{var}} = \frac{1}{MN} \sum_M \sum_N (g(i, j) - \bar{g})^2 \quad (5.13)$$

式中, $\bar{g} = \frac{1}{MN} \sum_M \sum_N g(i, j)$ 为图像块灰度均值。

4) 像素差和函数

像素差和函数的计算思想与方差函数类似, 相邻的像素灰度变化随着图像变清晰而增大。其公式为

$$F_{\text{smd}} = \sum_M \sum_N (|g(i, j+1) - g(i, j) + g(i+1, j) - g(i, j)|) \quad (5.14)$$

5) Laplacian 和函数

Laplacian 和函数的计算思想与 Sobel 梯度函数的一致, 实际上 Laplacian 算子与 Sobel 算子都是常用的边缘检测算子。考虑到 Laplacian 算子在某些情况下水平和垂直分量符号相反会造成数值相抵, 因此采用各分量的绝对值之和作为聚焦值计算。其公式为

$$F_{\text{smi}} = \sum_M \sum_N (|2g(i, j) - g(i, j+1) - g(i, j-1) + |2g(i, j) - g(i+1, j) - g(i-1, j)|) \quad (5.15)$$

6) Laplacian 能量函数

Laplacian 能量函数也是基于 Laplacian 算子, 其公式为

$$F_{\text{co1}} = \sum_M \sum_N \sum_M \sum_N (4g(i, j) - g(i, j+1) - g(i, j-1) - g(i+1, j) - g(i-1, j))^2 \quad (5.16)$$

7) 灰度相关函数

灰度相关函数的计算思想是, 像素间的相关程度随着图像变清晰而增大。存在两种近似的公式:

$$F_{\text{cor1}} = \sum_M \sum_N g(i, j)g(i+1, j) - \sum_M \sum_N g(i, j)g(i+2, j) \quad (5.17)$$

$$F_{\text{cor2}} = \sum_M \sum_N g(i, j)g(i+1, j) - MN \bar{g}^2 \quad (5.18)$$

4. 率域锐度评价函数

频域方法首先需要将图像数据转换到某一种频率域中, 然后提取出高频分量的信息来表达锐度或聚焦值。显然, 相比于空间域的方法, 频域方法首先需要将图像数据转换到频率域, 一般会需要更大的计算量。常用的频域变换方法有离散傅里叶变换 (Discrete Fourier Transform, DFT)、离散余弦变换 (Discrete Cosine Transform, DCT)、小波变换 (Wavelet Transform, WT) 等, 这里主要关注基于 DCT 的频域锐度评价函数。实际上, 目前应用于自动聚焦的频域锐度评价函数很多是基于 DCT 的, 这是因为 DCT 较 DFT、WT 等具有以下优点。

(1) 计算复杂度低, DCT 比快速傅里叶变换的计算量还小, 并且存在快速的 DCT 计算方法, 适合硬件实现。

(2) DCT 变换将绝大部分图像信息集中在少数低频的 DCT 系数上, 使得仅仅需要计算少量的系数就能很好地表达整幅图像, 这在相同的计算量前提下, 构建更为灵敏、更具分辨力的锐度评价函数提供了基础。

(3) 许多图像及视频压缩数据是基于块状 DCT 变换获得的 (比如 JPEG、MPEG2、H263, 以及 H264 等都是基于 DCT 的压缩算法), 因此设计基于块状 DCT 的锐度评价函数则可以直接采用已经获得的 DCT 块数据, 这可以较大地提高计算效率。

在介绍基于 DCT 的锐度评价函数之前, 先介绍 DCT 的计算方法。DCT 是一种空间域向频率域一对一的映射, 它将空间域的图像数据向一组 2D 正交基上投影而得到全为实数的系数。设图像大小为 $M \times N$, $g(x, y)$ 为在 (x, y) 的像素值, 则 DCT 可定义为

$$\mathcal{F}_c [g(x, y)] = \{D_{\omega\nu}(g(x, y))\} \quad (5.19)$$

系数 $D_{\omega\nu}(g(x, y))$ 对应像素 $g(x, y)$, 具体定义如下:

$$D_{\omega\nu}(g(x, y)) = \sum_{(x, y)} g(x, y) C_{\omega\nu}(x, y; M, N) \quad (5.20)$$

$C_{\omega\nu}$ 是正交基函数:

$$C_{\omega\nu}(x, y; M, N) = C_{\omega}(x; M) C_{\nu}(y; N) \quad (5.21)$$

其中

$$c_{\varphi}(z; A) = \alpha(\varphi; A) \cos\left(\frac{\pi\varphi(2z+1)}{2A}\right) \quad (5.22)$$

$$\alpha(\varphi; A) = \begin{cases} 1 & \varphi = 0 \\ \sqrt{A} & \text{其他} \\ \sqrt{\frac{2}{A}} & \text{其他} \end{cases} \quad (5.23)$$

一种普遍采用的方法是进行块状的 DCT 计算, 即首先将 $M \times N$ 大小图像分成 $n \times n$ 大小的块状, 然后分别对每一小块进行 DCT 变换。常见的 n 取值 8、4 为等。以 $n=8$ 为例, 8×8 -DCT 变换将每一 8×8 的图像块变换为 $\{D_{\omega\nu} | \omega = 0, 1, \dots, 7; \nu = 0, 1, \dots, 7\}$, 即 64 个系数, 其中 D_{00} 称为 DC 系数, 其他 63 个为 AC 系数。图 5-6 所示为这些系数的表示, 越靠近左上角低频分量就越多, 越靠近右下角高频分量越多。图 5-7 所示为对 lena 图进行 8×8 -DCT 变换的结果。

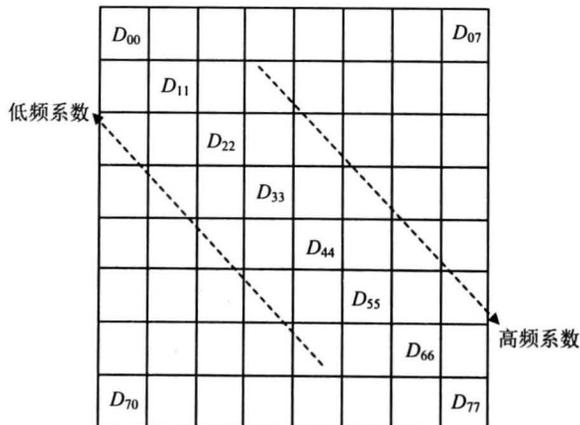


图 5-6 8×8 块状 DCT 系数

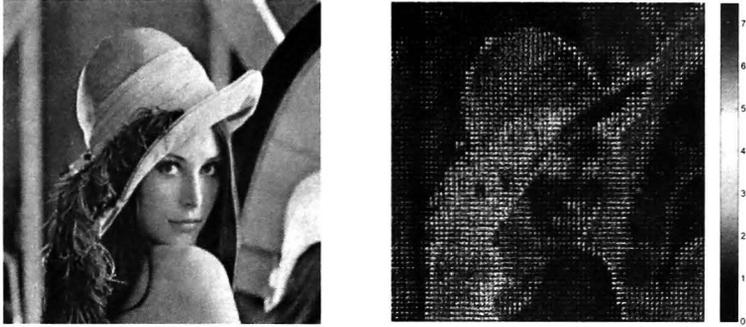


图 5-7 8×8 块状 DCT 变换

下面介绍几种基于 DCT 变换的频域锐度评价函数。

1) 贝叶斯熵 DCT 函数^[3]

贝叶斯熵 DCT 函数的计算思想是，归一化的 DCT 系数相当于一个概率分布，当聚焦越准确时，此分布越接近均匀分布。对每一 8×8 块，聚焦值计算公式为

$$F_{\text{bedct}} = 1 - \frac{\sum_{\omega+v \leq t} |D_{\omega v}|^2}{(\sum_{\omega+v \leq t} |D_{\omega v}|)^2} \quad (5.24)$$

式中，阈值 t 常设为 6。整体的聚焦值采用所有单个块聚焦值的平均值或所有值之和。

2) 基于中频系数的 DCT 函数^[4]

基于中频系数的 DCT 函数的计算思想是，经统计比较 DCT 某些系数平方值随着聚焦程度成正比例变化。此函数采用 $|D_{33}|^2$ 作为单个 DCT 块的聚焦计算值。考虑噪声的影响及计算方便，直接用图像与以下算子（记为 O_{mf} ）卷积得到 $|D_{33}|^2$ 的近似值：

1	1	-1	-1
1	1	-1	-1
-1	-1	1	1
-1	-1	1	1

即有

$$F_{\text{mfct}} = \sum_x \sum_y |g(x, y) * O_{\text{mf}}|^2 \quad (5.25)$$

3) 低高频比值 DCT 函数^[5]

低高频比值 DCT 函数的计算思想是，图像越清晰，细节越丰富，高频分量越多，与低频分量的比值就会越大。单位块的聚焦值计算公式为

$$F_{\text{rdct}} = \frac{\sum_{(\omega+v) \neq (0,0)} |D_{\omega v}|^2}{|D_{00}|^2} \quad (5.26)$$

4) 基于中频系数方向性选择的 DCT 函数^[6]

基于中频系数方向性选择的 DCT 函数的计算思想与第二种 DCT 函数类似，除了考虑某些 AC 系数随着聚焦程度有更明显的变化外，它还考虑某些系数根据图像内容不同可能呈现出方向性变化的特点。首先比较以下三个值：

$$\begin{cases} E(V) = |D_{01}| + |D_{02}| \\ E(H) = |D_{10}| + |D_{20}| \\ E(D) = |D_{11}| + |D_{12}| \end{cases} \quad (5.27)$$

然后根据其中的最大值来计算聚焦值，其公式为

$$F_{\text{direct}} = \begin{cases} D_{02}^2 + D_{03}^2 & \text{当 } E(V) \text{ 最大} \\ D_{20}^2 + D_{30}^2 & \text{当 } E(H) \text{ 最大} \\ D_{22}^2 + D_{33}^2 & \text{当 } E(D) \text{ 最大} \end{cases} \quad (5.28)$$

5) 基于系数重组的 DCT 函数

基于系数重组的 DCT 函数是一种基于 DCT 系数重组的锐度评价函数，其计算思想类似于 F_{mfct} 和 F_{direct} ，即从 AC 系数中提取某些随聚焦程度变化呈现更敏感、更大变化的系数，作为聚焦值计算的参与依据。首先按图 5-8 将每个 8×8 的 DCT 块分成 10 个子区域：如果将所有 8×8 块相同的 DCT 子区域组合到一起并按相应的位置放置，可以发现经此重组后的图像表示与经小波变换后的图像非常相似，如图 5-9 所示。这种 DCT 经重组

0	1	4	7
2	3		
5		6	
8			9

图 5-8 8×8 块状 DCT 系数分区

后体现出小波变换的相似性说明图像的高频分量主要集中在子区域 7、8、9 的 AC 系数上。由 DCT 特性知，边缘等高频信息主要反应在少数 AC 系数上，并且可以发现，其反应的高频信息主要集中在 7、8、9 区域的左上角系数位置。根据高频信息越丰富，图像会越清晰的规律，采用如下计算公式计算单位块的聚焦值：



(a) 无模糊的原图即高斯模糊 $\delta=0$

(b) 高斯模糊 $\delta=1.3$

图 5-9 两幅清晰度不同的图像对应的重组后的 DCT 图 (原图来源为公共图像数据库 LIVE)

$$F_{\text{rogdct}} = (D_{42}^2 + D_{41}^2 + D_{50}^2 + D_{51}^2) + (D_{44}^2 + D_{45}^2 + D_{54}^2 + D_{54}^2) + (D_{04}^2 + D_{05}^2 + D_{14}^2 + D_{15}^2) \quad (5.29)$$

取所有单个块聚焦值之和作为整体的聚焦值计算结果。

5. 锐度评价函数的比较和分析

锐度评价函数决定了聚焦值和聚焦值曲线，由于最大聚焦值对应的就是准确聚焦的位置，可以说锐度评价函数决定了聚焦过程的精度。一个好的锐度评价函数应该具备以下特性。

(1) 单峰性，即只存在一个最大峰值。

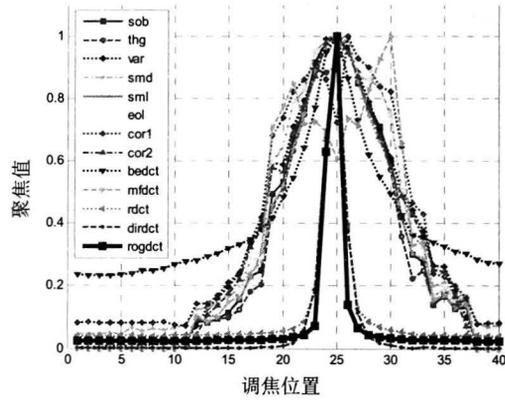
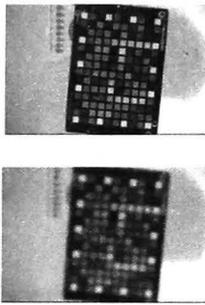
(2) 精确性，即最大峰值对应的为聚焦最精准的位置。

(3) 陡峭性，即从最大峰值往两侧都是单调递减的并且在峰值较小的邻域内函数值变换较大。

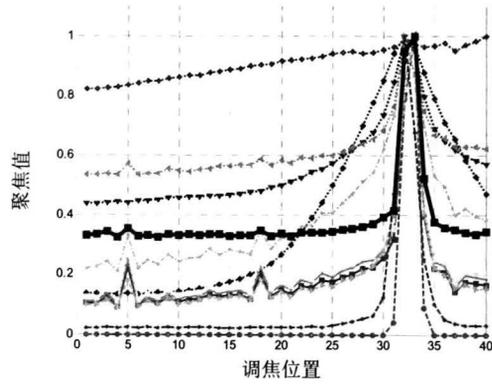
(4) 可实现性，即计算量较小，可以容易地在硬件上实现，并能满足实时性的要求。

为了比较和分析各锐度评价函数，用采集四组实际场景聚焦图像序列作为实验数据。每组图像序列由覆盖整个调焦范围、对应不同调焦位置的 40 幅分辨率为 1920×1080 像素的图像构成，各组聚焦图像序列对应的准确聚焦位置由观察比较得到。对每幅图像降采样为 960×540 ，取图像中间 196×128 块状区域作为聚焦值计算的图像窗口。参与比较的包括本节介绍的 13 种锐度评价函数，其中两种相关函数都包括在内；基于阈值的绝对梯度函数采取相邻像素差的平方作为计算依据。各聚焦曲线绘制采用归一化后的聚焦值数据。为了叙述方便，采用各函数公式对应的英文下标作为函数的表示。

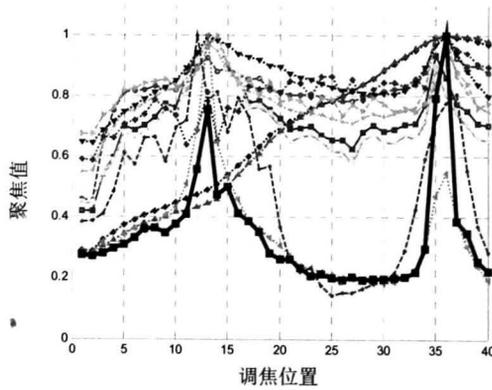
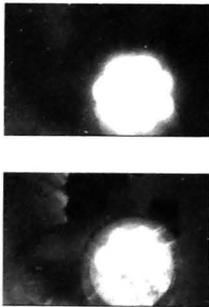
第一组图像如图 5-10 (a) 所示。聚焦目标物为纹理突出的彩色定标板，从图 5-10 绘出的各锐度评价函数的聚焦曲线可以看出，除了 `mfdct` 函数外各聚焦函数基本上都具备单峰性质。从精确性来看，有 6 种函数的最大峰值位置为准确聚焦的位置，也有 6 种空间域函数稍有偏差。4 个频率域函数包括 `bedct`、`rdct`、`dirdct` 以及 `rogdct` 具有更突出的陡峭性。第二组图像场景为阴天情况下普通的室外房屋，由图 5-10 (b) 聚焦曲线可见，除了 `cor1` 函数外各函数都表现出较好的单峰性。仔细比较可以发现，`rogdct` 最大峰值位置对应于准确聚焦位置，而其他 DCT 频域函数则有一个位置的偏差。第三组由直接把镜头对着主动光源（手电）而获得，由于光晕的影响以及背景相对较暗，使得光晕的边缘较强，对应于锐度值往往较高。由图 5-10 (c) 聚焦曲线可见，除了 `var`、`cor2` 外其他所有锐度评价函数都有两个峰值，其中各空域函数或对光晕输出更大的响应，或两峰值几乎无大小差异；而频域函数中 `bedct`、`mfdct` 也是输出两个无太大差别的峰值，`rdct`、`dirdct` 则在光晕阶段输出最大响应，只有 `rogdct` 输出对应于光晕的局部峰值和准确聚焦位置的最大峰值。有意思的是 `var` 和 `cor2` 两个函数能够在近乎准确聚焦位置上输出最大响应，这是由于两者锐度值计算都是基于像素灰度与整体灰度均值之间的差异，而不依赖边缘或梯度强度，所以光晕对它们影响很小。



(a) 第一组, 准确聚焦位置为 25

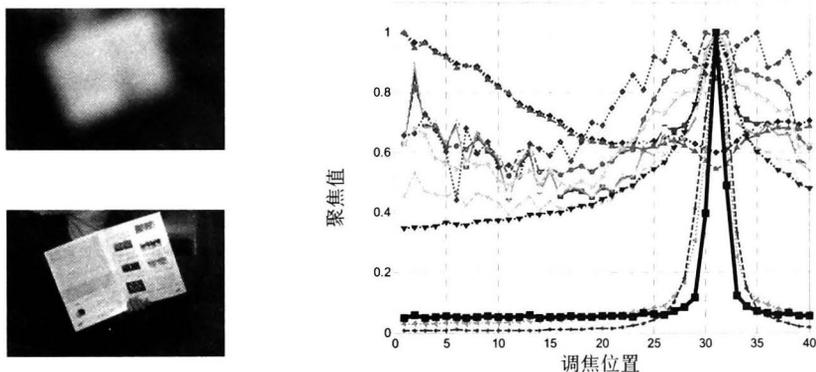


(b) 第二组, 准确聚焦位置为 33



(c) 第三组, 准确聚焦位置为 36

图 5-10 聚焦图像序列样图及各函数的聚焦曲线



(d) 第四组, 准确聚焦位置为 31

图 5-10 聚焦图像序列样图及各函数的聚焦曲线 (续)

第四组图像是在低照度情况下采集的, 相对于正常光照下, 低照度情况下采集的图像受噪声影响严重。由图 5-10 (d) 的聚焦曲线可以看出, 空间域的锐度评价函数普遍对噪声更为敏感, 所形成的聚焦曲线包含很多局部小波动, 甚至输出错误的最大峰值; 相反, 频率域函数整体上体现出较强的抗噪声能力。频率域函数的聚焦值计算都是基于高频分量的多少, 而噪声对高频分量影响较少, 因此频率域锐度评价函数自然体现出更强的抗噪声能力。

通过这四组实验可以发现, 一个锐度评价函数能在任何情况下都保持单峰性、精确性以及陡峭性不是一件容易的事情。实际上, 由于聚焦时可能会遇到各种不同的场景、光照条件、噪声干扰等, 一般的锐度评价函数很容易呈现出多个局部的峰值, 甚至错误的最大峰值等现象。相比较而言, 频域函数虽然计算量往往更大, 但整体上具备更强的抗噪声能力, 而空域函数尤其是方差、像素相关等函数对噪声更为敏感。空域函数中的 Sobel 和基于阈值的绝对梯度函数具备较为稳定的性能, 而 DCT 函数中基于中频系数方向性选择函数 (dirdct), 低高频比值函数 (rdct) 以及基于 DCT 系数重组函数 (rogdct) 体现出更健壮、更准确的性能。

5.2.2 聚焦峰值搜索

聚焦的准确位置对应于锐度评价函数的最大峰值, 也只有找到最大峰值才能知道准确聚焦的位置。如何确定最大峰值, 也就是采用什么样的方法来搜索聚焦峰值, 是基于对比度被动自动聚焦的另外一个需要考虑的问题。从某种意义上讲, 锐度评价函数决定了聚焦的精度, 而聚焦峰值搜索方法决定了自动聚焦的速度。下面介绍四种搜索方法。

第一种聚焦搜索方法为全行程搜索法^[2, 7], 也称全局搜索。最大峰值似乎总是需要将所有调焦位置的聚焦值计算出来后才能确定, 因此采用穷尽每一个调焦位置的方法或全行程搜索方法是自然的也是最直观的办法。如图 5-11 (a) 所示, 它从调焦范围的一端出发, 采用等长的步长 (一般为最小步长) 一步一步移动到调焦末端, 从而确定出一

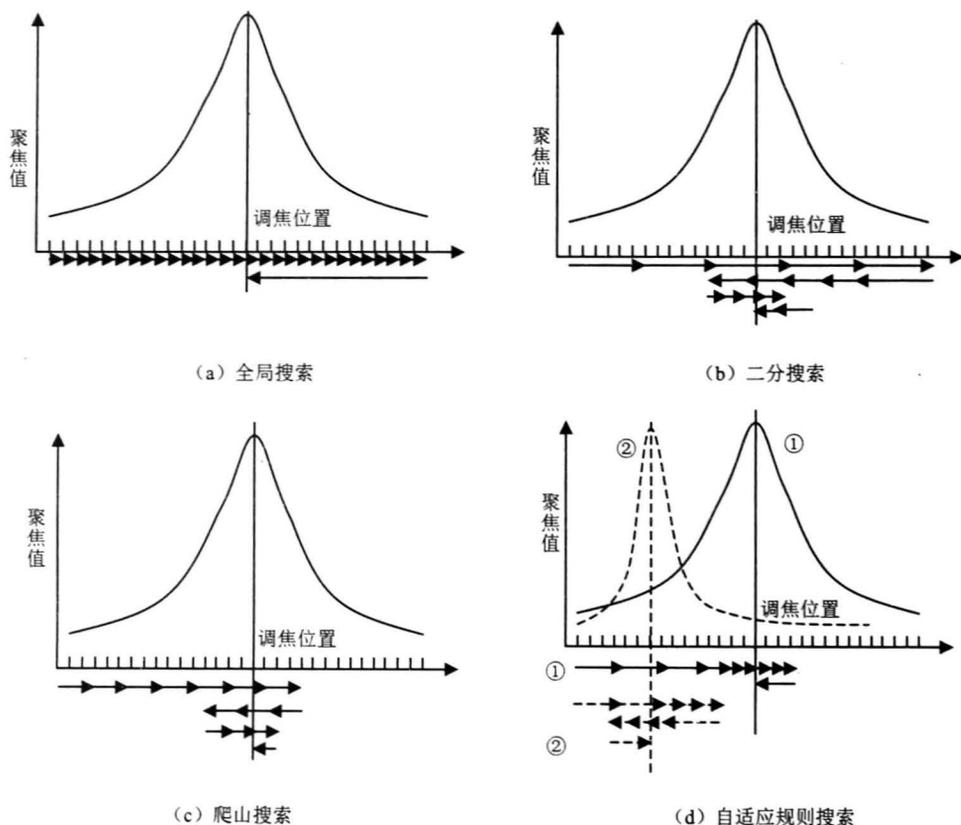


图 5-11 聚焦搜索示意图

个最大聚焦值位置。显然，这种全局搜索不是一个高效率的方法。实际上，对某些长焦镜头其调焦范围存在成千甚至上万个调焦位置，穷尽每一个位置会相当耗时，达不到快速聚焦的要求；并且，实际的成像系统聚焦往往要求在遇到聚焦最清晰位置之后，就能较迅速地判断而结束继续调焦，全局搜索方法显然达不到这个目的。此方法虽然效率不高，但是只要聚焦函数足够精确，就总是能够找到聚焦最准确的位置。这种简单而不失精度的方法在许多低端的数码相机产品中经常用到。

第二种聚焦搜索方法为二分搜索法^[7]，如图 5-11 (b) 所示，它是一种需要往返多次的搜索策略。与全局搜索类似，二分法从一端出发，首先以较大步长对整个调焦范围进行扫描一次，然后移动到当前最大聚焦值的周围邻域，再以较小步长进行搜索，如此往复，直到搜索的最大值变化足够小为止。与全局搜索相比，这种搜索策略需要较少的步骤，但是它存在的缺点也是明显的，一是来回反复调焦容易造成图像“振荡”现象，其视觉感官上的效果较差；二是对某些调焦装置来说，来回反复调焦在硬件上的消耗代价比较大，比如有的调焦模块往后调焦移动一步需要步进电动机移动多步，这有可能比全局搜索需要更多的步骤和消耗。

第三种聚焦搜索方法为爬山搜索^[8~10]，如图 5-11 (c) 所示，它的基本原理是在调焦过程通过前后的聚焦值信息来判定坡峰的位置。简单地说，爬山法首先以合适的固定步长出发，当遇到聚焦值显著变化的位置时，判定并记下处于爬坡状态；当聚焦值开始

连续下降时,判定已过坡峰位置,则调头再以较小步长从反方向进行爬坡一次,通过对坡峰如此反复地来回确定,直到最大聚焦值变化足够小为止。爬山搜索的优点在于不需要对整个调焦范围进行搜索,能以较少的调焦步骤迅速地确定准确聚焦的位置。缺点在于需要确定合适的阈值参数和每次爬坡的步长大小;容易陷入局部的坡峰位置而聚焦错误。存在一些基于爬山搜索的改进或类似方法,在此不再赘述。

第四种聚焦搜索方法是在规则搜索法^[7]基础上改进的一种方法,称之为自适应规则搜索,如图 5-11 (d) 所示。规则搜索^[8]的基本原理是,通过前后两帧图像的聚焦值来确定当前的搜索区域和步长大小,即将搜索区域划分为粗糙搜索区域、中间区域以及精细搜索区域。在粗糙搜索区域,每步长以 7~10 倍的最小步长搜索;中间区域则每步长为最小步长的 3~4 倍;精细搜索区域每步长即为最小步长。由于聚焦函数可能存在多个局部峰值,规则搜索没有对峰值作出明确的判定和取舍,一般也需要对整个调焦区间进行一次扫描来确定最大峰值的位置。为了避免对整个调焦区间进行搜索,增强聚焦逃离局部峰值的能力,提高聚焦的稳定性和准确性,基于规则搜索和爬山搜索方法,在自适应规则搜索方法中采取了以下的策略。

(1) 对全局坡峰的确定分两种情况处理:第一种如图 5-11 (d) 中①所示,即能够正常地判定爬坡和下坡两个状态,此时可以迅速确定坡峰位置;第二种如②所示,即由于某一步直接跨过坡峰位置,使得缺少爬坡状态,这时在下坡之后再返回到当前最大聚焦值位置,以小步长进行反向爬山搜索。这种对坡峰的搜索处理,可以避免爬山搜索法来回多次地重复搜索,同时以最小步长对坡峰区域的搜索也能保证聚焦精度。

(2) 由于对搜索区域的判定是基于前后两帧图像聚焦值之差,比如当来确定是否进入精细搜索区域,这使得阈值设定尤为重要。然而,由于实际聚焦场景千差万别,使得在某些情况下可能丢失对全局坡峰的判定,而造成对整个调焦区间覆盖搜索甚至多次往返搜索。对此,增加了对次坡峰的判定,即设定较小阈值使得当时标定为已搜索到一个次坡峰。当后续搜索没有发现更为显著的坡峰,一旦进入粗糙搜索区域时,则判定次坡峰为全局坡峰位置,这样可以避免对调焦区间覆盖搜索或多次往返聚焦。

(3) 由于实际聚焦场景存在各种干扰,比如夜间光源形成的光晕,使得仅仅依靠前后两帧图像聚焦值之差有时候无法准确判定搜索区域类型。因此,增加了对图像聚焦值和模糊度的要求,比如只有当,且同时当图像聚焦值超过规定的先验阈值以及模糊度低于规定的阈值时,才能判定当前已进入精细搜索区域。在 5.3 节,将对图像模糊度计算作介绍。

(4) 在中间区域和精细区域间增加一个“较好区域”并采用多种方法使得搜索能够逃离局部峰值区域。比如当判定已经进入精细搜索区域后,遇到聚焦值多次上下振荡或当连续爬坡的步骤过多时,则根据图像聚焦值和模糊度改变搜索区域,以用较大步长逃离此区域。

(5) 提高对下坡状态的判断要求,即只有当某一下坡步骤下使得聚焦值相对于前帧聚焦值超过一定比例的下降时,才确定是下坡状态;在下坡结束时只有当前聚焦值相对于最大聚焦值有超过一定比例的下降时,才可以进行结束聚焦或返回搜索等动作。添加下降度量值要求可以避免坡峰区域内的局部干扰,增加对全局坡峰的搜索准确度。

值得一提的是,由于实际目标场景多种多样,光照等外界干扰因素也很多,聚焦搜

索应尽可能在满足准确度的前提下提高速度。但是,在任何情况下都应该避免持续不断地往返调焦,而应该在最大限度的往返搜索次数内使得聚焦结束在最大聚焦值位置。对聚焦过程影响较大的一个重要因素是聚焦值计算时图像区域的选择,即选择哪块区域的图像作为聚焦值计算窗口。考虑计算量的约束,窗口的大小也是一个需要考虑的因素。一般的方法是选取图像中间合适大小的长方形区域。然而,中间区域有时候并不是聚焦的目标区域。有的自动聚焦方法选择有偏好的目标区域作为聚焦窗口,比如有以人脸作为最优先选择的区域^[11]。如何自适应地选择聚焦窗口也是一个值得研究的问题,在此对这方面不做更多的阐述。

5.3 图像模糊度评估

图像模糊度评估是图像质量评估的一种,可分为主观评估和客观评估。主观评估是基于人眼观察比较的主观打分,通过采集一定数目的打分样本统计得到。在很多应用上,人都是作为图像或视频输出的终极用户,其对图像的主观感受和评价被认为是最可靠的。主观评估的缺点显而易见,即耗时、耗力,无法应用到需要实时评估的场合。客观评估是通过对图像数据分析计算达到与人眼观测相关或近似的评估指标,根据计算条件可分为全参照(Full-Reference)、部分参照(Reduced-Reference)和无参照(No-Reference)三类方法。全参照评估需要原始无失真图像作为参照,其计算是基于原始图像与失真图像之间的误差,结果稳定且比较准确;部分参照方法利用原始图像的部分信息,比如常通过附加通道或水印等方式传输的原始图像特征数据,其计算基于原始图像特征与失真图像特征差异,减少了数据传输,同时也具有较高的可靠性;无参照方法不需要任何原始图像信息,这类方法灵活,具有最广泛的应用,是近十年里图像处理领域热门的研究问题之一。图像模糊是图像降质的一种,在这里主要关注图像模糊的客观无参照评估方法。

5.3.1 模糊度与锐度的关系

一般来说,图像锐度与图像模糊度是成反比例的,图像锐度越大,即边缘越尖锐、细节越清楚,则图像越容易分辨、越清晰,也就是越不模糊。所以在很多情况下,两者似乎是可以互相代替的。但是,严格来说两者之间还是存在差别。图像不模糊不一定就锐度大,反之亦然。比如准确聚焦的光滑纯黑色区域是不锐利的,但不能认为它是模糊的。

在前面介绍的锐度评价函数都是针对在自动聚焦应用中一序列图像场景内容相同,但离焦模糊度不同的图像数据,所获得的锐度值跟图像内容是很相关的,其大小在相对于整个聚焦图像序列才具有具体的意义。实际上,对于图像内容不同的图像,这些锐度评价函数所计算的锐度值与实际的模糊度往往是匹配不上的。例如,图 5-12 给出了公共图像数据库 LIVE^[12]中四幅相同分辨率图像,每幅图都有一定程度的高斯模糊,对应于不同大小的高斯模糊标准差 δ 。分别采用前面介绍的锐度评价函数对四幅图像作锐度值计算,均采用整幅图像作为计算窗口,得到的计算结果如图 5-13 所示,其中横轴对应图 5-12 (a)~图 5-12 (d) 所示的图像,纵轴给出的是模糊值,即归一化后锐度值的倒数。可以看出,所有锐度评价函数都不能够完全准确地分辨出内容不同的图像之间的绝对模糊程度,这是由于不同内容图像所呈现的局部图像特征(如亮度、边缘、频谱等)

不同，进而使得高频分量的绝对含量不同。比如，图 5-12 (a) 与图 5-12 (b) 相比具有更多纹理较少的平滑区域，而图 5-12 (b) 存在许多尖锐的边缘，除了 `rogdct` 与 `rdct` 函数能较明显地分辨出两者之间的模糊程度外，其他的锐度评价函数几乎都错误地将图 5-12 (a) 输出更大模糊响应。显然人眼主观上很容易分辨这两者之间的模糊程度的。同时也可以发现，`rogdct` 函数虽然能够分辨出图 5-12 (a) 与图 5-12 (c)、图 5-12 (d) 之间的模糊度差异，但也无法准确分辨出与图 5-12 (b) 间的细微模糊差别。

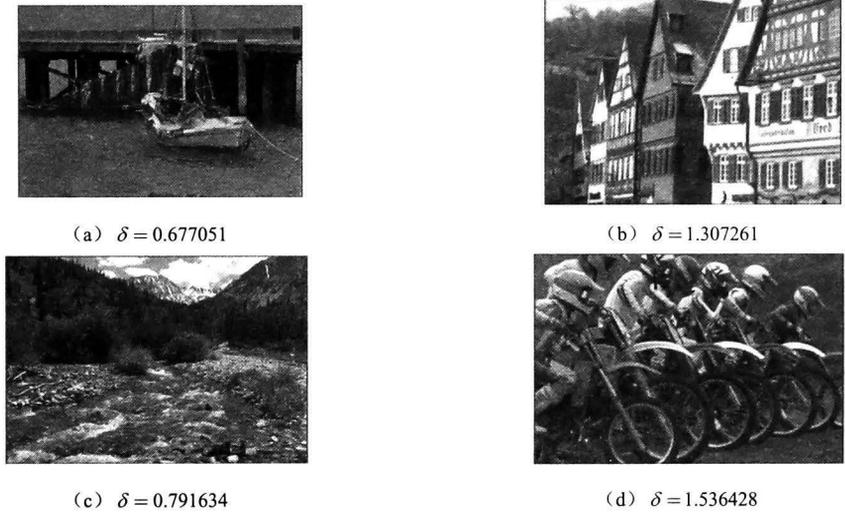


图 5-12 具有不同程度的高斯模糊图像， δ 为高斯模糊核标准差 (图像来源为公共图像数据库 LIVE^[12])

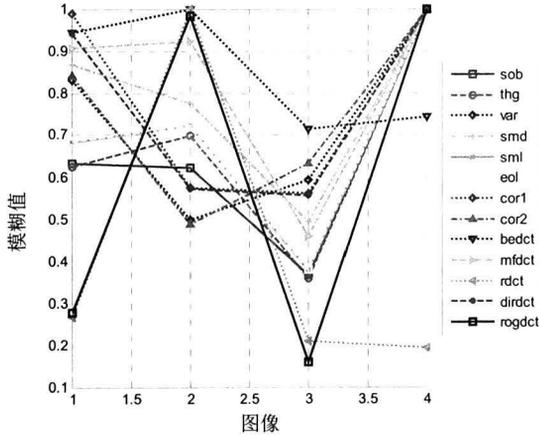


图 5-13 对图 5-12 各图像采用锐度评价函数计算得到的模糊度值

从以上分析可以看出，有必要研究能够满足对不同内容图像进行准确、客观估计模糊度的方法。同时模糊度评价对于自动聚焦应用也是有意义的，比如在判定触发聚焦时往往需要考虑图像模糊程度。将在 5.3.2 节介绍一种重要的模糊度评估方法。

5.3.2 基于视觉模糊阈的模糊度量

人类的视觉系统只能感知到边缘周围超过一定阈值的模糊度，这个阈值称为视觉可

辨模糊阈(Just Noticeable Blur, JNB)^[13],它是在视觉可辨差异(Just Noticeable Difference, JND)概念上提出来的。所谓视觉可辨差异是指能让视觉感知到灰度变化的前景与背景之间灰度的最小差异。因此,视觉可辨模糊也就是是指前景与背景对比度超过可辨差异时,在边缘周围可感知到的最小模糊度值,它的大小是与对比度大小紧密相关的。假定在灰度均匀为 I_B 的背景图像中,存在一个灰度均匀为 I_F 的前景长方块,并且对比度 $C = |I_F - I_B| > JND$ 。采用高斯平滑滤波对图像进行模糊化处理,将高斯滤波对应的标准差由低到高增加,直到能够在长方块边缘周边感知到模糊。于是, JNB 可用最小可察觉高斯模糊对应的标准差来表示,其具体值是可感知模糊概率为 63%所对应的高斯模糊标准差。在指定的对比度下,可感知模糊的概率符合心理函数,可采用以下模型来表示:

$$P = 1 - \exp\left(-\left|\frac{\sigma}{\sigma_{JNB}}\right|^\beta\right) \quad (5.30)$$

式中,常量参数 σ_{JNB} 和 β 可通过分析主观评价数据得到。一般 β 取值介于3.4~3.8,常取 $\beta = 3.6$ 。

对于一个边缘像素 e_i ,假定 $\omega(e_i)$ 表示边缘宽度, $\omega_{JNB}(e_i)$ 对应于可感知模糊的边缘宽度,则在边缘像素周边可感知模糊的概率为

$$P(e_i) = 1 - \exp\left(-\left|\frac{\omega(e_i)}{\omega_{JNB}(e_i)}\right|^\beta\right) \quad (5.31)$$

边缘宽度的具体计算可通过计算一个方向上灰度值增加的像素值,同时计算从另一个方向上灰度值递减的像素值,由这两个像素值之和减掉边缘像素即得到。得到单个边缘像素的可感知模糊概率的模型之后,采用概率和模型来表示整个区域 R 上可感知模糊的概率。采用概率和模型是基于两条假定条件:①模糊可感知到当且仅当至少一个边缘像素周边模糊可察觉;②每个像素的周边模糊检测概率是独立的;因此在 R 上能检测到模糊的概率为

$$P_{\text{blur}}(R) = 1 - \prod_{e_i \in R} (1 - P(e_i)) \quad (5.32)$$

也就是

$$P_{\text{blur}}(R) = 1 - \exp(-D_{(R)}^\beta) \quad (5.33)$$

其中,有

$$D_{(R)} = \left(\sum_{e_i \in R} \left| \frac{\omega(e_i)}{\omega_{JNB}(e_i)} \right|^\beta \right)^{\frac{1}{\beta}} \quad (5.34)$$

为了降低计算量,在求整幅图像可感知模糊的概率时采用块处理方法。首先将图像分为 64×64 的块状,对每一图像块进行边缘检测。如果某块的边缘数目比例超过阈值 T_{block} (常设为0.2%),则此块被标记为边缘块。对于每个边缘块 R_b 计算其可感知模糊概率,即

$$D_{(R_b)} = \left(\sum_{e_i \in R_b} \left| \frac{\omega(e_i)}{\omega_{JNB}(e_i)} \right|^\beta \right)^{\frac{1}{\beta}} \quad (5.35)$$

至此，对于整幅图像，其可感知模糊概率则可表示为

$$P_{\text{blur}}(I) = 1 - \exp(-D^\beta) \quad (5.36)$$

其中

$$D = \left(\sum_{R_b} |D_{(R_b)}|^\beta \right)^{\frac{1}{\beta}} \quad (5.37)$$

至此，得到基于视觉模糊阈的模糊度量公式（ L 为边缘块总数目）：

$$B = \frac{D}{L} \quad (5.38)$$

图 5-14 给出了基于视觉模糊阈模糊度量算法的流程图^[14]。

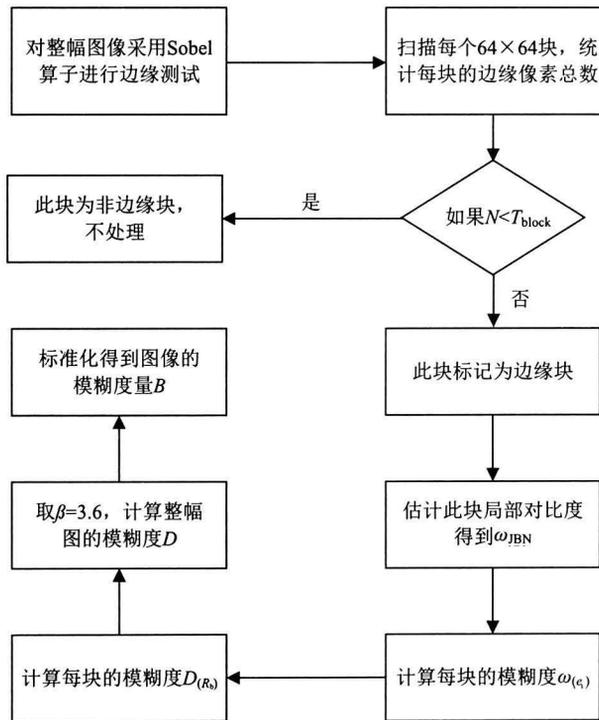


图 5-14 基于视觉模糊阈的模糊度量算法的流程图

为了简单地验证视觉模糊阈模糊度量算法的有效性，用它来估计图 5-12 中 4 幅图像的模糊度。在此算法中需要设定的参数有边缘块阈值 T_{block} 以及 Sobel 边缘检测算子中梯度强度阈值 T_{pixel} （只有当像素的梯度强度超过此阈值时才标定位边缘像素）。设定 $T_{\text{block}} = 0.2$ ， $T_{\text{pixel}} = 150$ ，得到如图 5-12 (a) ~ 图 5-12 (d) 所示的模糊度值分别为 11.20、19.29、11.51、33.73。由此可见，采用此算法能够定量地辨别出 4 幅图像的模糊程度。

基于视觉阈值模糊度量方法具有计算量小，对图像模糊度的估计与人类视觉机制较为相关。然而，此算法是基于空间阈图像边缘检测后的数据计算的，因此算法性能很大程度上取决于边缘检测的结果，对噪声会比较敏感；并且对参数 T_{pixel} 与 T_{block} 值非常敏感，需要仔细调节参数才能够获得准确的结果。因此，有必要研究解决对边缘检测敏感、过于依赖于具体参数设定的问题。

5.4 总 结

近些年来,研究更灵活、更稳定的基于数字图像处理和分析的自动聚焦方法以及研究客观无参照的图像模糊度评估方法都是图像处理和计算摄影领域里的热门问题。本章主要介绍了自动聚焦技术中基于图像分析的对比度检测自动聚焦方法,对其中涉及的两个关键问题,即聚焦值计算和聚焦峰值搜索进行了较为翔实的阐述。锐度评价函数决定了自动聚焦所能达到的精度,而聚焦峰值搜索则决定了自动聚焦的速度。因此,只有研究或采用适应性强、抗噪声能力强的锐度评价函数,以及设计稳定而高效的峰值搜索方法,才能够满足自动聚焦迅速、精确的要求。针对某些成像系统需要解决自动触发自动聚焦功能的问题,也简要介绍了能对不同场景内容的图像进行绝对模糊度估计的一种方法。

参 考 文 献

- [1] Santos A, Ortiz de Solorzano C, Vaquero J J, et al. Evaluation of autofocus functions in molecular cytogenetic analysis[J]. *Journal of Microscopy*, 1997, 188(3): 264-272.
- [2] Ng Kuang Chern N, Neow P A, Ang Jr M H. Practical issues in pixel-based autofocusing for machine vision[C]//Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on. IEEE, 2001, 3: 2791-2796.
- [3] Kristan M, Perš J, Perše M, et al. A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform[J]. *Pattern Recognition Letters*, 2006, 27(13): 1431-1439.
- [4] Lee S Y, Kumar Y, Cho J M, et al. Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning[J]. *Circuits and Systems for Video Technology*, IEEE Transactions on, 2008, 18(9): 1237-1246.
- [5] Shen C H, Chen H H. Robust focus measure for low-contrast images[C]//Consumer Electronics, 2006. ICCE'06. 2006 Digest of Technical Papers. International Conference on. IEEE, 2006: 69-70.
- [6] Jeon J, Lee J, Paik J. Robust focus measure for unsupervised auto-focusing based on optimum discrete cosine transform coefficients[J]. *Consumer Electronics, IEEE Transactions on*, 2011, 57(1): 1-5.
- [7] Kehtarnavaz N, Oh H J. Development and real-time implementation of a rule-based auto-focus algorithm[J]. *Real-Time Imaging*, 2003, 9(3): 197-203.
- [8] Ooi K, Izumi K, Nozaki M, et al. An advanced autofocus system for video camera using quasi condition reasoning[J]. *Consumer Electronics, IEEE Transactions on*, 1990, 36(3): 526-530.
- [9] He J, Zhou R, Hong Z. Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera[J]. *Consumer Electronics, IEEE Transactions on*, 2003, 49(2): 257-262.
- [10] Lee J S, Jung Y Y, Kim B S, et al. An advanced video camera system with robust AF, AE, and AWB control[J]. *Consumer Electronics, IEEE Transactions on*, 2001, 47(3): 694-699.
- [11] Rahman M, Kehtarnavaz N. Real-time face-priority auto focus for digital and cell-phone cameras[J]. *Consumer Electronics, IEEE Transactions on*, 2008, 54(4): 1506-1513.
- [12] H.Sheikh, Z. Wang, L. Cormack ,et al. LIVE Image Quality Assessment Database Release 2[EB/OL]. 2005. <http://live.ece.utexas.edu/research/quality>.
- [13] Ferzli R, Karam L J. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)[J]. *Image Processing, IEEE Transactions on*, 2009, 18(4): 717-728.

第 6 章 自动曝光与自动白平衡技术

6.1 概 述

6.1.1 自动曝光与自动白平衡的必要性

从第 2 章的介绍中知道摄像机拍摄需要调节的参数包括光圈、焦距、曝光时间（快门）、感光度 ISO、白平衡等。要调节好这些参数，拍出各种精美的艺术效果，摄影师需要掌握摄像机工作的基本原理及相应的摄影技术技巧。这并不是一件容易的事，往往需要多年的摄影积累，包括摄影理论与实际拍摄技巧两方面。对于广大普通用户而言，通常不会拥有丰富的摄影经验，甚至对摄影理论毫不知晓。他们使用摄像机常常只是为了留下一张相片或一段视频。因此“傻瓜摄像机”应运而生。它结构简单，使用方便，即使不懂摄影的人，只要装上电池，打开镜头盖，正确取景，按下快门，就能获得一张不错的相片。区别摄像机是否“傻瓜”的关键点，就是真正的傻瓜机应该具有自动聚焦、自动曝光、自动白平衡等自动调节参数功能。自动化的调节，大大方便了用户的使用。当然，某些特殊的艺术效果仍然需要摄影师手动调节，但是目前看来，傻瓜摄像机的自动调节功能已经能够满足用户的普通摄影需求。

另一方面，对于视频拍摄，例如，监控摄像、无人驾驶感知、航拍摄影等，用户需要拍摄一段时间内的连续影像。这段时间内，拍摄场景的距离、光照、色温等都可能发生变化。这时有人参与的手动调节摄像机参数已不再适用。因为可能人还没有调节好摄像机参数，感兴趣的目标就已经消失。此时，自动调节的重要性就凸显出来。目前的自动调节功能通常能够在数秒甚至更短的时间内调节到位，其调节速度是手动调节不可比拟的。目前，除非场景是固定的，可以预先设置好摄像机参数，其他情况下，几乎所有的视频拍摄都使用自动调节功能。

在第 5 章中，介绍了自动聚焦技术，本章将把目光放在自动曝光与自动白平衡技术上。

6.1.2 曝光的基本概念

1. 什么是曝光

摄影中的曝光可以理解为被摄影物体发出的光线，通过摄像机镜头投射到图像传感器 CCD 或 CMOS 上，形成原始图像的过程。这个过程类似于人们睁开眼睛，光线通过瞳孔进入眼球，在视网膜上成像的过程。在拍照时，按下快门按钮，快门打开，过一段时间再关上，在这段时间内，光线从镜头进入摄像机，照在感光器件上，感光器件上的感光像素记录照射到它上面的光。于是，就形成了一张照片。这也就是一次曝光过程。其中，快门控制镜头的开闭，它从开到闭的时间就是曝光时间。而光通过镜头的孔径大

小则由光圈大小决定。曝光时间越长，光圈越大，则进光量越大，图像越明亮。

有时在低照度的条件下，即使镜头的光圈完全打开，仍然无法得到明亮的图像，此时需要增大胶片或传感器的感光度 ISO。感光度 ISO 就是胶片或传感器对光线的敏感程度。感光度越高，相同的曝光量得到的相片亮度越高，即感光材料对光线越敏感，感光能力越强。例如，ISO 200 的胶片感光速度是 ISO 100 的两倍。ISO 400 大概是 ISO 200 的两倍。常用的还有 ISO 800，甚至 ISO 1600 的高感光度胶片。传统的胶片感光度是通过改变胶片的化学成分，来改变它的光线的敏感度。而数码摄像机的感光器件是不变的，它是采用电子信号放大增益技术实现了感光度的调整。感光度越高，增益也越大。通常以 ISO 100 为标准值，对应 1 倍增益，即不放大信号，而感光度为 ISO 200 时，感光度提高一倍，需要提高增益幅度。具体增益幅度，需要根据具体不同的传感器，由厂家测定并给出其对应关系。

总之，曝光是一个与光圈、曝光时间、感光度密切相关的综合过程。曝光正常的相片，画面明暗适当，过曝的相片，画面很亮，欠曝的相片，画面很暗，如图 6-1 所示。在图 6-1 中，欠曝时，整体昏暗，特别是树林深处，看不清树叶细节。过曝时，树叶反光太强，草地泛白。正常曝光时，明暗适中，能看见树叶细节。



(a) 过曝

(b) 正常曝光

(c) 欠曝

图 6-1 不同曝光的图像

2. 如何调节曝光

那么究竟该如何调节与曝光相关的因素呢？是不是光圈越大越好，曝光时间越长越好，感光度越大越好呢？答案是否定的。因为这三个参数不仅与曝光相关，而且与图像质量的其他方面也密切相关。

光圈与景深密切相关。光圈越大，景深越小。在景深范围内的物体可以清晰聚焦。景深小意味着能清晰聚焦的距离范围很窄。有时候，为了突出前景，需要小景深，即大光圈设置；有时候需要看清全局，也需要大景深，小光圈设置。所以光圈的设置不能随意设置，并不是越大越好。

曝光时间除了与图像亮度相关外，还与图像的运动模糊程度相关。当摄像机曝光时，运动物体会在传感器上留下一串影像。曝光时间越长，积累的影像就越多，模糊就越大。除了运动模糊外，使用手持式摄像机时，长时间曝光还容易引起抖动模糊。因此，曝光时间不宜设置过长。当然追求特殊艺术效果的应用除外。所以，曝光时间并不是越长越好。

感光度 ISO 除了与图像亮度相关外，还与图像的噪声密切相关。由于数码摄像机通过加大增益幅度的方式实现感光度的增加，所以它在放大图像信号的同时，也放大了噪声信号。感光度越高，噪声越大。所以，感光度也不是越大越好。

既然光圈、曝光时间、感光度 ISO 三者都与图像质量的其他方面相关，那么在调节曝光时，应遵循什么样的原则呢？这个问题的答案是不确定的。准确地说，应该根据用户需要拍摄的效果进行调节。例如，当拍摄前景突出的图像时，需要使用大光圈，此时应适当调小曝光时间与 ISO；当拍摄高速运动的物体时，需要使用短的曝光时间，此时应适当增大光圈与 ISO；当拍摄灯光流动效果时，需要使用尽量长的曝光时间，此时应适当调小光圈与 ISO；当拍摄视频时，曝光时间受到帧速率的约束，一般连续视频需要达到每秒 24 帧，即每帧图像曝光时间至多为 $1/24\text{s}$ ，约 41.7ms ；而在没有特殊需求时，人们总是希望得到清晰、无噪声的图像，即光圈不能太大、曝光时间不能太长，ISO 也不能太高。究竟应该如何调节，应该根据实际拍摄的场景与期望的艺术效果进行调节。但不管怎么调节，都需要得到正常曝光的图像，既不能曝光过度，也不能曝光不足。

为了使用户能够简便地使用摄像机，自动曝光诞生了。它代替了人的操作，由摄像机根据人类的视觉特点，自动对所拍摄场景进行曝光，获得具有正常亮度的图像。自动曝光工作时首先会估计当前场景的亮度，然后调节摄像机的光圈、曝光时间和感光度，使得到图像的亮度与预设的目标亮度一致。在摄像机参数调节策略上，自动曝光一般不会优先调节感光度 ISO，因为增大它，将带来噪声。而对于光圈与曝光时间（即快门速度），主要有三种方式：光圈优先方式、快门优先方式，或者为不同亮度等级设定固定的光圈快门组合方式。当对运动模糊要求较高时，通常选择光圈优先方式，即优先调节光圈，然后调节曝光时间，最后调节感光度。当对景深要求高时，通常选择快门优先方式，即优先调节曝光时间，然后调节光圈，最后调节感光度。光圈快门组合方式则是二者的折中。当然有些自动曝光策略也会将感光度与曝光时间综合考虑，因为一些人认为曝光时间过长导致的运动模糊是难以忍受的，另一方面过高的 ISO 带来的噪声也是难以忍受的，所以需要二者综合考虑。究竟采取哪种方式，目前并没有一个固定的标准，需要根据用户的拍摄需求决定。但是无论采取什么样的策略，自动曝光的目标都是使画面正常曝光，不过曝，也不欠曝。

对于自动曝光调节参数的方式，一般摄像机都会提供一种默认方式，和数种可选方式，用户可以根据自己的需求进行选择。但是无论采取哪种方式，自动曝光也只能满足大多数的日常拍摄需求。在面对一些特殊的拍摄需求时，自动曝光就显得无能为力。以高速摄影为例，要拍摄运动的火车，曝光时间需要在 $1/1000\text{s}$ 以内，此时就只能限制曝光时间，当曝光量不够时，就必须增大光圈。而如果恰好又需要大景深，则光圈的大小也必须限制，那么就只能调节感光度 ISO。虽然 ISO 会带来噪声，但这也比拍摄到一张黑暗的图像好。显然，自动曝光无法知道它面对的是不是一个高速运动的物体，也不会知道拍摄者是否一定要拍摄清晰的运动火车，或者只要背景清晰即可。要完成这样的工

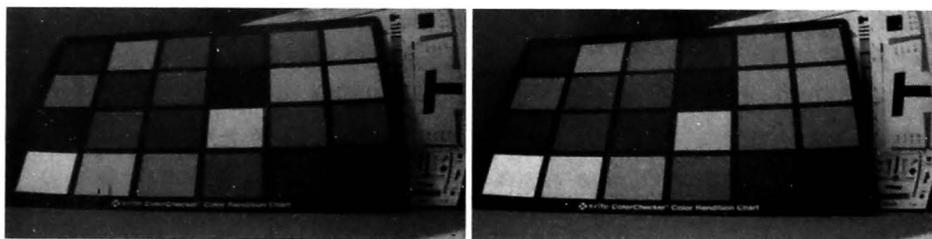
作，就只能依靠手动调节完成。

自动曝光存在的意义就在于它简化了普通用户拍摄的难度。另一方面，在视频拍摄时，面对亮度变化的场景，自动曝光则是必不可少的功能之一。自动曝光主要可以分为拍摄场景的亮度估计，即测量场景光照，与摄像机参数调节策略两个方面。其中，场景光照测量是否准确是影响自动曝光性能的主要方面。6.2节中将对自动曝光技术，特别是测光技术进行详细介绍。

6.1.3 白平衡的基本概念

1. 什么是白平衡

当人们用眼睛观察自然世界时，在不同的光线下，对相同颜色的感觉基本是相同的，比如在早晨旭日初升时，人们看一个白色的物体，感觉它是白的；而人们在夜晚昏暗的灯光下，看到的白色物体，感觉它仍然是白的。这是由于人类从出生到成长的过程中，大脑已经对不同光线下的物体的彩色还原有了适应性。这种现象称为颜色恒常性^[1]。不幸的是，CMOS或CCD等感光器件没有这样的适应能力。白色物体在暖色灯光照射下，拍摄的图像呈橘红色，如图6-2(a)所示；在冷色灯光照射下，拍摄的图像呈淡蓝色，如图6-2(b)所示。



(a)

(b)

图 6-2 偏色的图像

为什么在不同色温的光源下，物体会出现偏色呢？这是因为图像传感器只是记录了所有投射到其上的光线，其本身并不能分辨投射到其上的色光是由物体本身的色彩反射而成，还是由偏色的环境光造成。所以传感器是没有颜色恒常性能力，不能够适应这种色光的变化。所以当它真实呈现出所拍摄到的图像时，就出现了偏色。

为了使得摄像机也具有颜色恒常性能力，需要使用白平衡技术。所谓白平衡(White Balance)，简单地说就是去除环境光的影响，还原物体真实的颜色，把不同色温下的白颜色调整正确。从理论上说白颜色调整正确了，其他色彩就都准确了。即在红色灯光照射下，白色物体依然呈白色，在蓝色灯光照射下也呈现白色。

讲到白平衡离不开一个概念——色温。色温(Color Temperature)是表示光源光色的尺度，单位为K。英国著名物理学家开尔文认为，假定某一纯黑物体(称为“标准黑体”)，能够将落在其上的所有热量吸收，而没有损失，同时又能够将热量生成的能量全部以“光”的形式释放出来的话，它便会因受热的高低而变成不同的颜色。如图6-3所示，将标准黑体从绝对零度开始加热，温度逐渐升高，光度亦随之改变，黑体呈现出由

红变化为橙红、黄、黄白、白、蓝白、蓝的过程。而一天当中，阳光的色温也随时间变化。日出后 40min 光色较黄，色温在 3000K 左右；正午阳光雪白，色温上升至 4800~5800K，而阴天正午时分，则色温大约为 6500K；日落时分，阳光又偏红，降至 2200K 左右。

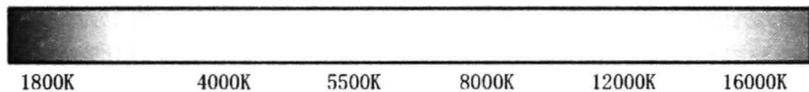


图 6-3 色温变化示意图

注意：色温的温度标准不是摄氏度，而是开氏度。摄氏度的符号是“℃”；而开氏度的符号是“K”。它们的换算关系是：开氏度=摄氏度+273 度。例如，20℃=293K。不同的色温会给人带来不同的心理感受。大于 5000K 的色温，光色呈带蓝的白色，给人一种清凉的感觉，属于比较冷的气氛；3300~5500K 之间的色温，光色呈比较正的白色，给人一种爽快的气氛；小于 3300K 的色温，光色呈带红的白色，给人一种温暖的感觉，属于比较稳重的气氛。

2. 如何调节白平衡

在讲述如何调节白平衡以前，首先需要明确一点，摄影师并不是总是一味地追求色彩的还原。有时为了达到某种艺术效果，常常需要将拍摄的照片故意调整为某种色调。例如，追求怀旧效果时，常将相片设为棕色；烘托温暖效果时，常将相片设为偏黄的暖色调等。除了特殊的艺术追求外，通常情况，人们还是希望通过白平衡得到真实的色彩信息。

白平衡的调节一般是通过改变红、绿、蓝三色的比例关系实现。当环境光为纯白光时，白色物体本身就呈现白色，红、绿、蓝三色的感光电路电子放大比例是相同的 1:1:1 的关系。当出现色光后，白色出现了偏色，改变后的红、绿、蓝三色比例为 $\alpha:\beta:\gamma$ ，那么感光电路电子放大比例将做相应的逆向调整 $1/\alpha:1/\beta:1/\gamma$ ，使其仍然保持 1:1:1 的关系。这样白色就依旧是白色。那么如何才知道红、绿、蓝三色偏色的比例 α 、 β 与 γ 呢？那就需要对环境色温进行测光。如果已知环境色温，那么就确定了 α 、 β 与 γ 之间的比例关系，就能够进行相应的白平衡调节。

目前，白平衡的调节主要有三种方式：预设白平衡、手动白平衡与自动白平衡。

预设白平衡是指在已知环境色温的情况下，人工把摄像机的色温设置调整到已知的环境色温，以达到正确的白平衡效果。例如，当环境色温为 3200K 时，可以把摄像机色温滤光片设置在 3200K，景物可以得到正确的色彩还原；当环境色温为 5600K 时，摄像机色温滤光片设置在 5600K，景物可以得到正确的色彩还原。一般情况下，摄像机有多个挡位可供选择，使用时，可以根据环境情况，预先设置滤光片的挡位。预设白平衡需要色温计，以正确地测量环境色温。

手动白平衡是指摄影师把摄像机对准白纸拍摄，这时，白纸充当标准白色，摄像机需要通过内部自动调整，即估计出红、绿、蓝三色偏色的比例为 $\alpha:\beta:\gamma$ 并做相应的调整，使拍摄出的画面呈现纯白色，以此达到白平衡的目的。估计出偏色比例后，以此作为对环境色温的估计，以后的拍摄都以这个色温为准。手动白平衡测量时，白纸需要占画面面积的一半以上。

在手动白平衡中，白纸的选择也很有奥妙。白纸具有不同的色度，有偏青的复印纸，

也有偏黄的白报纸，不同的白纸色度对画面色调有不同的影响，其规律是白纸的色度和画面的色度成补色关系，如图 6-4 所示。如果用浅黄色纸调白平衡可以使画面呈现蓝色调，这是拍夜景经常用的方法。相反选择蓝青色纸调白平衡可获得暖色调。掌握这个规律还可以有意识地创作出带有主观色调的片子，烘托艺术气氛。

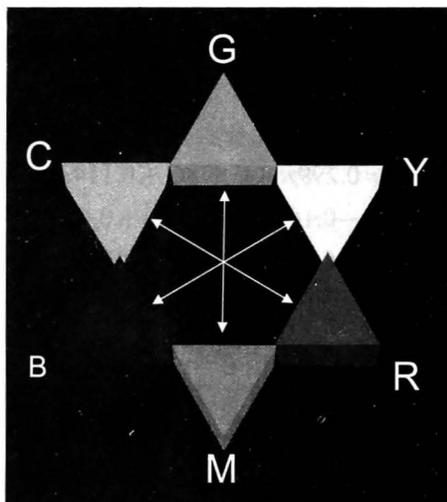


图 6-4 白纸的色度和画面色度成补色关系

显然，手动白平衡方法，虽然脱离了色温计的束缚，但是却不能适应色温变化的环境。一旦色温变化就需要重新调整，非常麻烦，不利于连续视频的拍摄。

自动白平衡比手动白平衡更进一步，达到了全自动的白平衡调节。如柯达公司的广告语“*You Press the Button, We Do the Rest.*”所述。用户只需按下按钮，摄像机能够自动把不同色温下的白颜色调整正确。摄像机采用的这种技术就是自动白平衡技术。当摄像机对着被摄体时，随着照明光的色温不同，摄像机的白平衡被自动调整，而不必手动控制。

从预设白平衡，到手动白平衡，再到自动白平衡，整个白平衡技术的关键都围绕着能否正确地估计环境色光上。自动白平衡技术的关键点也在于此。不同白平衡技术采用了不同的方法来寻找画面中的白平衡基准点，以此来达到白平衡调校。决定白平衡基准点的不同，就产生了不同的自动白平衡算法。在 6.3 节中将重点介绍不同的自动白平衡算法。

与自动曝光一样，自动白平衡技术也只能满足一般普通用户的拍摄需求，当需要追求某些特殊的艺术效果时，仍然需要手动的色温调节。这里手动的色温调节指人工设置红、绿、蓝三色间的比例关系，而不是指上述的拍摄白纸的手动白平衡技术。

6.2 自动曝光技术

本节将详细介绍自动曝光技术。自动曝光可以分为亮度获取、亮度分析、曝光调节三个步骤。从实现方式上，自动曝光又可以分为光学方式与电子学方式。光学方式的自

动曝光需要使用单独的光照传感器来测量光照度，从而进行场景亮度获取。获取亮度后，再结合图像信息与当前光圈值、曝光时间和感光度值信息进行分析，判断当前曝光参数是否合适。如果不合适则进行曝光调节。曝光调节指调节光圈、曝光时间和感光度这三个曝光相关参数。传统的胶片摄像机以及部分数码摄像机采用的就是这种方式。电子学方式直接从图像传感器输出的图像信号中获取亮度信息，还能够额外提供对场景的分析能力。相比于光学方式，它节省了测光设备，目前已经被绝大多数数码摄像机采用。由于电子学方式是直接通过摄像机镜头测光，因此这种方式也被称为 TTL (Through The Lens) 方式。下面将重点介绍这种方式。

$$\begin{cases} Y = 0.299R + 0.578G + 0.114B \\ U = -0.147R - 0.289G + 0.436B \\ V = 0.615R - 0.515G - 0.100B \end{cases} \quad (6.1)$$

对于彩色图像，图像传感器获得的信号属于 RGB 色彩空间。经过式 (6.1) 的色彩空间转换，可以得到 YUV 色彩空间，其中亮度分量 Y 包含了 93% 的图像能量。获取了每个像素的亮度后，需要对图像的整体亮度情况进行分析，判断其是否过曝，或欠曝。如何判断其是否过曝或欠曝呢？需要设置一个理想的视觉亮度作为目标。那么究竟怎么样的亮度才是理想的视觉亮度呢？数码摄像机做了一个假设，采用了一个最简单也是最有效的标准，就是假设被摄物体具有某一固定的反光率，而这个固定的反光率为 18%。人眼产生的视觉效果中 18% 刚好是中间灰度。对于 8 位的摄像机系统，具有 255 个灰度级，128 就是它的中间灰度。数码摄像机在执行自动曝光时需要达到的目标就是使得曝光后的图像具有接近中间灰度的整体亮度水平，这也就是理想视觉亮度了。需要指出的是，18% 或者中间灰度的假设在大多数情况下是可行的，但是不可避免也有失效的时候，这时候就需要采用曝光补偿技术了，这部分内容后面会讲到。将得到的图像亮度和目标亮度进行比较后，就可以根据分析结果进行曝光调节了，电子学自动曝光基本框架如图 6-5 所示。

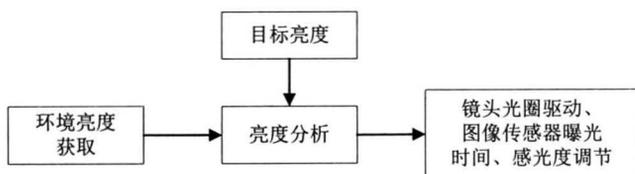


图 6-5 自动曝光基本框架

曝光调节需要遵循一定的优先原则，如 6.1.2 节所述，包括光圈优先、快门优先等。摄像机厂商一般会提供若干模式，供用户选择。如果已知需要的曝光量，那么根据用户选择的调节策略进行调节已经是一件非常成熟的技术了。现在问题的重点集中在如何对环境亮度，即当前图像亮度进行分析，从而确定正确的曝光量？如何判断当前曝光值是否过曝或欠曝？是对整个图像，还是图像中的某一块区域呢？这就需要深入到测光技术中。本节接下来的部分首先将介绍常用的测光技术，然后会介绍一些考虑了背光及正面强光等特殊场景的曝光控制算法，这些算法较之常用测光技术研究性更强，而常用测光技术基本已经用于商业摄像机中。

6.2.1 常用测光技术

一般具有测光功能的摄像机上都有至少三种基本的测光方式：中央重点平均测光、中央部分测光和点测光。

中央重点平均测光又叫做中央平均测光。中央平均测光主要是考虑到一般摄影者习惯将拍摄主体也就是需要准确曝光的东西放在取景器的中间，所以这部分拍摄内容是最重要的。在进行亮度分析时将中央部分亮度的权重分配较大的比例（约 70%），周围部分分配较小的比例（约 30%）进行平均亮度计算：

$$B_{\text{est}} = \frac{\sum_{i \in I(i)} w_i y_i}{N} \quad (6.2)$$

式中， B_{est} 为估计的场景亮度； i 为图像中的像素位置； $I(i)$ 为整个图像范围； N 为图像中像素的总数目； y_i 为各像素的亮度； w_i 为每个像素点赋予的权重。中央平均测光法中，越靠近中央的像素，权重 w_i 越大。显然，当各点的 w_i 相等时，中央平均测光退化为全局平均测光。中央平均测光方式简单实用，但是如果画面主体不在中央，以及取景范围内明暗分布不均时，这种方式确定的曝光量往往不够准确，如图 6-6 所示。



图 6-6 中央重点平均测光的曝光效果

中央部分测光又称为局部测光，测量取景框中央长方形或圆形（椭圆形）范围内的平均亮度：

$$B_{\text{est}} = \frac{\sum_{i \in C(i)} w_i y_i}{N_C} \quad (6.3)$$

式中， $C(i) \subset I(i)$ ，为选取的中央测光范围，大约占整个画面的 3%~12%； N_C 为测光范围内的像素总数。在 $C(i)$ 范围内，各像素权重 w_i 相等。中央部分测光可以得到拍摄主体准确曝光的照片，避免了环境光线的影响，能够确保主要表现对象所需的曝光量，如图 6-7 所示。

点测光的测光方式与中央部分测光类似，但是测光范围 $C(i)$ 更小，选取画面的 1%~3% 的范围求取平均亮度。测光范围的选取也不完全局限于中央部分。点测光基本不受

测光区域外其他景物亮度的影响，能够做到对表现主体的准确曝光，如图 6-8 所示。

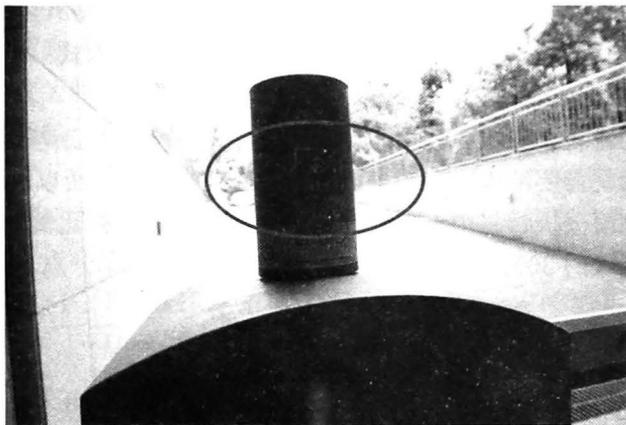


图 6-7 中央部分测光的曝光效果

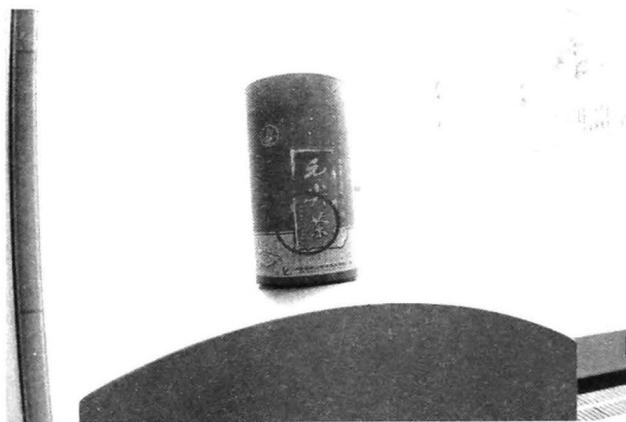


图 6-8 点测光的曝光效果

三种测光模式各有所长，在不同的拍摄场景下，需要灵活使用。中央平均测光保持了明亮与暗淡区域的平衡；中央部分测光使画面中央区域正确曝光；点测光使被曝光主体正确曝光。

6.2.2 更先进的自动曝光控制

图 6-9 所示是室内、室外高对比度条件下的三种测光模式效果对比。中央平均测光，室内曝光较准确，但是窗外明亮部分的细节几乎损失。点测光对焦于室外部分，但是室内明显曝光不足。中央部分测光，兼顾了部分室外与部分室内，取了二者间的一个平衡。但是事实上无论哪种测光方式都很难兼顾到整个画面的亮度。当光照条件是背光（逆光）或是强正面光时，由于场景中最亮和最暗区域具有强烈的对比度，如果仍然采用 6.2.1 节所述的常用测光方法，虽然整体的图像亮度仍然是“中间灰度”，但是图像中重要的主体区域就肯定遭遇曝光不足或者曝光过度，所以必须采用更先进的自动曝光控制方法。

一种常用的方法是使用人工反光板进行光照补偿。这种方式常见于各种艺术拍摄



(a) 中央平均测光



(b) 点测光



(c) 中央部分测光

图 6-9 高对比度条件下的测光

中。这种方式简单实用，但是对于个人摄影爱好者，需要携带额外的设备，显得较为麻烦。除了人工光照补偿外，还可以采用手动设置曝光补偿的方式。现在大部分数码相机都有曝光补偿设置。曝光补偿是指在现有的曝光参数基础上再对其施加一定的补偿值。常见的是对曝光时间的补偿。在背光情况下，为突出正面主体增加曝光时间，在正面强光下，减少曝光时间。除了曝光参数的补偿外，还可以采取目标亮度补偿，即在“中间灰度”的基础上增加一个补偿值，达到相同的效果。

曝光补偿后的效果如图 6-10 所示。对比左侧未进行曝光补偿和右侧进行了曝光补

偿的照片，可以看出进行了曝光补偿后，拍摄主体——树木的亮度改善了很多，视觉效果更好了。

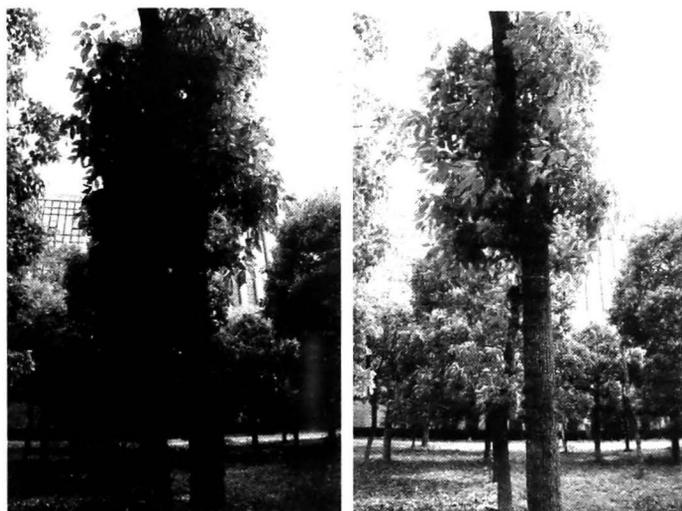


图 6-10 曝光补偿前后效果图

那么除了手动的曝光补偿外，有没有自动的曝光补偿，甚至说在自动曝光时能够主动识别主体的曝光控制算法呢？答案是肯定的。除了工业界已经商业化使用的一些算法外，学术领域也有许多研究成果。例如，分区加权测光、基于亮度直方图的测光等。

1) 分区加权测光

分区加权测光最早由尼康（Nikon）公司提出，也称为评价测光。它与中央平均测光相比最大的不同就是分区加权测光将取景画面分割为若干测光区域，每个区域独立测光后再整体加权计算出整体的曝光。目前，除了尼康外，佳能、美能达、宾得等品牌的摄像机都推出了各自的分区加权测光算法。它们的区别在于测光区域分布、加权分析算法不同。例如，尼康将测光区域分割为八个部分，而佳能顶级机器上设计了 21 个区域的 TTL 测光模式。除了商业摄像机外，在学术研究领域也有许多针对分区加权的算法。

Haruki 与 Kikuchi^[2]提出了一种基于模糊逻辑的算法，首先将屏幕分割为六个区域，通过赋予各区域亮度累加值不同的动态权值将曝光重点放在屏幕中央。Kuno 等人^[3]通过跟踪目标来动态地改变分区权值矩阵，确保对动态目标的合适曝光。Lee 等人^[4]也提出了一种健壮性的基于固定分区的测光方法。此外，还有根据各区域亮度累加值等参量与各种典型场景的对应关系来进行曝光控制的研究^[5]，这需要图像数据库的支撑。下面重点介绍一种适用于大动态范围场景的自动曝光控制算法^[6]。

文献[6]的主要思想是将图像分为若干小区域，通过区域间的相邻对比度变化，找出图像中的拍摄主体的背光区域或正面强光区域，实现主体与背景的分割。然后对主体平均亮度与背景平均亮度进行加权平均，求得整个图像的平均亮度。其权重则由主体与背景的对对比度决定。其具体步骤如下。

第一，将图像平均分为 $M \times N$ 个小块，如图 6-11 所示。图中的灰色区域为预置的“主体区域”。即预计被拍摄主体只会出现在“主体区域”内，搜索时也在主体区域范围内搜索。每一小块的亮度由该块内像素的平均亮度 B_{ij} 表示。两个相邻块之间的对比度由垂直

和水平方向亮度的一阶导数表示。如果对比度接近 0，那么相邻两块是连续的，它们很可能属于图像中的同一区域。如果对比度大于某一特定值，则相邻两块的亮度不连续。

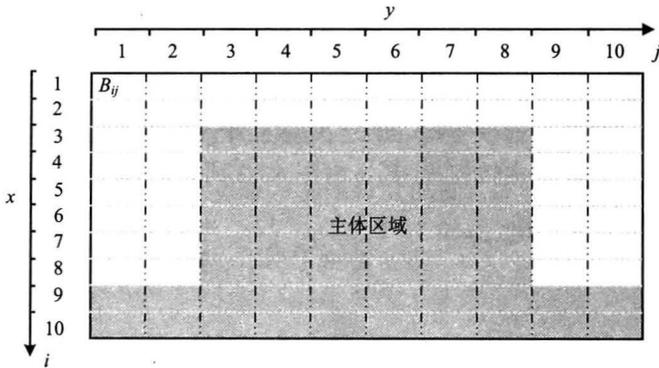


图 6-11 图像窗口分割^[6]

第二，根据阈值 T 将对比度 C_{ij} 分为三种类型。连续亮度 $|C_{ij}| < T$ ，正向突变 $C_{ij} \geq T$ ，负向突变 $C_{ij} \leq -T$ 。

第三，通过对比度类型对图像的垂直和水平方向进行搜索，找出彼此具有强对比度的连续亮度区域。这些连续区域很可能就是图像的主体和背景。搜索时，首先进行水平方向搜索，然后进行垂直方向搜索。搜索过程中，如果负向突变和连续亮度相邻出现，说明后续的块有可能属于暗的区域。在算法中对这种情况以一个暗区标志标记。如果在这个块后面的若干块后有连续亮度和正向突变相邻出现，而暗区标志也同时有效，那么所被标记的块就被确定为此方向上的连续暗区。另一方面，如果在暗区标志有效后的若干块后，出现了连续亮度和负向突变相邻的组合，也就是说对比度朝同一个方向改变，那么暗区标志将会被取消，而此前被标记的块也不会被确定为连续暗区，因为它们并不属于一个封闭的亮度区域。同样的，在正向突变和连续亮度以及连续亮度和负向突变间的块将被确定为该方向上的连续亮区。图 6-12 所示的是搜索后的结果。其中搜索只在灰色区域进行，黑色的圆点为连续亮区，块数量为 N_b ，白色圆点为连续暗区，块数量为 N_d 。

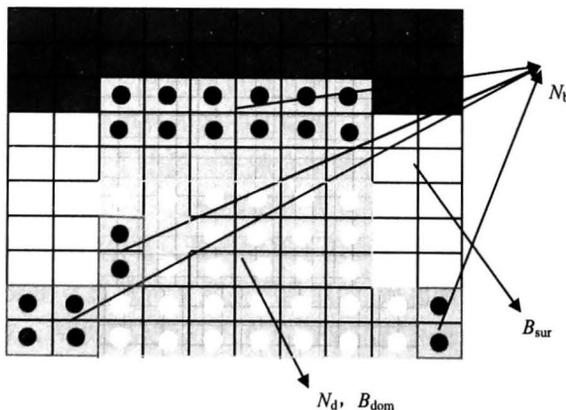


图 6-12 连续亮度区域搜索结果^[6]

第四，光照条件判断。如果灰色的主体区域，由连续暗区 N_d 主导，那么为背光，

主体就是主体区域中的连续暗区；如果主体区域由连续亮区 N_b 主导，那么为强正面光，主体就是主体区域中的连续亮区；否则为普通光照。在图 6-12 中，连续暗区 N_d 为主导，确定为背光，那么主体就是 N_d 所在区域。而 B_{dom} 为主体的平均亮度。注意这里的主体只是连续暗区，而不是整个灰色区域代表的主体区域。除去主体之外的区域为背景，其平均亮度为 B_{sur} 。

第五，对图像区域亮度加权平均，确定图像的整体亮度。权重由主体与背景的对比值 C 决定。 $C = B_{dom} - B_{sur}$ 。主体权重 W_{dom} 由图 6-13 所示的函数确定。当 $C < C_L$ 时 $W_{dom} = 0.5$ ，当 $C_L \leq C \leq C_H$ 时， $W_{dom} = 0.5 + 0.5 \frac{C - C_L}{C_H - C_L}$ ，当 $C > C_H$ 时， $W_{dom} = 1$ 。背景权重

$$W_{sur} = 1 - W_{dom}$$

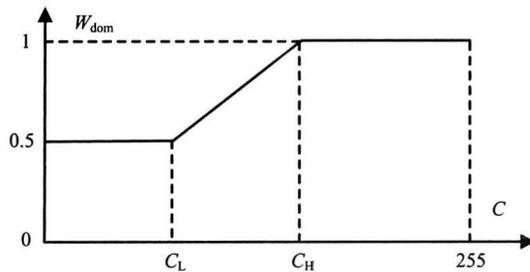


图 6-13 主体权重函数^[6]

加权平均后图像的亮度为

$$BI = B_{dom} \times W_{dom} + B_{sur} \times W_{sur}$$

得到平均亮度后，就可以与预设的目标亮度比较，确定是过曝还是欠曝，再通过曝光调节模块调整，即可实现自动曝光。

文献[6]提出的方法是一种典型的通过分析分区亮度，动态确定权重的分区加权测光方法。

2) 基于亮度直方图的测光

除了分区加权测光外，研究者还提出了基于图像统计信息，如亮度直方图的测光方法。Tao 等人^[7]提出用多权值模板和直方图来克服采集图像过程中因野外环境中亮度变化大、动态范围广等造成的曝光不足或曝光过度等问题。文献[8, 9]利用亮度直方图的几何分布信息调整摄像机的曝光参数，使输出图像的亮度直方图满足一定的要求。而谷明琴等人^[10]则将分区加权与亮度直方图结合起来，提出了一种新的自动曝光方法。下面重点介绍这种方法，特别是其中的亮度直方图部分。

该方法是面向无人驾驶系统，主要用于路面场景。首先将画面平均划分为 9 个区域，然后在 9 个区域中确定出感兴趣区域；然后利用亮度 V 分量直方图来判断图像的曝光情况，继而依据感兴趣区域的位置和曝光情况来选择 9 个区域权重矩阵，对感兴趣区域和曝光不足的区域分配更多的权重；最后根据预设的目标亮度、当前帧的亮度值以及当前帧的摄像机曝光参数来调整摄像机的新的曝光，完成车载摄像机的自动曝光。可以看出，该算法的主体框架还是分区加权测光，但是在每个区域分析时，采用了对亮度直方图分析的方式。

该算法确定的感兴趣目标是道路中的交通标志，利用文献[11]的方法精确检测交通

标志,并分析交通标志在各个区域的面积大小,将所占面积最大的区域作为感兴趣区域。当在所有区域中都未检测到交通标志时,则以正中间区域为感兴趣区域。对大量含有交通标志的图像,统计交通标志在9个区域中出现的概率。概率最大的即为感兴趣区域。

分析各区域曝光情况时,首先将RGB颜色空间转换为HSV空间,并计算亮度V分量的直方图。直方图中每一级的统计值对应着能量强度,级数越高,图像亮度越大。即可根据能量的聚集区域来判断图像的曝光情况。为此,将V分量直方图均分为256份,计算 $S_1 = \sum_{i=1}^{128} N_i$ 及 $S_2 = \sum_{i=129}^{256} N_i$,V分量直方图的峰值位置 $V_{LP} = L\left(\max_{i=1,2,\dots,256} N_i\right)$ 。通过计算直方图左、右两部分的面积比来判断图像的曝光情况:

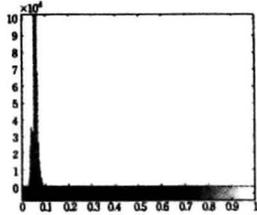
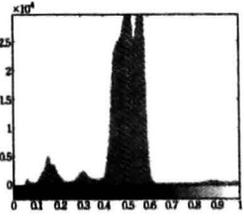
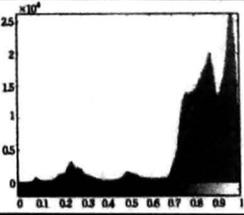
$$P = \begin{cases} S_2 / S_1, & S_1 \neq 0 \\ \infty, & S_1 = 0 \end{cases} \quad (6.4)$$

当 S_1 为0时, P 为无穷大,此时一定为曝光过度。通过分析在不同场景下,用不同曝光值采集的图像,根据 P 和V分量直方图的峰值位置 V_{LP} ,来划分不同的曝光情况:

$$\text{曝光} \begin{cases} \text{不足, } V_{LP} < 0.47 \cap P < T_1 \\ \text{过度, } V_{LP} > 0.53 \cap P > T_2 \\ \text{正常, 其他} \end{cases} \quad (6.5)$$

为了确定 T_1 和 T_2 两个阈值,在相同场景下,将摄像机的曝光时间从最小变化到最大,能观察到一个场景中图像从曝光不足到曝光过度, P 值的逐渐变化过程。通过比较不同场景下的 P 值变化曲线,阈值 T_1 和 T_2 分别取0.75和2时,能有效地区分图像的曝光类型。表6-1给出了三种曝光情况下 P 值和V分量直方图的示例。

表6-1 三种曝光情况示例^[10]

曝光程度	P 值	V分量直方图
曝光不足	0	
曝光正常	0.546	
曝光过度	11.428	

当感兴趣区域以及各区域曝光情况确定后,就可以选择相应的权值矩阵了。文献[10]为不同情况提供了多种权值模板。模板中的每一个矩阵都增大曝光不足区域和感兴趣区域的权值,减少了背景区域的影响,例如在图像顶部的天空背景。最后通过加权平均求得最后的图像平均亮度。

3) 其他测光方式

除了上述的分区加权测光和基于亮度直方图的测光方式外,还有许多其他的测光方式。有实现自动曝光补偿的方法,如 Shimizu 等^[12]提出的算法为每帧图像计算出三个亮度门限,按照全帧图像中亮度值大于不同的门限的像素比率来区分背光、正面强光等特殊场景,并进行相应的曝光补偿。Murakami 等^[13]提出的自动曝光系统在对图像主要拍摄对象进行曝光补偿的同时使用图像的色彩信息保持主体与背景的协调,在通过模糊逻辑处理系统补偿摄像机曝光时要保留背景的有用信息。各种对比度提升技术也被用来进行曝光控制^[14~17]。还有基于图像分辨率和对比度最大化的算法^[18]。甚至机器学习中基于神经网络的方法也被用于自动曝光^[19, 20]。

文献[20]的神经网络方法结合了分区加权、直方图分析。该方法认为图像亮度和曝光值之间存在一种非线性关系,可以通过神经网络训练得到。图像被分为如图 6-14 所示的五个区域:中间区域、上面区域、下面区域以及两个侧面区域。这种划分方法是为了使自动曝光适用于风景照、人物照等不同类型的照片,不同类型的照片所需要的曝光重点不一样。每一个区域的直方图决定了各自的光通量的平均值、方差、最小值以及最大值。光通量的平均值与平均亮度以及对比度有关,对比度是测定图像清晰度的。计算已有图像帧的同样空间信息,将光通量和亮度的变化信息综合到神经网络中去。实验确定了决定最小均方差的隐藏层神经元个数。所采用的神经网络包括两个隐藏层,隐藏层中共有 20 个处理单元,其中 15 个确定了最小总均方误差。输出神经元的个数取决于训练神经网络所得到的曝光值。更详细的介绍,可参考本章后列出的参考文献。

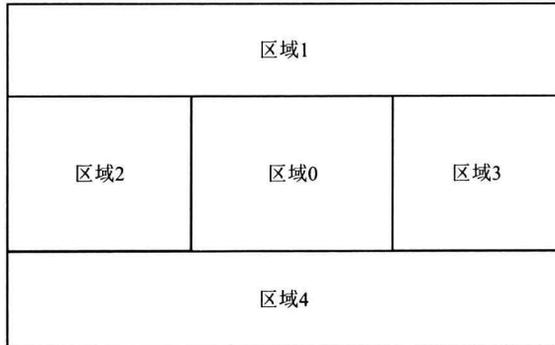


图 6-14 图像分区示意^[20]

这里需要指出的是,无论是哪种测光技术,简单的或复杂的,自动曝光的目标都是将拍摄主体曝光准确。但是如果画面主体有多个目标,自动曝光将无能为力。在图 6-10 中,既要看清窗外,也要看清室内时,自动曝光将无法满足这种需求。此时,面对高动态范围场景,需要新的计算摄影技术 HDR 成像技术的帮助,这部分内容将在第 7 章中介绍。

6.3 自动白平衡技术

6.1.3 节已经讲述了三种白平衡调节方式：预设白平衡、手动白平衡与自动白平衡。本节中，将就自动白平衡技术做进一步的探讨。首先，从理想朗伯特模型(Lambert Model)出发，发现色温对成像的影响以及校正方法。

朗伯特反射模型（又称为理想漫反射模型）是对不光滑表面的一种理想化近似。朗伯特模型假设所有方向上的反射光线强度相同，不随入射方向或观测方向的改变而改变。根据朗伯特反射模型^[21, 22]，图像的成像过程主要受三个因素影响：光源、物体的反射材质属性和摄像机的光谱响应函数。因此，图像成像模型可表示为

$$\rho(x) = \int_{\omega} S(x, \lambda) L(\lambda) C(\lambda) d\lambda \quad (6.6)$$

式中， x 为图像的空间坐标位置； λ 为光谱波长； $\omega \in \mathfrak{R}$ 为可见光波长的范围（380~780nm）； $S(x, \lambda): \mathfrak{R} \times \mathfrak{R} \rightarrow [0, 1]$ 为场景 x 处的光谱反射函数； $L(\lambda)$ 为光源的光谱功率分布； $C(\lambda) = [C_r(\lambda), C_g(\lambda), C_b(\lambda)]^T$ 为成像设备对于红色 r 、绿色 g 和蓝色 b 三色波段的光谱响应函数，如图 6-15 所示； $p(x)$ 为 x 处对应的 RGB 像素值，即 $p(x) = [p_r(x), p_g(x), p_b(x)]^T$ 。

不同色温的光源有着自己不同光谱功率分布 $L(\lambda)$ ，如图 6-15 所示。 $L(\lambda)$ 的不同，导致了最终成像的 RGB 像素值的不同。假设没有物体反射，只是记录光的颜色，那么光照颜色可以表示为

$$e = \int_{\omega} L(\lambda) C(\lambda) d\lambda \quad (6.7)$$

光谱功率分布 $L(\lambda)$ 的不同，导致了光的颜色的不同，如图 6-15 所示。当然，其实光的颜色也只是人眼对光谱的反应。光谱不同，人眼就看到了红光、黄光、绿光、蓝光等。所以白光也只是人眼对某一种特定光谱的反应。将人眼看到的白光对应某一特定的光谱 $L_c(\lambda)$ ，定义为标准光源，色温在 5500K 左右。显然，如果能够知道当前光源的光谱 $L(\lambda)$ 和标准光源 $L_c(\lambda)$ ，就能够还原出物体在标准光源下的色彩。但不幸的

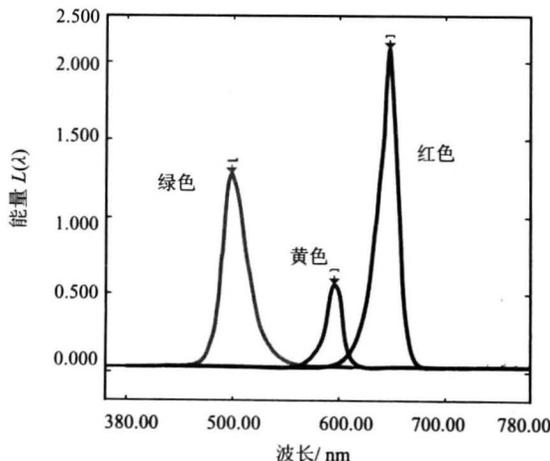


图 6-15 不同颜色色温的光谱

是, 光谱分布 $L(\lambda)$, 以及与摄像机响应 $C(\lambda)$, 甚至场景反射率 $S(x, \lambda)$ 都是无法进行严格的定量描述。所以色彩的白平衡校正就成为一个病态问题。

自动白平衡的目标其实就是将未知光源下的图像调整到该场景在标准光源下的图像, 即

$$f: \mathfrak{R}^3 \rightarrow \mathfrak{R}^3$$

$$\rho^u(x) = \int_{\omega} S(x, \lambda) L^u(\lambda) C(\lambda) d\lambda \rightarrow \rho^c(x) = \int_{\omega} S(x, \lambda) L^c(\lambda) C(\lambda) d\lambda, \forall x \quad (6.8)$$

其中, $\rho^u(x)$ 表示未知光源下的图像; $\rho^c(x)$ 表示标准光源下的图像; $f \in \mathfrak{R}^3$ 为自动白平衡的调整校正方法。当假设摄像机满足窄带宽响应时, 可以近似地认为, 在同一光源下的 f 是常量^[23]。通常将 f 定义为对角矩阵 $A = \text{diag}(f)$, 即 Von Kries 对角模型。将未知光源下的像素调整到标准光源下只需乘以 A , 即

$$\rho^c(x) = A^{u,c} \rho^u(x) \quad (6.9)$$

所以自动白平衡最重要的求解校正矩阵 $A^{u,c}$ 与光照估计密切相关。在窄带宽响应假设下, 令光照颜色 $e = (\alpha, \beta, \gamma)$, 无色偏的场景中 x 点的 RGB 观测值为 (R_x, G_x, B_x) , 则加上光照颜色后, 该点的 RGB 值为 $(\alpha R_x, \beta G_x, \gamma B_x)$ 。令有色偏的场景中 x 点的 RGB 观测值为 (R'_x, G'_x, B'_x) , 则有 $(R'_x, G'_x, B'_x) = (\alpha R_x, \beta G_x, \gamma B_x)$ 。对于有色偏图像的校正, 需要将光照的偏差去掉, 所以进行色偏校正时需要在 x 点的 RGB 观测值上分别乘以相应的系数, 使得 R_x, G_x, B_x 的系数比值为 1。由于绿色分量在各种光照下一般变化不大, 所以一种常用的简单色偏校正方法是将绿色分量系数置为 1, 则校正后的 RGB 值为 $((\beta/\alpha)R'_x, G'_x, (\beta/\gamma)B'_x)$, 即

$$A^{u,c} = \begin{bmatrix} \beta/\alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \beta/\gamma \end{bmatrix} \quad (6.10)$$

此外, 还有一种比较常用的校正方法是:

$$A^{u,c} = \begin{bmatrix} \sqrt{\alpha^2 + \beta^2 + \gamma^2} / \sqrt{3}\alpha & 0 & 0 \\ 0 & \sqrt{\alpha^2 + \beta^2 + \gamma^2} / \sqrt{3}\beta & 0 \\ 0 & 0 & \sqrt{\alpha^2 + \beta^2 + \gamma^2} / \sqrt{3}\gamma \end{bmatrix} \quad (6.11)$$

那么如何解决光照估计这一病态问题呢? 目前, 主要有两类方法: 一是基于图像统计特征, 直接对场景添加约束或假设; 二是通过学习训练发掘场景和光源相关的先验知识, 对新的未知样本进行预测。下面分别对这两类方法进行介绍。

6.3.1 基于图像统计特征的方法

基于图像统计特征的方法通常较为简单, 易于实现。最简单常用的方法是灰世界 (Gray World) 法。

灰世界法假设认为: 场景中所有物理表面的平均反射是灰色的, 即可以取 RGB 三个通道的平均颜色作为光照估计, 即

$$\alpha : \beta : \gamma = \frac{\sum_{x=1}^N R_x}{N} : \frac{\sum_{x=1}^N G_x}{N} : \frac{\sum_{x=1}^N B_x}{N} \quad (6.12)$$

式中, N 为图像的像素数。

灰世界校正的效果如图 6-16 所示。实验原图来自 Gehler 等人^[24]提供的 Color-Checker 数据集。



(a) 未经白平衡校正

(b) 灰世界法校正

图 6-16 灰世界法自动白平衡效果

灰世界法特别适合反射面足够丰富, 细节很多的场景。然而当场景中出现大面积单色物体时, 如飘扬的红旗、绿色的草地、黄色的沙漠等, 灰世界方法将把单色物体错误地估计为环境色光。如图 6-16 所示, 天空由于红色墙面的影响被校正为偏蓝的色彩。为了提高灰世界方法的性能, 主要的改进方法有: ①将图像分块, 在分块图像上分别计算平均颜色^[25]; ②识别灰色面^[26]或灰色点^[27], 在灰色区域内计算平均颜色。显然, 分块图像只能在一定程度上抑制大面积单色物体影响, 在块内部对光照的估计仍然不准确, 而识别灰色面或灰色点的方法, 当场景中灰面或灰点数目很少甚至找不到时, 效果也不好。

白斑 (White Patch) 法是另一种常用的方法。白斑法假设认为: 一幅图像中, RGB 颜色通道的最大响应是由场景中的白色表面引起^[28]。因此, 白斑法直接取 RGB 三个颜色通道的最大值作为对光照的估计, 即

$$\alpha : \beta : \gamma = \max R_x : \max G_x : \max B_x \quad (6.13)$$

这是一种非常简单快速的方法。但是根据白斑假设, 场景中至少要存在一个标准白光下的像素点, 即三个颜色通道的最大反射率相等, 该算法才有较高的光照估计精度。而实际上, 该条件很难满足。

为了克服灰世界法和白斑法的问题, Weijer 等人^[29]提出了灰边缘 (Gray Edge) 法。灰边缘法假设认为: 场景中所有物理表面的平均反射的差分是无色差的。

为了得到更具有一般性的灰边缘算法引入, 闵科夫斯基范式——灰边缘算法, 同时为了减少噪声影响, 灰边缘算法还引入了高斯平滑操作, 并且将颜色导数推广到高阶, 得到了一个更为通用的白平衡计算框架:

$$\left(\int \left| \frac{\partial^n f^\sigma(x)}{\partial x^n} \right| dx \right)^{1/p} = k e^{n \cdot p \cdot \sigma} \quad (6.14)$$

式中， e 表示光照颜色 (α, β, γ) ，每个颜色通过分别计算即可得到 (α, β, γ) 值； n 指颜色导数的阶数； p 是闵科夫斯基范式参数； $f^\sigma(x) = f \otimes G^\sigma$ ，表示图像 f 与高斯滤波器 G^σ 的卷积。随着选择不同的 n 、 p 、 σ 参数，式 (6.14) 包含了灰世界法、白斑法等现有的一些自动白平衡方法，如表 6-2 所示。

表 6-2 统一框架下的不同白平衡方法

方法名	符号
灰世界法 (Gray World)	$e^{0,1,0}$
白斑法 (Max RGB)	$e^{0,\infty,0}$
高阶灰世界法 (Shades of Gray) [30]	$e^{0,p,0}$
通用灰世界法 (General Gray World)	$e^{0,p,\sigma}$
1 阶灰边缘法 (Gray Edge)	$e^{1,p,\sigma}$
最大灰边缘法 (Max Edge)	$e^{1,\infty,\sigma}$
2 阶灰边缘法 (2 nd Order Gray Edge)	$e^{2,p,\sigma}$

目前的研究表明灰边缘法假设能够有效克服大面积单色物体影响，且实验效果较好[29, 31~33]。然而灰边缘算法在取得较好效果时 (2 阶灰边缘) 必须做高斯平滑和计算高阶导数，这涉及图像的缓存与多次卷积等操作，会消耗较多的计算资源，不便于硬件实现。

根据灰边缘法假设，提出了一种能够在硬件上实现的简化灰边缘方法[34]。对于高斯平滑而言，其本质是做高频滤波，其他的高频滤波方法完全可以取代它。显然，最简单的滤波方法是均值滤波。考虑到图像传感器图像数据逐行传输的特点，如果做横向、纵向双向均值滤波，则必须缓存多行图像。为了满足实时性要求，只采用横向均值滤波。但是，由于不采用纵向滤波，滤波效果将下降。为了进一步滤除高频信息，考虑到图像降采样也是一种去高频方法，所以采用横向均值平滑降采样的滤波方法。

一般情况下，均值平滑后的图像需要除以相同的平滑模板大小 N ，将像素值归一化到 0~255 的范围 (像素深度为 8 位时)。但是，白平衡算法只需要得到最后光照颜色估计，并不需要得到完整的归一化平滑图像，因此为了简化算法，可以省去该除法操作。设 N 阶降采样滤波模板为

$$A = \underbrace{[1 \quad 1 \quad \cdots \quad 1]}_{N \uparrow 1} \tag{6.15}$$

则横向均值平滑降采样后的图像为

$$f^N(m', n) = \sum_{m=(N-1)m'}^{Nm'-1} f(m, n) \tag{6.16}$$

式中， $N \in [1, \text{width}(f(m, n))]$ ，即降采样模板应大于 1，而小于图像宽度。横向均值平滑降采样如图 6-17 所示。

若要求梯度导数，它需要在横向、纵向，甚至斜向求差分。同样地，为了不缓存图像

行，而又达到求图像梯度的效果，本章采取横向一阶差分的梯度求取方式。差分模板为

$$T=[1 \quad -1] \quad (6.17)$$

则差分图像为降采样图像与差分模板的卷积 $|f^N \otimes T|$ 。其中， f^N 由式 (6.16) 得到。

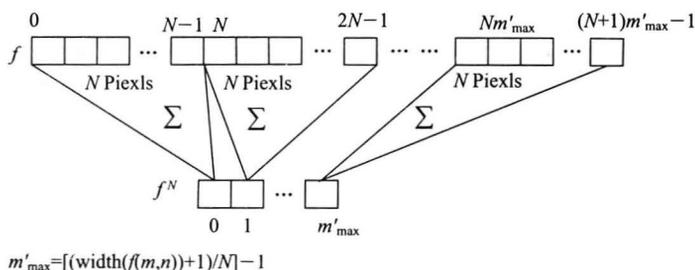


图 6-17 横向均值平滑降采样操作

对于闵科夫斯基范数 p 值，为了进一步减小计算复杂度，从而减少乘法与开方操作，所以将 p 值取 1，即不引入闵科夫斯基范数。

估计光照颜色时，一般要求差分图像的颜色均值。此时 RGB 三个通道都需要除以相同的图像像素数。事实上，如果不做除法，仅仅是光照颜色的 α, β, γ 值会被成比例线性放大，但它们之间的比例关系不会发生变化。同样为了简化算法，可以省去光照估计时的除法。最后，可以得出场景的光照估计为

$$\int f_r^N(x) dX = ke^N \quad (6.18)$$

式中， $f_r^N(x) = |f^N \otimes T|$ ； e 表示光照颜色 (α, β, γ)，每个颜色通过分别计算即可得到 (α, β, γ) 值。

在硬件实现方面，该简化算法有着天然的优势。下面以 FPGA 为例，做进一步说明。

在 FPGA 中，做加法与减法运算非常方便。对于乘法，虽然有专用的乘法器硬件，但是一片 FPGA 芯片，乘法器资源是有限的。对于除法，则没有专用的除法器硬件，只有通过替代算法实现，但其所占资源较多，而且时序控制也比较复杂。本书算法式 (6.18) 只需要做加法与减法，非常适合 FPGA 的硬件实现。此外，在算法流程中没有必要在得到整幅降采样图像后才进行卷积，可以将降采样与卷积进行流水线作业。

对于某颜色通道，算法流程设计如图 6-18 所示。其中，像素个数计数 cnt，用于计数均值平滑降采样所需的 N 个值；tmp_now 累加均值平滑降采样所需的 N 个像素的和；tmp_last 缓存上一次降采样的结果；gain 累加一帧差分图像的像素和作为光照估计值。算法主要分为两个步骤。

第一，等待传感器信号，做平滑累加。判断是否有行结束、场结束标志。如果一场图像结束，则一次光照估计流程结束。如果一行图像结束，则将计数器 cnt，累加器 tmp_now、tmp_last 清零，等待下一次信号到来。如果不是行结束或场结束（即是有效像素），则累加计数器 cnt 和像素和 tmp_now。

第二，做梯度差分。判断是否累加到了 N 个像素，如果是，则将当前累加和 tmp_now 和上一次累加和 tmp_last 做比较。取差分的绝对值累加到光照估计累加和 gain 上。然后保存此次累加和 tmp_now 的值到 tmp_last 中，再将计数器 cnt、累加器 tmp_now 清零，等待下一次信号到来。

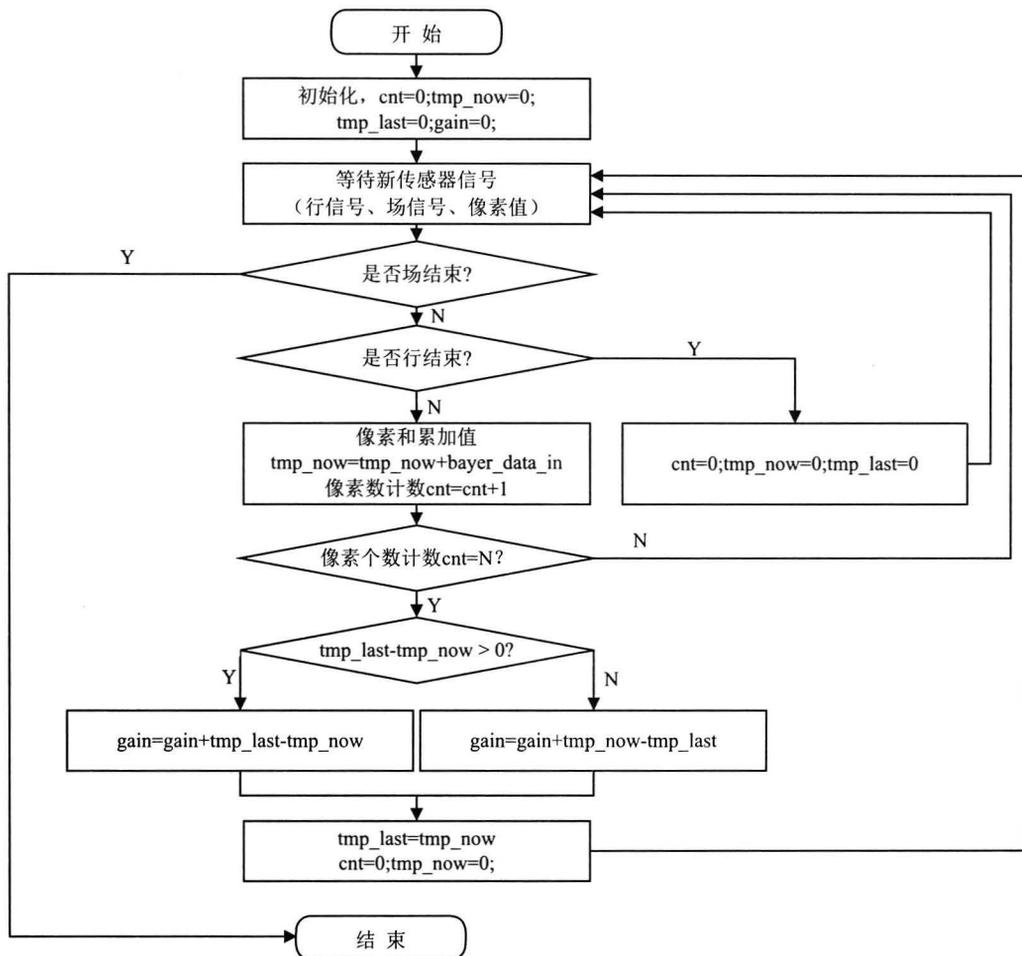


图 6-18 简化的灰边缘算法流程图

从资源消耗的角度看，RGB 三个通道需要一个计数器 cnt 、三个累加器 tmp_now 、三个缓存累加器 tmp_last 和三个光照估计累加器 $gain$ ，共 10 个寄存器即可。最后，在得到光照估计值后，利用式 (6.10) 做颜色校正，只需要做两次除法与两次乘法，而这也是无法避免的。

简单灰边缘算法实验结果如图 6-19 所示。



(a) 未经白平衡校正

(b) 灰世界法校正

(c) 上述简单灰边缘法校正

图 6-19 简单灰边缘算法自动白平衡效果对比

6.3.2 基于学习训练的方法

基于学习训练的方法需要通过训练集训练才能得到训练模型。基于学习训练的方法常常使用贝叶斯统计理论或神经网络^[35]等机器学习方法。虽然神经网络看上去能够合理地解决问题，但实际上它缺乏对颜色恒常性问题的深入描述，在实际应用中其泛化能力较差。

贝叶斯方法^[24, 36~38]期望找到某种概率分布作为先验知识。但不幸的是，这种分布通常难以准确找到。早期的 Bayesian 算法^[36]将其假定为高斯分布，然而，这并不符合现实情况，这也导致了算法的性能较差。虽然 Finlayson 等人^[37]引入相关颜色，解决了相关矩阵预计算的困难，但它只能针对预先定义的 11 种光源进行训练和估计，并不实用。Rosenberg 等人^[38]采用非高斯模型得到了较好的效果。但他也指出，虽然非高斯模型的 Bayesian 算法的性能平均要优于色域匹配 (Gamut Mapping) 算法，但很多的情况下，色域匹配算法仍要优于非高斯模型的 Bayesian 算法。

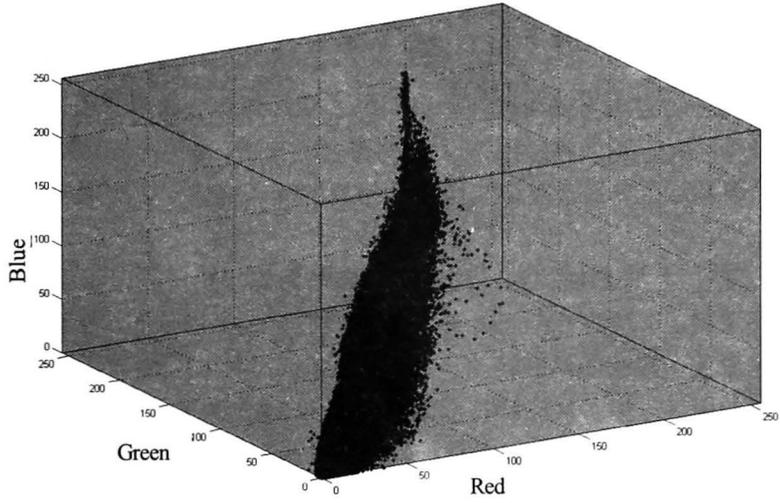
色域匹配^[39]由 Forsyth 提出。该方法指出不同条件下观察到的像素都有它们各自不同的特点。例如，暖光源照射下的场景普遍呈现红色或橘色，几乎观察不到蓝色，如图 6-20 所示。本小节的实验图库来自 B. Funt 等人提供的 Grey-Ball 数据集^[40]。



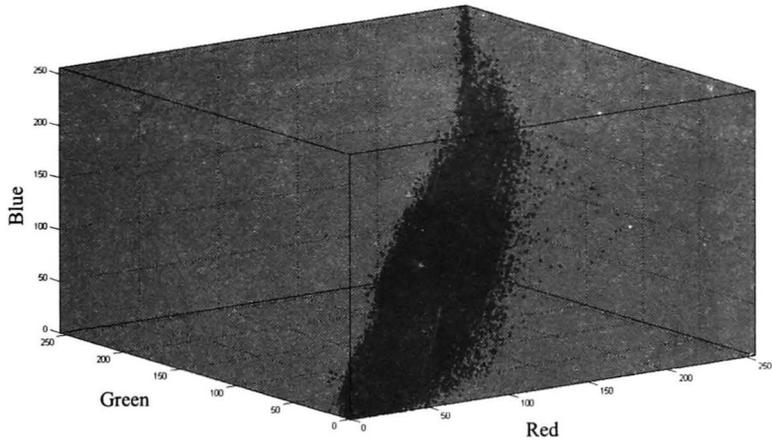
图 6-20 暖光源下的场景

色域匹配假设在一定光源下观察到的数据（即像素）应该只有有限数量的色彩，即整个色彩空间的一个子集，如图 6-21 所示。

定义在标准光源下观察到的颜色集合为标准色域。Forsyth 证明标准色域是一个凸壳 (Convex Hull)。如图 6-22 所示（见插页），图中绿色的点为标准色域下图像的色彩值，红线将最边界上的点连接了起来。色域匹配首先确定所有能够将观察图像色域变换到标准色域的映射，然后再根据训练结果选择出最佳映射。但是色域匹配只是简单地使用了凸集上的若干个较少的边缘点组成的凸壳，或者说只使用了凸壳上的那些点之间的联系，对于凸壳内部的数据并没有很好地利用起来。Gijssenij 等人^[41]基于以数据的灰边缘形式研究色域匹配，也取得了一定的研究成果。



(a) 冷光源下的色域



(b) 暖光源下的色域

图 6-21 不同光源下的色域

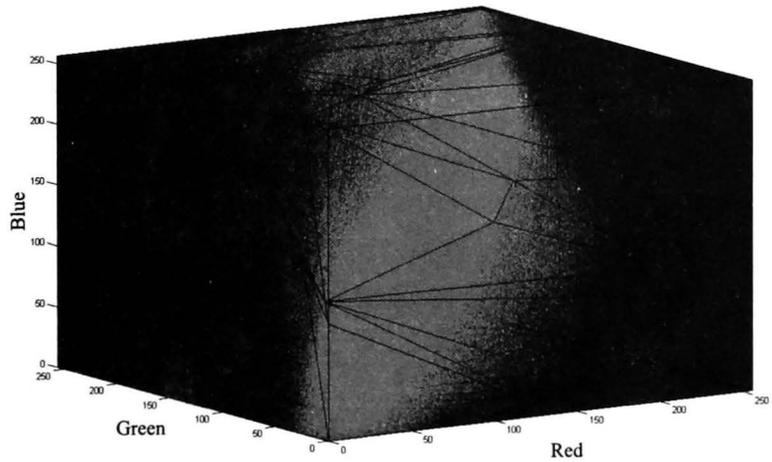


图 6-22 标准色域的凸壳

除了贝叶斯方法和色域匹配 (Gamut Mapping) 方法外, 还有基于支持向量机回归的方法^[42]、基于场景语义的方法^[43, 44]等。

基于色域数据与核方法, 提出了一种新的自动白平衡方法。从一个全新的视角审视观察到的色域数据, 提出了基于数据驱动的自动白平衡方法。通过分析图像的成像机理, 发现同一光源下的色域数据可以构成一个以该光源为法向量参数确定的超平面。这个超平面将看似毫无规律的色域数据联系起来, 找到了同一光源下色域的共同规律。基于此发现, 利用核方法, 设计类似 SVR 的数据源特征挖掘算法——SVF, 对标准光源下的色域数据进行训练, 隐式地求出标准光源对应的法向量。与色域匹配方法不同, 这种方法在训练时利用了整个色域数据, 而不仅是凸壳上的少数点。当给定一幅未知光源下的图像时, 通过最优化方法求出图像校正矩阵, 使得该光源下的图像经过校正矩阵调整后映射到高维空间, 能够最佳地符合标准光源超平面。由于该方法完全由光源照射下的图像数据决定, 不需要额外的其他信息, 因此称为基于数据驱动的自动白平衡方法。

根据 Fang-Hsuan 等人的研究^[45], 光谱功率分布函数 $L(\lambda)$ 可以表示为

$$L(\lambda) = \sum_{i=1}^n e_i E_i(\lambda) \quad (6.19)$$

式中, $E_i(\lambda)$ 是描述光谱功率分布的基函数; e_i 是对应的权重; n 是基函数的个数。将式 (6.19) 代入式 (6.6) 中可得

$$\begin{aligned} \rho(x) &= \int_{\omega} L(\lambda) S(x, \lambda) C(\lambda) d\lambda \\ &= \sum_{i=1}^n e_i \int_{\omega} E_i(\lambda) S(x, \lambda) C(\lambda) d\lambda \end{aligned} \quad (6.20)$$

将上式两边分别乘以 $\frac{1}{\rho(x)}$, 并移项, 可进一步得到

$$\left(-\sum_{i=1}^n e_i \int_{\omega} \frac{E_i(\lambda) S(x, \lambda)}{\rho(x)} dx \right) + 1 = 0 \quad (6.21)$$

通过观察式 (6.21) 发现, 它可以写为超平面的形式:

$$\langle w, \chi \rangle + 1 = 0 \quad (6.22)$$

式中, $\langle \cdot, \cdot \rangle$ 表示内积; $w = [e_1, e_2, \dots, e_n]^T$; $\chi = \left[\int_{\omega} \frac{-E_i(\lambda) S(x, \lambda) C(\lambda)}{\rho(x)} d\lambda \right]_i^n = 1$ 。因

此, 如果能够找到映射到高维空间 H 的函数 $\Phi: \mathcal{H}^3 \rightarrow H$, 使其满足

$$\Phi(\rho(x)) = \left[\int_{\omega} \frac{-E(\lambda) S(x, \lambda) C(\lambda)}{\rho(x)} d\lambda \right]_{i=1}^n, \forall x \quad (6.23)$$

则图像中的任何一个点经函数 Φ 都可以投影到由法向量 w 确定的超平面上。而法向量 w 仅与光源参数 e_i 相关。当然, 正如前面所述, 由于物体光谱反射函数 $S(x, \lambda)$ 以及摄影机的光谱响应函数 $C(\lambda)$ 都无法严格定量测定, 所以要直接求出映射函数 Φ 是很困难, 甚至几乎是不可能的。因此, 只有通过近似的方法表示出该映射。

另一方面, 为了减少训练样本的数量, 需要去除冗余的训练样本, 因此考虑是否能将整个图像缩小到图像的色域范围, 即去除场景空间位置因素。观察式 (6.23) 可以发

现, $\int_{\omega} E_i(\lambda)S(x, \lambda)C(\lambda)d\lambda$ 可以看成是摄像机在光谱功率分布基函数 E_i 的特殊光源照射下成的像, 因此式 (6.23) 可变换为

$$\begin{aligned} \Phi(\rho(x)) &= \left[\int_{\omega} \frac{-E(\lambda)S(x, \lambda)C(\lambda)}{\rho(x)} d\lambda \right]_{i=1}^n \\ &= \left[\frac{-1}{\rho(x)} \int_{\omega} E(\lambda)S(x, \lambda)C(\lambda)d\lambda \right]_{i=1}^n \\ &= \left[\frac{-\rho^{E_i}(x)}{\rho(x)} \right]_{i=1}^n \end{aligned} \quad (6.24)$$

式中, $\rho_{E_i}(x)$ 为光谱功率分布基函数 E_i 的特殊光源照射下的 RGB 值。类似于 $\rho(x) = [\rho_r(x), \rho_g(x), \rho_b(x)]^T$, $\rho^{E_i}(x) = [\rho_r^{E_i}(x), \rho_g^{E_i}(x), \rho_b^{E_i}(x)]^T$ 。通过式 (6.24), 可知只需要对图像中像素点对应的 RGB 像素值做超平面高维映射即可, 与具体的位置并无关系。因此, 可以只在光源色域数据上进行分析。

现在通过新的数据源特征挖掘算法核方法 SVF (Support Vector Fit) 对光源特征进行估计。与 SVM、SVR 不同的是, 这两者都是监督学习算法, 需要知道特征数据对应的输出, 而 SVF 是无监督学习算法。SVF 旨在发现观察数据源内在的特征。由前所述, 对光源的特征参数的估计可以转化为将光源照射下的色域数据映射超平面后, 对该超平面法向量的估计。

假设已知光源 c 的色域为 $\Omega^c = \{\rho^c(i)\}_{i=1,2,\dots,P}$, 其中 P 为色域中点的个数, $\rho^c(i)$ 为色域中某点对应的 RGB 值。经过映射函数 Φ 后的色域数据 $\chi_i^c = \Phi(\rho^c(i))$ 应该尽可能地满足该光源对应的超平面 Γ^c :

$$\Gamma^c: \langle w^c, \chi_i^c \rangle + 1 = 0 \quad (6.25)$$

式中, w^c 为光源参数, 即超平面法向量。

由于直接求出映射函数 Φ 几乎不可能, 而另一方面核函数可以在原空间表示向量在映射到高维空间后的内积, 即 $K(a, b) = \langle \Phi(a), \Phi(b) \rangle$, 所以采用核函数对 Φ 进行逼近。因此在式 (6.25) 中, 要完成的任务是求解最优的 w^c , 使得式 (6.25) 能最大可能地成立。

类似于 SVR 的推导, 定义损失函数 $\zeta = L(0, \langle w^c, \chi^c \rangle + 1)$, 则目标是使总的损失最小, 即

$$w^c = \arg \min_{w^c} \sum_{i=1}^P \zeta_i \quad (6.26)$$

满足

$$\zeta_i = L(0, \langle w^c, \chi_i^c \rangle + 1)$$

类似于 SVM, 对于观测数据中的噪声, 选择线性 ε 不敏感函数来处理:

$$L(0, \langle w^c, \chi^c \rangle + 1) = |\langle w^c, \chi^c \rangle + 1|_{\varepsilon} \quad (6.27)$$

此时, 该问题转化为求优化目标函数的最小化问题:

$$w^c = \arg \min_{w^c} \frac{1}{2} \|w^c\|^2 + \lambda \sum_{i=1}^P (\zeta_i + \hat{\zeta}_i) \quad (6.28)$$

$$\text{满足} \quad \begin{cases} \langle w^c, \chi_i^c \rangle + 1 \leq \varepsilon + \zeta_i \\ \langle w^c, \chi_i^c \rangle + 1 \geq -\varepsilon - \hat{\zeta}_i \\ \zeta_i, \hat{\zeta}_i \geq 0 \end{cases} \quad i=1, 2, \dots, p$$

式中, $\|\cdot\|$ 表示二范数 (2-norm); ζ_i 、 $\hat{\zeta}_i$ 为松弛因子; 常数 $\lambda > 0$ 表示对超出误差 ε 的惩罚程度。

通过 Lagrange 对偶理论, 引入核函数, 可以将式 (6.28) 转换为如下二次规划问题:

$$W(\hat{\alpha}, \alpha) = \arg \min_{\hat{\alpha}, \alpha} \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P (\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) K(\rho^c(i), \rho^c(j))$$

$$\varepsilon \sum_{i=1}^P (\hat{\alpha}_i + \alpha_i) + \sum_{i=1}^P (\hat{\alpha}_i - \alpha_i) \quad (6.29)$$

$$\text{满足} \quad \begin{cases} 0 \leq \hat{\alpha}_i \leq \lambda \\ 0 \leq \alpha_i \leq \lambda \end{cases} \quad i=1, 2, \dots, P$$

式中, $K(\cdot, \cdot)$ 为核函数, 本章选用高斯径向基核函数 $K(a, b) = e^{-\frac{\|a-b\|^2}{\sigma^2}}$, 且由推导式 (6.29) 的过程, 可以得出

$$w^c = \sum_{i=1}^P (\hat{\alpha}_i - \alpha_i) \Phi(\rho^c(i)) \quad (6.30)$$

即通过求解式 (6.29) 可以得到参数 $\hat{\alpha}_i$ 、 α_i , 它们可以隐式地表示出光源 w^c 。虽然 w^c 还与映射 $\Phi(\rho^c(i))$ 相关, 但是后面可以推导出已知 $\hat{\alpha}_i$ 、 α_i , 就已经可以求出另一未知光源下的校正矩阵了。

此外, 根据支持向量的特点, 由于采用了不敏感函数, 求得的 w^c 将是非常稀疏的。即是说, 虽然 w^c 是 P 维 (P 代表色域中点的个数), 但是实际求得的 w^c 中存在大量的 0 值, 不为 0 的维数, 正是支持向量的个数。区别于监督学习的 SVR, 将这种数据源特征挖掘算法称为 SVF (Support Vector Fit)。

通过 SVF 算法, 对标准光源 c 下的图像色域数据 $\rho^c(i)$ 进行训练, 可以得到标准光源 c 的特征 w^c 。当有一幅未知光源 u 下的图像 $\rho^u(x)$ 时, 由式 (6.9) 可知, 可以通过对角矩阵 $A^{u,c} = \text{diag}(f) \in \mathcal{R}^{3 \times 3}$ 对 $\rho^u(x)$ 进行校正。现在, 可以将问题描述为使用校正矩阵 $A^{u,c}$ 调整后的图像色域映射到高维空间后, 能够最好地服从标准光源下的超平面 Γ^c , 即转化为如下的最优化问题:

$$A^{u,c} = \arg \min_{A^{u,c}} \sum_{j=1}^Q \frac{(\langle w^c, \Phi(A^{u,c}, \rho^u(j)) \rangle + 1)^2}{\|w^c\|^2} \quad (6.31)$$

式中, $\rho^u(j)$ 为未知光源图像的色域中某点 j 的 RGB 值; Q 为该色域点的个数。式 (6.31) 中, 当 $A^{u,c} = 0$ 时, 显然成立, 等式退化。因此, 在求解过程中, 需要加入约束条件, 确保优化过程的非退化。本章加入的约束为 $A^{u,cT} A^{u,c} \geq 3$, 即校正矩阵系数的平方和大于 3。其物理意义为图像经过校正后, 保证亮度的正常, 不会变得过暗 ($A^{u,c} = 0$ 时, 退化为全黑)。此外, w^c 在 3.2 节的标准光源训练中已经求出, 并由式 (6.30) 表示。由该公式可知, $\|w^c\|^2$ 为常数, 且 $w^c = \sum_{i=1}^P (\hat{\alpha}_i - \alpha_i) \chi_i$, 并且由高斯径向基核函数的性质,

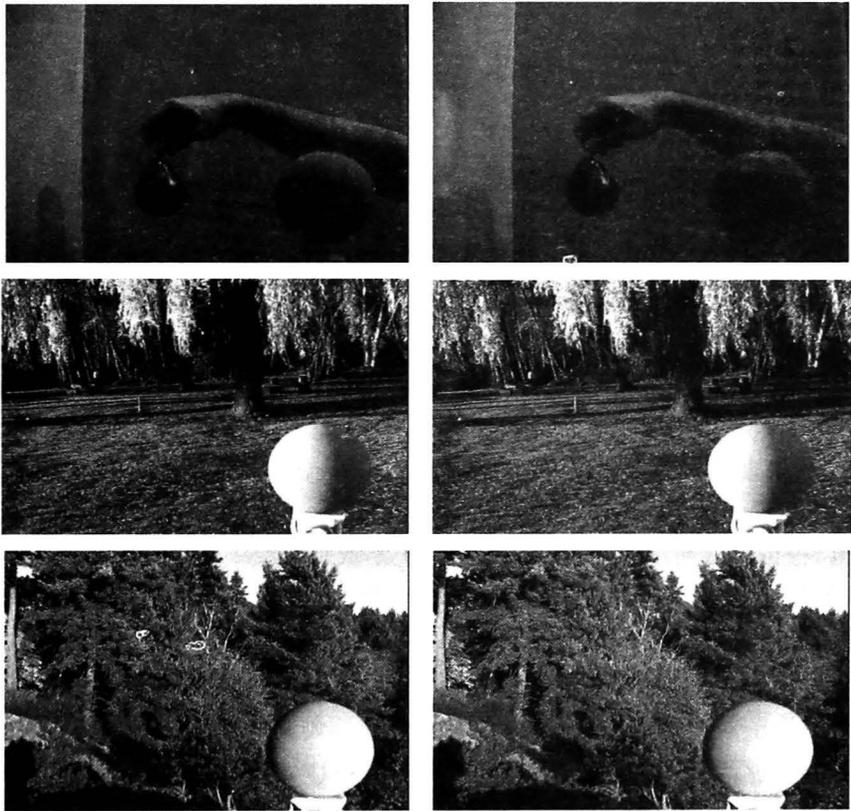
式 (6.31) 可写为

$$A^{u,c} = \arg \min_{A^{u,c}} \sum_{j=1}^Q \sum_{i=1}^P (\hat{\alpha}_i - \alpha_i) \left(e^{-\frac{(A^{u,c} \rho^u(j) - \rho^c(i))}{2\sigma^2}} + 1 \right) \quad (6.32)$$

满足

$$A^{u,cT} A^{u,c} \geq 3$$

该基于数据驱动自动白平衡算法效果如图 6-23 所示。图中圆球为一个非彩色无色偏的灰色球，将它连接到拍摄的摄像机上，用来记录场景真实的光照信息。从图中可以看出，经过校正后色偏有了明显改善。



(a) 原始图像

(b) 校正后的图像

图 6-23 本章基于数据驱动的自动白平衡效果

6.4 总 结

本章介绍了自动曝光与自动白平衡技术。曝光与白平衡是摄像机中两项最基本的设置功能。曝光决定了图像的明暗程度，白平衡决定了图像的偏色程度。要调节好它们通常需要具备一定的摄影知识以及丰富的摄影经验。这对于普通用户的使用非常不便。自动曝光与自动白平衡技术解决了这一问题。它们不仅可以方便用户的使用，而且在摄像过程中，能够面对变化的场景进行自动调节。

自动曝光是否准确,最重要的是对环境光的测量,估计出光照强度。目前常用的测光方式包括中央重点平均测光、中央部分测光、点测光等。此外,面对逆光或强正面光的高对比度场景,带曝光补偿功能的测光方式正在被研究使用。其中,主要的技术包括分区加权测光、基于亮度直方图的测光等。自动白平衡是否准确,最重要的是对环境光色温的估计。自动白平衡方法主要分为两大类。第一类是基于图像统计特征的方法,包括灰世界法、白斑法、灰边缘法等。这类方法效果明显、计算简单,是目前摄像机中常用的方法。第二类是基于学习训练的方法。这类方法需要通过训练集学习得到关于光照的先验信息,然后再对未知光源下的色光进行估计,主要包括基于贝叶斯的方法、色域匹配(Gamut Mapping)法以及本书提出的基于数据驱动的方法。这类方法计算复杂,但是准确度相对较高。

未来,自动曝光与自动白平衡技术的发展主要还是在于提高处理复杂场景的能力。例如,逆光、强正面光,甚至多个强光源、多色温光源、极端色温光源等。其次,就是研究算法的硬件化、实时化问题,因为大多数效果更好的算法会伴随着较大的计算复杂度。

参 考 文 献

- [1] EBNER M. Color constancy [M]. Chichester: John Wiley & Sons, 2007.
- [2] HARUKI T, KIKUCHI K. Video camera system using fuzzy logic [J]. IEEE Transactions on Consumer Electronics, 1992, 38(3): 624-634.
- [3] KUNO T, SNGIURA H, MATOBA N. A new automatic exposure system for digital still cameras [J]. IEEE Transactions on Consumer Electronics, 1998, 44(1): 192-199.
- [4] LEE J, JUNG Y, KIM B. An advanced video camera system with robust AF, AE and AWB control [J]. IEEE Transactions on Consumer Electronics, 2001, 47(3): 694-699.
- [5] SAMPAT N, VENCATARAMAN S, YEH T. System implications of implementing auto-exposure on consumer digital cameras [C] // SPIE Conference on Sensors, cameras, and Applications for digital photography, San Jose, January 23, 1999. Bellingham: SPIE, c1999: 100-107.
- [6] 梁佳毅, 洪志良. 适用于大动态范围场景的自动曝光控制算法[J]. 光电工程, 2008, 35(5): 89-92.
- [7] TAO J, KUHNERT K, NGUYEN D. Multiple templates auto exposure control based on luminance histogram for onboard camera [C]// IEEE International Conference on Computer Science and Automation Engineering, Shanghai, June 10-12, 2011. New York: IEEE, c2011: 89-92.
- [8] 杨海涛, 常义林, 王静, 等. 一种基于亮度直方图的自动曝光控制方法[J]. 光学学报, 2007, 27(5): 841-847.
- [9] 倪国瑜, 黄昶. 基于亮度直方图宽度的自动曝光方法研究[J]. 光学仪器, 2009, 31(6): 21-24.
- [10] 谷明琴, 蔡自兴, 易亮. 面向无人驾驶车感知系统的自动曝光方法[J]. 计算机科学, 2013, 40(6): 300-302.
- [11] CAI Z, GU M. Traffic sign recognition algorithm based on shape signature and dual tree-complex wavelet transform [J]. Journal of Central South University of Technology: English Edition, 2013, 20(2): 433-439.
- [12] SHIMIZU S, KONDO T, KOHASHI T. A new algorithm for exposure control based on fuzzy logic for video cameras [J]. IEEE Transactions on Consumer Electronics, 1992, 38(3): 617-623.
- [13] MURAKAMI M, HONDA N. An exposure control system of video cameras based on fuzzy logic using color information [C]// the Fifth IEEE International Conference on Fuzzy System, New Orleans, September 08-11, 1996. New York: IEEE, c1996: 2181-2187.
- [14] STARK J. Adaptive image contrast enhancement using generalizations of histogram equalization [J]. IEEE Transactions on

- Image Processing, 2000, 9(5): 889-896.
- [15] SU M, GUO J, LIN D. New compensation algorithm for color backlight images[C]// 2002 International Joint Conference on Neural Networks, Honolulu, May 12-17, 2002. New York: IEEE, c2002: 1396-1400.
- [16] NEUMANN L, MATKOVIC K, PURGATHOFER W. Automatic exposure in computer graphics based on the minimum information loss principle[C]// International Conference on Computer Graphics, Hannover, June 22-26, 1998. New York: IEEE, c1998: 666-677.
- [17] MATKOVIC K, NEUMANN L. Interactive calibration of the mapping of global illumination values to display devices[C]// Twelfth Spring Conference on Computer Graphics, Bratislava and Budmerice, June 5-7, 1996. Bratislava: Comenius University, c1996.
- [18] CHEN C, TSENG C, HUNG C, et al. Combinational AE-AF system with fuzzy climbing search servo [C]// IS&T/SPIE 18th Annual Symposium Electronic Imaging, San Jose, January 15-19, 2006. Bellingham: SPIE, c2006: 60690B.
- [19] NASSER K, OH H, SHIDATE I, et al. New approach to auto-white-balancing and auto-exposure for digital still cameras [C]// SPIE Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III, San Jose, January 19, 2002. Bellingham: SPIE, c2002: 268-276.
- [20] 徐培风, 李正明, 孙俊. 基于图像的自动曝光算法研究[J]. 光学仪器, 2005, 27(2): 59-61.
- [21] FORSYTH D A, PONCE J. Computer Vision: A modern Approach [M]. New Jersey: Prentice Hall, 2003.
- [22] BARNARD K. Practical colour constancy [D]. Burnaby: Simon Fraser University, 1999.
- [23] FINLAYSON G, DREW M, FUNT B. Diagonal transforms suffice for color constancy [C]// IEEE Conference on Computer Vision, Berlin, May 11-14, 1993. New York: IEEE, c1993: 167-171.
- [24] GEHLER P, ROTHER C, BLAKE A, et al. Bayesian color constancy revisited [C]// IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, June 23-28, 2008. New York: IEEE, c2008: 1-8.
- [25] BARNARD K, MARTIN L, COATH A, et al. A comparison of computational color constancy algorithms; part ii: Experiments with image data [J]. IEEE Transactions on Image Processing, 2002, 11(9): 985-996.
- [26] LI B, XU D, XIONG W, et al. Color constancy using achromatic surface [J]. Color Research and Application, 2010, 35(4): 304-312.
- [27] HUO J, CHANG Y, WANG J, et al. Robust automatic white balance algorithm using gray color points in images [J]. IEEE Transactions on Consumer Electronics, 2006, 52(2): 541-546.
- [28] LAND E. The retinex theory of color vision [J]. Scientific American, 1977, 237(6): 108-128.
- [29] WEIJER J, GEVERS T, GIJSENIJ A. Edge-based color constancy [J]. IEEE Transactions on Image Processing, 2007, 16(9): 2207-2214.
- [30] FINLAYSON G, TREZZI E. Shades of gray and colour constancy [C]// IS&T/SID 12th Color Imaging Conference, Scottsdale, November 2004. Springfield: IS&T, c2004: 37-41.
- [31] CHEN H, SHEN C, TSAI P. Edge-based automatic white balancing with linear illuminant constraint [C]// SPIE Conference on Visual Communications and Image Processing, San Jose, January 29, 2007. Bellingham: SPIE, c2007.
- [32] GIJSENIJ A, GEVERS T, WEIJER J. Physics-based edge evaluation for improved color constancy [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, June 20-25, 2009. New York: IEEE, c2009: 1-8.
- [33] GIJSENIJ A, GEVERS T, WEIJER J. Improving Color Constancy by Photometric Edge Weighting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(5): 918-929.
- [34] TAN X, LAI S-M, WANG B, et al. A simple gray-edge automatic white balance method with FPGA implementation [J]. Journal of real-time image processing, 2013.
- [35] CARDEI V, FUNT B, BARNARD K. Estimating the scene illumination chromaticity using a neural network [J]. Journal of Optical Society of American, 2002, 19(12): 2374-2386.
- [36] BRAINARD D, FREEMAN W. Bayesian color constancy [J]. Journal of Optical Society of America A., 1997, 14(7):

- 1393-1411.
- [37] FINLAYSON G, HORDLEY S, HUBEL P. Color by correlation: a simple, unifying framework for color constancy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1209-1221.
 - [38] ROSENBERG C, MINKA T, LADSARIYA A. Bayesian color constancy with non-Gaussian models [C]// the Seventeenth Annual Conference on Neural Information Processing Systems, Vancouver, December 8-13, 2003. Cambridge: MIT Press, c2004.
 - [39] FORSYTH D. A novel algorithm for color constancy [J]. International Journal of Computer Vision, 1990, 5: 5-36.
 - [40] CIUREA F, FUNT B. A large images database for color constancy research [C]// IS&T/SID Eleventh Color Imaging Conference, Scottsdale, November 2003. Springfield: IS&T, c2003: 5-36.
 - [41] GIJSENIJ A, GEVERS T, WEIJER J. Generalized gamut mapping using image derivative structures for color constancy [J]. International Journal of Computer Vision, 2010, 86(2-3): 127-139.
 - [42] WANG N, XU D, LI B. Edge-based color constancy via support vector regression [J]. IEICE Transactions on Information and Systems, 2009, E92-D(11): 2279-2282.
 - [43] GIJSENIJ A, GEVERS T. Color constancy using natural image statistics and scene semantics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(4): 687-698.
 - [44] BIANCO S, CIOCCA G, CUSANO C, et al. Improving color constancy using indoor-outdoor image classification [J]. IEEE Transactions on Image Processing, 2008, 17(12): 2381-2392.
 - [45] CHENG F, HSU W, CHEN T. Recovering colors in an image with chromatic illuminant [J]. IEEE Transactions on Image Processing, 1998, 7(11): 1524-1533.

第 7 章 高动态范围成像技术

7.1 概 述

现实世界中的场景丰富多彩、绚丽多姿，这些场景有着非常高的动态范围和对比度。场景的动态范围被定义为场景中最亮部分的亮度值与最暗部分的亮度值之比。传统的显示设备，如 CRT 显示器因为所能显示的动态范围（在第 4 章已经做了定义）非常有限，大约为 10^2 数量级，因此无法真实再现现实的高动态范围（High Dynamic Range, HDR）场景。然而，由于人眼特有的视觉自适应机能，人们才能够同时感知高动态范围场景的不同部分。数字图像随着科技的发展在人们的日常生活中变得越来越重要，对高质量数字图像的需求推动着数字成像技术的飞速发展。特别是随着高动态范围成像技术（High Dynamic Range Imaging, HDRI）的出现，使在传统的低动态范围显示设备上再现现实场景成为可能。HDRI 是一种可以表示现实场景高动态范围亮度信息的成像技术，通过这种技术可以准确地记录现实场景的亮度信息，并能在传统的显示设备上显示出来。近年来在数字摄影、电影特技、计算机游戏、人类视觉系统的模拟、卫星遥感图像、运动模糊、交互 3D 应用等领域得到了广泛的应用。

虽然高动态范围 HDR 图像可以通过硬件方式直接获得，但是成本非常昂贵，并且传统的显示设备只支持低动态范围图像的输出，所以在其应用和推广方面遇到了挑战，而这些挑战也正是高动态范围研究的主要领域。HDRI 技术的研究领域主要有：高动态图像的获取、高动态范围图像的合成技术和色调映射技术。通过多曝光度图像序列的对齐技术和高动态范围图像的合成技术，可以通过合成同一场景中不同曝光度的图像序列来获得现实场景的照度图。

如果能将 HDR 图像在显示设备上显示出来，就可以给人们呈现原始场景的真实感受。有两种方式来显示 HDR 图像：一种是使用新的具有高动态范围显示能力的显示系统，在这种情况下，HDR 图像的照度图可以直接在设备上再现出来，而不使用任何压缩技术。另外一种方式则是通过色调映射技术在传统的显示设备，如 CRT 显示器上显示高动态范围图像，比如具有大约两个数量级动态范围的 CRT 显示器、打印机。色调映射算法在 HDR 图像的显示过程中起着决定性的作用，目前色调映射技术分为两类，即全局色调映射技术和局域色调映射技术。通过 HDR 图像的色调映射技术可以把照度图显示在传统低动态范围显示设备上，真实再现现实场景。

本章在国内外学者的研究基础上，简要介绍 HDRI 技术的各个关键步骤，并重点介绍了 HDR 图像的显示技术——色调映射技术，色调映射算法也正是 HDRI 技术研究的重点。

7.2 HDRI 技术的处理机制和步骤

HDR 图像是指可以表示现实世界场景中高动态范围亮度信息的图像,其像素亮度值正比于场景中对应点的实际亮度值,而传统数字图像中的像素亮度值代表的仅仅是显示的色彩值,与实际场景的亮度不存在线性的对应关系。因此可以说高动态范围图像是“场景相关”的,而传统数字图像是“输出相关”的。正是由于这一特性,与传统的数字图像相比,高动态范围图像所能表示的层次更丰富,色彩空间更高,因此现实场景中的亮度区域和黑暗区域的细节信息都能很好地保留下来。类似于计算机图形学的渲染管线,高动态范围成像技术的每个算法处理步骤清晰并且独立,其处理流程如图 7-1 所示。



图 7-1 高动态范围成像技术处理流程

7.2.1 HDR 图像的获取简介

尽管现成的数码照相机能直接采集到照度图,并且输出 12 位、16 位或者 32 位的数据,但是这种方式需要高端数码相机才得以实现,成本相对比较昂贵。所以一般使用另外一种低成本的方式,即使用一组在不同曝光度下拍摄同一场景的低动态范围的图像合成得到。在图 7-2 中展示了同一场景中不同曝光度下所拍摄得到的图像序列。在低曝光度的图 7-2 (a) 中,尽管可以清晰地看到阳光照射进来的部分,但是由于曝光度比较低,在较暗的区域丢失了很多细节信息而没法辨认。在图 7-2 (d) 中,由于过高的曝光度,虽然得到了大量较暗部分的细节信息,但是在亮的部分由于过于饱和也丢失了很多信息。所以可以通过合成这一组在不同曝光度下拍摄的图像获得高动态范围的照度图。通过这种方式合成的照度图基本上还原了原始场景的亮度信息。并且通过合成多张不同曝光图像序列的方式获得照度图是目前比较常用的方式。对于如何通过多曝光图像序列获得照度图以及中间涉及的相关技术,本章将在 7.3 节进行比较详细的描述。



图 7-2 原场景四张不同曝光度图像序列^[1]

7.2.2 全亮度图的合成简介

采集多张不同曝光的图像序列以后,这些图像序列包含了图像暗处和亮处的信息,下一步工作就是将这一系列图像合成为一张全亮度的 HDR 图像。HDR 的图像数据英文

为 Radiance Map, 被译为照度图或光谱图, 照度图的存储一般采用 32 位的浮点数表示, 使用 32 位的浮点数能够精确记录现实场景的实际亮度范围或者相对亮度范围。通常情况下, 照度图可以通过合成多张具有不同曝光度的普通数字图像而得到, 文献[2, 3]中分别介绍了合成照度图的不同方法。

另一个需要提到的是, 照度图的存储对象是浮点型数据, 所以无法使用传统的数字图像存储格式来进行存储。HDR 图像的存储指标包括大范围、高精度、占用低存储空间等。目前常用的照度图存储格式有 IEEE TIFF 格式、RGBE 格式、SGILogLuvTIFF 格式和 openExR 格式。在实际的应用过程中, 由于实际需要, 很多公司都对其做出了重要贡献。在 7.4 节中将对这几种格式进行简要的介绍和对比。

7.2.3 色调映射技术简介

尽管照度图能够精确记录现实场景的亮度信息, 但是照度图一般用 32 位的浮点数表示, 是不能直接在传统的显示设备, 如打印机、CRT 显示器等上直接显示的。要在普通的显示设备上显示照度图, 再现真实的原始场景, 必须把 32 位浮点数格式照度图映射成 8 位格式的普通图像, 这一过程称为色调映射 (Tone Mapping) 或色阶重建 (Tone Rep Reduction), 本章统一称为色调映射。现实场景能否真实再现, 色调映射算法起着决定性作用, 这也是本章研究的重点内容, 如图 7-2 所示的色调映射结果如图 7-3 所示。关于色调映射技术的详细研究状况和算法细节, 将在 7.5 节中作详细介绍。



图 7-3 高动态范围图像再现图 7-2 中原始场景结果^[1]

7.3 HDR 图像的获取

传统数码相机在采集场景的图像数据时用二维的亮度值来表示图像, 并且在每一个通道上只可能有 256 个不同的整数值。由于在成像过程中的量化和非线性的映射, 原始场景的大量信息被丢失。数码相机中的非线性映射决定了如何将不同曝光度 X 转换成与之对应的图像亮度值, 其中曝光度 X 和式 (7.1)、式 (7.2) 所决定的辐照度相关。在成像过程中被丢失的图像信息非常困难或者根本不可能从单张 8 位的图像数据中得到恢复。另外, 传统数字图像的像素亮度值很少是现实场景中相关辐照度的真实度量, 这也

就意味着不同的像素亮度值并不一定对应着原始场景中的两个不同的照度值,这种现象正是传统数码相机在辐照度和像素亮度值之间的非线性关系造成的。

一幅照度图是由浮点数构成的二维数组,数组中的每一个像素代表着原场景中绝对的或者相对的辐射值。和传统的8位数字图像相比,照度图有着能够精确记录原场景信息的优势。这是因为辐照度的巨大变化可以用浮点型数据表示出来。在场景中,照度图可以被认为原场景的一个虚拟表示,即记录的是人眼所看到的,而不是显示设备所能显示的。

当采集或者记录一个场景时,可以选择使用传统的能够产生8位数字图像的照相机,但是采集来的数据质量完全依赖于照相机和成像技术。另外,也可以选择使用照度图来记录场景,也就意味着原场景中的信息将被准确地记录下来。使用照度图来记录原始场景,使得原场景通过一定的技术可以再现成8位的普通数字图像,并在传统显示设备上显示,这种方式是通过使用相机给定曝光度的反射函数或者现有的色调映射算法来拍摄的一张虚拟的照片来实现的。当然,也可以使用更高质量的数码相机或者色调映射算法来再现照度图。显然,照度图提供了很多的灵活性和可能性去再现原场景的影像,而所得到的结果也是很令人满意的。

尽管现成的数码照相机能直接采集到高动态范围景色,并且输出12位、16位或者32位的数据,但是这种方式却很昂贵。另外一种低成本的方式是使用一组在不同曝光度下拍摄同一场景的低动态范围的图像序列来获得。在高曝光度的照片中,尽管由于高饱和度和在亮的区域会丢失很多的信息,但是在暗的区域的信息则可以很好地保留下来。同理,在低曝光度图像中,则可以更好地保留下明亮区域的信息。因此可能通过一组相互补充的低动态范围图像序列合成与原始场景直接相关的照度图。其过程如图7-4所示。

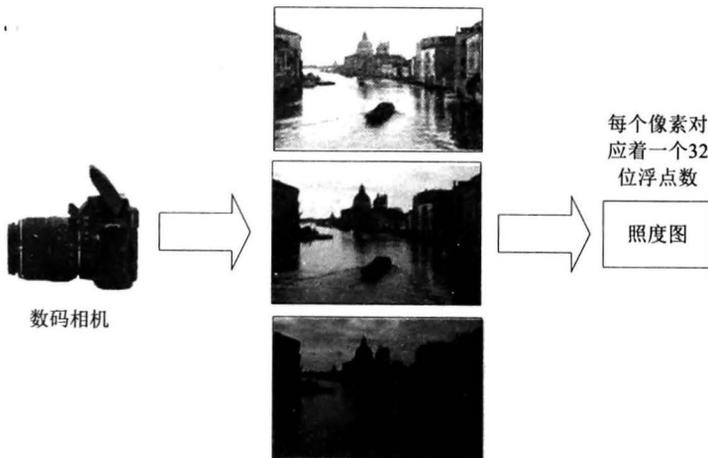


图7-4 通过拍摄不同曝光度的图像序列合成高动态范围图像的过程

在一个典型的图像成像系统,比如相机中,辐照度 E 和场景亮度 L 的关系如文献[3]中所述:

$$E = L \frac{\pi}{4} \left(\frac{d}{h} \right)^2 \cos^4 \varphi \quad (7.1)$$

式中, h 是镜头的焦距; d 是镜头的口径; φ 是主射线到主光轴的夹角。因此,恢复辐照

度 E 也就相当于恢复场景亮度 L ，曝光度 X 和辐照度 E 关系如下：

$$X = E \cdot \Delta t \quad (7.2)$$

式中， Δt 是曝光时间。在典型的成像系统中，由 CCD 产生的电子信号和曝光度成线性相关关系。但是从 CCD 的输出到最终的数字值却是非线性关系，这主要是 Gamma 映射、A/D 转换、感光噪声等造成的。曝光度到最终的数字值 Z 之间的关系如式 (7.3) 所示，这种函数关系被称为照相机的响应函数 f ：

$$Z = f(X) = f(E \cdot \Delta t) \quad (7.3)$$

照相机的响应函数通常假定为单调的，因此数字值 Z 到曝光度 X 的关系可以表示出 f 的反函数：

$$X = f^{-1}(Z) = g(Z) \quad (7.4)$$

一旦得到相机的响应函数，就可以通过一系列多曝光的图像序列来估算原始场景的照度图。然而，尽管人们可以使用多曝光度图像序列来获得照度图，但因为多曝光度图像在拍摄过程中可能有晃动等人为因素，使得多张图合成时也出现鬼影、伪色彩等现象，所以需要相应的技术，比如图像序列的配准技术来解决这些问题。在 7.4 节中将分别介绍照度图的合成技术和与之相关的多曝光度图像序列的配准技术。当然也有一些不用做配准的宽动态图像的采集技术，比如控制像素级曝光的 HDR 图像采集方式，以及通过分光装置在不同 CCD 上采用不同曝光设置的 HDR 图像采集方式，由于这些方法还没有得到普及，这里就不详细叙述了。

7.4 全亮度图合成

对于同一场景，如果只拍摄一张图像，往往会出现黑暗区域曝光度不足和明亮区域曝光过度的情况，而这些区域图像信息很难得到恢复。为此，可以通过拍摄多张具有不同曝光度的图像，然后进行综合分析，可以最大限度地还原原始场景。这就是高动态范围图像合成技术的基本原理，如图 7-5 所示。

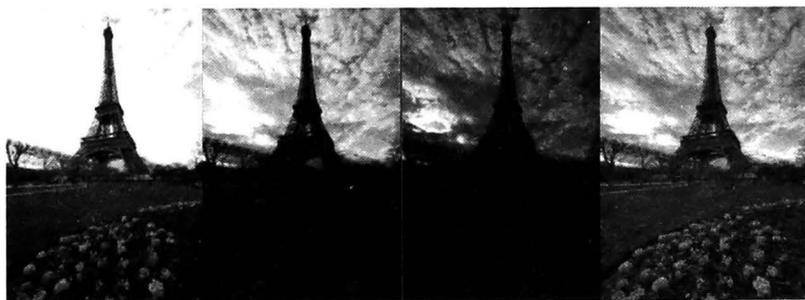


图 7-5 同一场景的不同曝光度图像序列和通过合成之后再再现的结果图

图 7-5 中，前三张是曝光度依次减小的图像，最后一张是通过前面三张合成照度图之后通过色调映射算法再现的结果图。

多曝光度图像序列合成照度图的基本原理可以分为以下几个部分。首先，恢复照相机的相应曲线，也就是使用该曲线得到图像对应场景的真实照度；其次，选取合适的 HDR 存储格式来记录所恢复的照度图；最后，合成高动态范围的图像。下面对这几步

分别加以介绍。

7.4.1 照相机响应曲线的估计

一般的数码照相机成像包括曝光、光电转换、信号放大、模/数转换、重构等几个重要的步骤。一定量的原始场景亮度 L 在数码照相机中透过透镜之后, 成为一定量 E 的照度, 然后经过感光成像、模/数转换以及重构等步骤成为每一个像素位置对应一定灰度值的数字图像。在这个过程中, 其实每一步都是一个未知的非线性关系。然而, 为了方便理解, 可以把所有这些非线性关系叠加起来, 用一个总括的非线性映射函数来表示 E 和 Z 之间的关系:

$$Z = f(E) \quad (7.5)$$

这个函数关系式 f 由特定的照相机决定, 通常被称为照相机响应曲线 (Camera Response Function, CRF)。而这个函数关系式也说明了基本的软件成像方法的基本原理: 将最终得到的灰度级 Z 映射回相对稳定的场景亮度空间, 即先估计照相机的响应曲线函数, 然后求其反函数, 从而可以由像素灰度值 Z 推导出的照度值 E 。

下面介绍一种代表性的 Debevec & Malik 算法, 这一算法由 Paul Debevec 提出, 其基本思想是利用不同的曝光度图像序列结合尺度上的相关性估计出照相机的响应函数。具体过程如下^[2]:

$$Z_{ij} = f(E_i \Delta t_j) \quad (7.6)$$

式中, Z_{ij} 表示像素值; i 是对应像素在图像中的位置标识; j 是曝光时间标识。根据实验经验, 假设照相机响应函数 (CRF) 是一个单调函数, 其反函数表示为

$$f^{-1}(Z_{ij}) = E_i \Delta t_j \quad (7.7)$$

对方程两边取对数, 且令 $F(Z_{ij}) = \ln f^{-1}(Z_{ij})$, 可得到

$$F(Z_{ij}) = \ln E_i + \ln \Delta t_j \quad (7.8)$$

式中, i 的范围是像素点的个数; j 表示具有多少张不同曝光度的图像。并且 Z_{ij} 和 Δt_j 是已知的, F 和 E_i 是未知, 这也是算法所要求解的。

由于传统数据图像的像素值范围为 $0 \sim 255$, 因此不要求出表达式 F , 只要求出 256 个 $F(Z)$ 的值即可, 最大和最小像素值分别用整数 Z_{\max} 和 Z_{\min} 表示, N 表示像素的个数, M 表示图像的张数。则问题可归结为求解最小二乘目标函数:

$$O = \sum_{i=1}^N \sum_{j=1}^M [F(Z_{ij}) - \ln E_i - \ln \Delta t_j]^2 + \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} F^M(Z)^2 \quad (7.9)$$

方程的前面部分用于求解曲线, $\lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} F^M(Z)^2$ 部分则用于平滑该曲线。其中

$F^M(z) = F(z-1) - 2F(z) + F(z+1)$, 而 λ 用于控制平滑部分相对于数据求解部分的权重, 需要参照 Z_{ij} 中噪点情况来设定。Debevec 使用单值分解 (Singular Value Decomposition, SVD) 方法来求解这个线性方程组。

图 7-6 中横坐标 Z_{ij} 表示像素值, 纵坐标 $\ln E_i$ 表示曝光量对数。图 7-6 (a) 中相同形状的点表示不同图像中的同一组点, 每一组点都对应一条特征曲线。由于并不知道曝光量的绝对数值, 这使得特征曲线可在曝光量即纵坐标上移动。图 7-6 (b) 表示通过平移

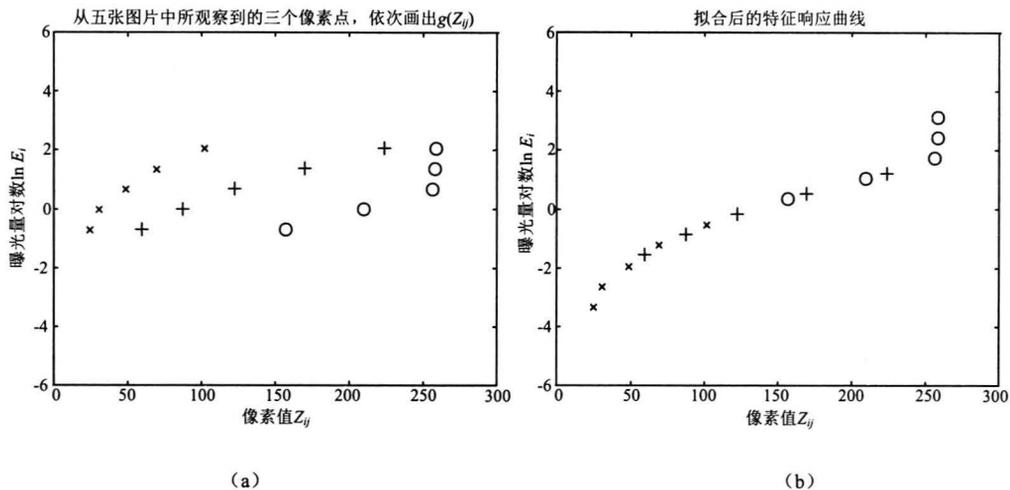


图 7-6 照相机响应曲线的拟合过程^[2]

几条曲线拟合出的特征响应曲线。

作者在文献[2]中做拟合光线响应曲线时, 为了曲线在坐标轴上的位置固定不变, 引入附加条件 $F(Z_{mid}) = 0$, 其中 $Z_{mid} = \frac{1}{2}(Z_{min} + Z_{max})$ 。这些限制条件使得位于 Z_{max} 和 Z_{min} 之间的中值像素点具有了单位曝光量。因为光线响应曲线在接近极限, 即 Z_{max} 和 Z_{min} 时, 曲线会变得陡峭, 不平滑。为了使曲线尽量满足这种特征, 引入权重函数 $\omega(z)$:

$$\omega(z) = \begin{cases} z - Z_{min} & z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & z \geq \frac{1}{2}(Z_{min} + Z_{max}) \end{cases} \quad (7.10)$$

于是式 (7.9) 改写为

$$O = \sum_{i=1}^N \sum_{j=1}^M [F(Z_{ij}) - \ln E_i - \ln \Delta t_j]^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} [\omega(z) F^M(Z)]^2 \quad (7.11)$$

对方程进行求解即可得到照相机响应光线曲线。

7.4.2 HDR 合成图像存储格式

通常用浮点型表示的 HDR 图像的亮度值范围超过 100 000 : 1, 这使得用 8 位表示的每个通道上只有 256 个离散整数值图像相形见绌。因此很多不同的技术被提出来对 HDR 的图像信息进行编码, 使之能进行有效的存储和传输, 并能保留足够的精度。

1. IEEE TIFF 格式

也许最简单和最直接的 HDR 图像存储方式是 IEEE TIFF 格式, 这种格式用三个通道 RGB 存储照度图, 每一个通道用 32 位的浮点数表示。这种存储方式意味着每一个像素必须占用 96 位。尽管这是一种非常精确的照度图存储方式, 但其缺点也很明显: 这种格式需要大量的存储空间。另外, 这种存储方式也不适合压缩, 因为浮点数格式

存在差熵特征和噪声等现象。因此, 尽管有着其精确性, 但这种格式并没有得到广泛的应用。

2. RGBE 格式

Ward 在 1991 年提出了 RGBE 格式来编码照度图^[4]。这种格式用 4B 来存储每一个像素, 这比 IEEE TIFF 格式节约了大量的存储空间, 并且这种存储方式所占空间与仅仅需要 3B 的传统图像格式相差不多。这种技术的思想是每一个通道使用一个独立尾数, 而所有的通道共用一个 8 位的指数。而尾数一般通过乘以一个以 2 为底的指数来一般化, 这个指数取值需要使得各通道上面最大值的范围为 0.5~1.0, 另外两个范围为 0~1.0。

RGBE 这种编码方式涵盖了超过 76 个数量级的范围。知道编码规则之后, 恢复编码前的数值也就不是一件难事了。然而, 尽管提出的 RGBE 文件格式在精确度和动态范围方面优于 24 位的文件格式, 但这种方法也有着很多的缺点。第一, 编码的动态范围大大超过了人眼所能感知的范围, 这导致了这一数量级未充分利用。如果有一种格式在同样的比特数量下提供了较小的数量级, 却有着更好的精度将会更好。第二, 对所用的 RGB 值使用一个通用的指数将对主色产生影响, 从而导致色差。因为当最大颜色指数一般化到 0.5~1 时, 另外两个可能为了共用同一个通用的指数而变成大约为 0 的值。第三, 非负的编码值并非覆盖了任何一个实原色的色域。尽管 RGBE 文件格式存在这些缺点, 但是在有效准确的共享高动态范围信息方面得到了最广泛的应用。

3. SGI LogLuv TIFF 编码

提出这种文件格式的作者的目的是改正 RGBE 中的缺点, 希望为照度图的编码提供一个工作标准。这一动机导致了 24 位和 32 位的 LogLuv 像素编码^[4, 5]。这一编码的原则是基于人类感知而设计的, 因此量化的步骤和人类对对比度和颜色的检测相匹配。32 位的 LogLuv 的编码使用 15 位对亮度信息进行编码, 16 位对色度信息进行编码。另一个则是符号位, 用来表示亮度是正值还是负值。亮度和色度信号可以使用下面的转换从 RGB 值得到。24 位的编码使用 10 位对亮度信息进行编码, 剩下的 14 位用来对色度信息进行编码。在这种情况下, 没有符号位。

尽管和 32 位编码方式比较起来, 24 位的编码每个像素只保存了一个字节的空, 很多学者还是推荐使用 32 位的格式来对高动态范围图像进行编码, 因为 32 位的编码方式更简单并且能提供更大的动态范围和精度。在 Sam Leffler 流行的 TIFF I/O 库中, LogLuv 编码以一种新的 SGILOG 的压缩方式得到了实现, 并且是公开和免费的。TIFF 格式的图像的读和写可以通过 LogLuv 的编码解码器(压缩/解压)模型来达到。

7.4.3 照度图的合成方法

在计算出相机光线响应曲线和选择存储格式之后, 可以由像素值得到原始场景的亮度值, 并由此得到一幅照度图。然而, 值得注意的是, 在得到了特定相机的光线响应曲线之后, 此曲线可以应用于此相机拍摄的所有图像, 以由此来确定这些图像的照度值。

根据式(7.8), 公式两边取对数得到

$$\ln E_i = F(Z_{ij}) - \ln \Delta t_j \quad (7.12)$$

由式(7.10)可以得到每一个像素对应的照度值,使用恢复函数的 M 幅不同曝光度图像来计算像素的照度值:

$$\ln E_i = \frac{\sum_{j=1}^M [F(Z_{ij}) - \ln \Delta t_j]}{M} \quad (7.13)$$

由权重函数式(7.10)可得

$$\ln E_i = \frac{\sum_{j=1}^M [\omega(Z_{ij})F(Z_{ij}) - \ln \Delta t_j]}{\sum_{j=1}^M \omega(Z_{ij})} \quad (7.14)$$

通过权重函数,让像素值接近响应函数中部的曝光量,得到更高的权重值。这样,过暗和过饱和的像素值就会被忽略掉,这样会使合成的照度图效果更好。而使用多张不同曝光度的图像生成的最终照度值可以减少噪点,降低图像质量问题带来的影响。然而,最终得到的是照度值的相对值,由前面所述可知,它与亮度值成正比,即得到的是亮度值的相对值,而不是绝对值。但是可以利用某一已知亮度来标定获得亮度的绝对值。

灰度图像的合成:对于一般的灰度图像,在前面 CRF 的恢复部分,算法中所用到的是图像各点的像素值,可以直接使用其灰度值来计算。

彩色数字图像的合成:上面所述的合成方法实现了单通道灰度图的合成,如果要合成彩色图像,只需要在此基础上加入彩色信息即可。对于彩色图像来说,因为具有 RGB 三个颜色通道,可以分别在 RGB 三个通道上计算 CRF 曲线,通过三次计算就可以求出三个通道上各自的照度值。按照灰度图像的做法,合成彩色图像时,每个颜色通道都用 Z_{mid} 来对应单位曝光量。这样具有 RGB 值 (Z_{mid} , Z_{mid} , Z_{mid}) 的像素点的 RGB 三通道具有相同的亮度值。这种方法的好处是 CRF 曲线分别求解,照度值准确,但是计算量呈 3 倍增长,速度很慢。另外,白平衡处理的图像色彩可能不是真实的色彩,造成了一定的颜色失真。因此,需要一种更好的方法来处理颜色失真问题。

HSV 色彩模型与 RGB 三原色模型相比的优点在于它可以把色彩信息同亮度信息分开。根据相机成像过程可知,在多曝光过程中,图像的色彩信息应该保持不变,变动的只是亮度,也就是说多曝光图像的差别只在 V 通道中,因为 V 通道是图像亮度的度量。因此如果将 HDR 图像表示为 HSV 模型,则 V 通道为高动态, H 和 S 保持低动态 8 位深度即可。照度图的合成也可以只在 V 通道上进行操作。主要步骤如下。

- (1) 色彩空间转换:这一步主要是提取各曝光图像的 V 通道。
- (2) CRF 曲线计算,即根据 V 通道计算照相机的响应曲线 CRF。
- (3) 求得的 CRF 和 V 通道进行加权平均合成,作为合成图像的 V 通道。
- (4) 通道 HS 的整合:目前使用的较为简单的整合算法是将各张的 HS 通道的平均值作为最终合成的 HS 通道。

7.5 色调映射技术

尽管高动态范围图像有其优越性,但是,高动态范围图像的数据格式仍然是一个高动态范围的数据格式,是不能在通常8位的显示设备上直接显示的。因此,不管高动态范围图像储存的数据有多么精确,如果不能在8位的低动态范围显示设备上使其表达的影像场景可视化,此图像不过仅仅是一些没有太大意义的数字而已。为了应对这种挑战,色调映射技术(Tone Mapping)便应运而生。所谓的色调映射技术就是如何地把高动态范围图像映射为通常的低动态范围显示设备可以接受的8位的整数格式,并使其显示出来的图像更加准确、逼真地再现原始影像场景,与原始场景看上去有一致的明暗程度。

高动态范围图像的色调映射技术在高动态范围的显示流程中起着举足轻重的作用,色调映射算法的好坏决定着最终结果的优劣,一个好的色调映射算法能够很好地再现原始场景。学术上一般把色调映射算法分为两类:一是基于色阶重建曲线的算法,即全局色调映射算法(Global Tone Mapping);二是基于色阶重建算子的算法,即局域色调映射算法(Local Tone Mapping)。下面将对两类算法进行简要介绍。

7.5.1 全局色调映射算法

全局色调映射算法是指使用单调映射函数将现实场景的亮度映射到显示亮度的算法。这个单调的映射函数可以定义为 $L_d = T(L_w)$,其中 L_d 是显示亮度, L_w 是现实场景的亮度。因为这是一个单调的函数,所以对于两个场景亮度 L_{w1} 和 L_{w2} ($L_{w1} < L_{w2}$),会得到两个完全不同的 L_{d1} 和 L_{d2} ,并且无论这两个像素点在图像中的位置是在什么地方,都有 $L_{d1} < L_{d2}$ 成立。全局色调映射算法有着计算复杂度低、运算速度快、容易使用等优点,并且在保持原始亮度顺序不变的同时还不会产生光晕等人造的效应。然而全局色调映射算法在映射过程中,会丢失图像的局部对比度和细节信息。Tumblin、Rushmeier和Ward作为这方面的先驱,分别在他们的论文^[6, 7]中阐述了色调映射在计算机图像中存在的问题,并提出了各自的全局色调映射算法。Tumblin和Rushmeier的全局色调映射算法旨在保持观察者对显示图像和真实场景的亮度信息在感觉上的一致性。而Ward^[8]也通过另外一个方式解决了这个问题,他的全局色调映射算法旨在保持观察者对显示图像与真实场景的对比度信息在感觉上的一致性。在此之后,Ferwerda等人在他们的论文^[9]中同样提出了一种新的全局色调映射算法,他们在这种算法中成功地模拟了一些重要的视觉效果,比如视觉适应性、色貌、视觉灵敏度等。之后,Larson等在文献[10]中提出基于直方图调节的改进算法,该算法利用直方图均衡的方法把现实世界的亮度映射到显示级别上,并通过人眼对对比度的敏感性做了进一步的优化。Drago在文献[11]中使用对数映射的策略来显示高动态范围图像,Drago的这种方法利用对数曲线在自变量较小时比较陡峭,而自变量较大时比较缓慢的特点,能够增加低亮度部分的对比度的同时,降低高亮度部分的对比度。在文献[12]中,段江等提出了一种新颖的基于神经系统的色调映射算法。而在文献[13]中,作者通过对多种全局色调映射算法得到的结果进行分析,结合起来显示高动态范围图像。然而文献[13]所述的方法,虽然能够利用各种全局色调映射算法各自的优点来显示高动态范围图像,但是这注重的只是信息的可视化,并非这

里所说的色调映射算法。在文献[14]中，段江等提出了一种综合了线性映射和直方图均衡的色阶重建曲线，得到了非常好的可视效果。下面对几种重要的全局色调映射算法进行简要介绍。

1. 指数映射算法

最简单的色阶重建曲线当属文献[15]中所介绍的指数映射算法了。在文献[15]中，作者使用式(7.15)把现实场景的亮度映射到显示亮度，即

$$D = D_{\max} \times \left(1 - e^{-\frac{1}{I_{\text{ave}}}} \right) \quad (7.15)$$

式中， I_{ave} 是场景的平均亮度； D_{\max} 是所能显示的最大亮度。指数映射曲线如图7-7所示。正如作者所说，这种指数方法的目标并不是与人眼的视觉反应保持一致，而是提供一种压缩方法，使原始场景的低亮度值部分压缩程度低，高亮度值部分压缩程度高。尽管这种方法有着自己的优点，但是使用这种方法往往把场景的平均亮度映射到 $0.635 \times D_{\max}$ 的显示亮度，这样就使得非常亮的场景区域看起来非常暗，而非非常暗的地方看起来非常亮。

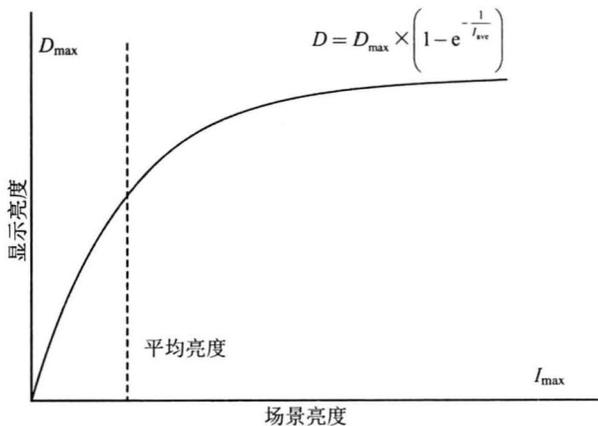


图 7-7 指数映射曲线^[15]

2. 对数映射算法

另外，在文献[11]中，作者 F. Drago 等使用一种自适应的对数映射策略来显示高动态范围图像。他们首先使用式(7.16)做一次对数映射：

$$I_d = \frac{\log(I_w + 1)}{\log(I_{\max} + 1)} \quad (7.16)$$

式中， I_d 和 I_w 分别表示显示亮度和场景亮度； I_{\max} 表示场景的最大亮度。并且，F. Drago 等注意到，通过改变式(7.16)中对数的底数，可以得到不同的映射结果图。然而，却不能通过使用一个固定的底数来达到平衡不同结果图的目的。但对数映射函数却有这样一个特点：比较陡峭的较小底数的对数函数能够增加低亮度值的对比度和明亮度，使得在黑暗区域有着较好的对比度和可见性；同样，比较平缓的较大底数的对数函数却能比

较彻底地压缩高亮度区域的值。在一个场景中，每一个不同底数的对数函数总是比较适合映射原始场景的某一亮度范围。在接下来的工作中，F. Drago 等人研究出一种新的方法，将不同对数函数对应的理想部分提取出来，组成一个更加有效的映射函数。其思想是：较小的亮度值需要较小的底数，而较大亮度值需要较大的底数。他们使用前面介绍的指数映射算法达到了这样的目的。

3. 快速色调映射算法

除了上面的两种方法之外，最为知名和有效的全局色调映射算法当属段江等在文献[14, 16, 17]中介绍到的全局色调映射方法，在这三篇论文中所介绍的全局色调映射算法的基本思想大致一样，文献[16]在文献[14]的基础上做了一定的扩展，文献[17]又在文献[16]的基础上做了一定的优化。图 7-8 表示的是高动态范围照度图的直方图，而色调映射算法应该给密集区域分配较多的显示值，松散区域分配较少的显示值，同时又保持着原始场景的相对对比度。这样的操作意味着直方图对应的松散区域压缩得多，密集区域压缩得少，甚至有时在松散区域还有轻微的扩张。

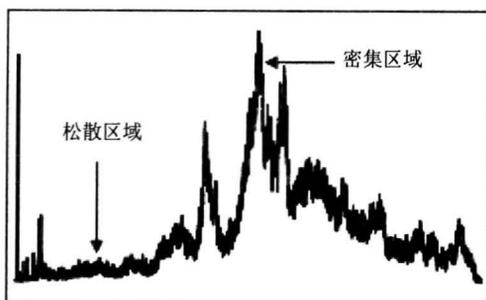


图 7-8 照度图的对数直方图^[14]

在文献[17]介绍算法中，使用式(7.17)把高动态范围图像 I 的亮度压缩到需要显示的亮度 D 上面：

$$D(I) = (D_{\max} - D_{\min}) \frac{\log(I + \tau) - \log(I_{\min} + \tau)}{\log(I_{\max} + \tau) - \log(I_{\min} + \tau)} + D_{\min} \quad (7.17)$$

式中， I_{\min} 和 I_{\max} 表示场景的最小和最大亮度值； D_{\min} 和 D_{\max} 表示显示设备所能够显示的最大亮度值和最小亮度值，其实 D_{\min} 和 D_{\max} 也就定义了显示设备的动态范围，通常分别是 0 和 255； τ 是用来控制图像总体明暗程度的关键值。 τ 值越大，图像越暗； τ 值越小，图像越亮（见图 7-9）。不同图像应取不同的 τ 值，需要通过不断的实验得到，为了自动求得 τ 值，文献[17]提出了一种自动求得适合不同图像 τ 值的算法。这个算法对于大量高动态图像都适用，可以将图像的整体亮度水平调节到一个合适的水平。该算法受到文献[18]思想的启发。在文献[18]中，作者认为每张图像亮度的对数平均值应该映射到不同的关键值（Key Value），图像越亮，这个点就应该越靠近显示设备的可显示动态范围的亮端，反之亦然。具体步骤如下。

第一步：由式(7.18)求得图像亮度的对数平均值为

$$I_{\text{ave}} = \exp \left[\frac{1}{N} \sum_{x,y} \log(\varepsilon + I(x, y)) \right] \quad (7.18)$$

式中, N 是图像像素点的总数; 因为对于完全黑暗的像素点, $I(x, y)$ 可以为 0, 为了使得取对数始终可行, 将所有 $I(x, y)$ 加上一个小值 ε , 而这并不影响最终结果。

第二步: 利用式 (7.19) 求得图像的关键值, 此值应在 0~1, 为

$$K = A \times B^{((2 \log I_{\text{ave}} - \log I_{\text{min}} - \log I_{\text{max}}) / (\log I_{\text{max}} - \log I_{\text{min}}))} \quad (7.19)$$

式中, A 和 B 是常数, 根据经验通常将其设为 0.4 和 2, 所以 K 应该在区间 [0.2, 0.8] 上。由上面的公式可以看出, 图像越亮, 其亮度对数平均值就越大, 则其关键值也就越大, 反之亦然。

第三步: 利用式 (7.20) 求得 τ , 使得 I_{ave} 能够映射到关键值 K 。

$$K = \frac{\log(I_{\text{ave}} + \tau) - \log(I_{\text{min}} + \tau)}{\log(I_{\text{max}} + \tau) - \log(I_{\text{min}} + \tau)} \quad (7.20)$$

上述方程可由数值计算解得。在文献 [17] 中使用了牛顿法来求 τ 值, 一般使用 20 次迭代即可得到合适的 τ 值。如图 7-9 所示, 图 (d) 中的 τ 值为自动求得值, 可以看到, 映射后的图像亮度值很适中。

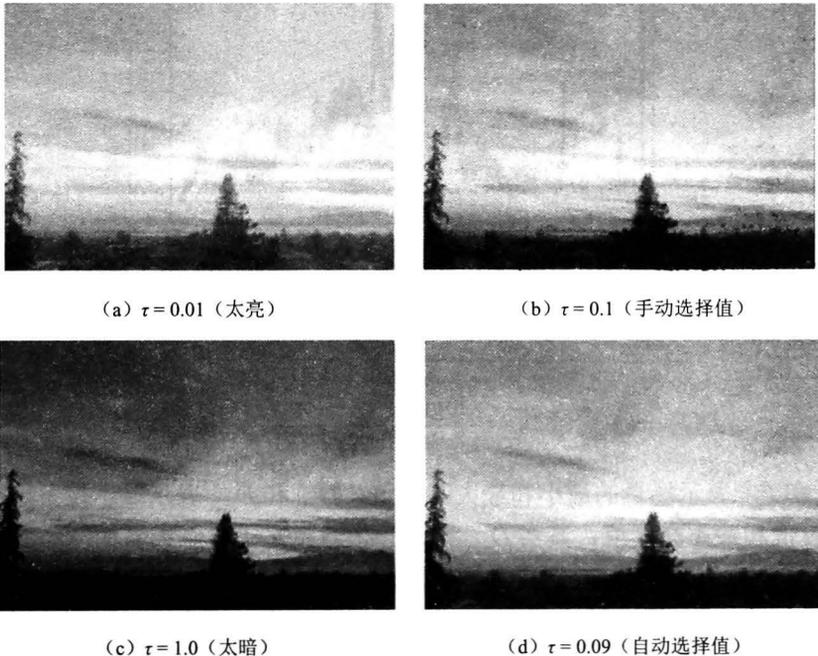


图 7-9 用式 (7.18) 得到的高动态图像的色调映射图

尽管运用式 (7.18) 可以将照度图映射到一个合适的亮度, 但是映射后的图像往往对比度很低, 很多细节不能保留, 原因如图 7-9 所示。在图 7-9 中, $D(I)$ 被线性划分成 256 个区间, 区间分割点为 l_n , 同一个区间的亮度值将被映射到同一个整数亮度值 d , d 的范围是 [0, 255], 这其实是一个量化的过程, 并且只考虑到了亮度的具体值, 而忽略了亮度的分布特性。因此, 在分布很密集的区域, 很多亮度值将被映射成同一个整数值,

造成映射后的图像细节丢失,对比度低;相反,在分布很稀疏的区间,较少的亮度值被映射到同一个整数值,造成了浪费。一个通常的解决方案是直方图均衡,这个方案将亮度值的分布也考虑了进去。 $D(I)$ 同样被划分成 256 个区间,区间分割点为 e_n ,每个区间的像素点总数相同。这种分割方式仅仅只考虑到了亮度值的分布,而没有考虑亮度的具体值。和前述相同,同一个区间的亮度值将被映射到映射后图像中的同一个整数亮度值 d 。这样,现实设备的动态范围可以很好地被利用。但是这样也存在问题,在分布较密集的区域,映射后的对比度将被夸大,而在分布较稀疏的区域,又存在过分压缩的问题,导致细节的丢失。

然而,线性量化和直方图均衡两种方法是优缺点互补的,文献[17]的全局映射算法旨在两者之间寻求平衡,取得更好的映射结果,区间分割点为 le_n 。把这个算法称为基于直方图调节的线性均衡量化算法(Histogram Adjustment-based Linear to Equalized Quantizer, HALEQ),它能够简单快速地实现线性量化和直方图均衡两种方法的平衡,使得映射后的图像效果最佳,具体原理可以参考相关资料。

7.5.2 局部色调映射算法

全局色调映射算法在显示高动态范围图像时操作简单,能得到不错的显示效果。然而这类算法会造成图像细节信息的丢失,且这些细节信息直接影响着图像的对比度,对图像的质量却是非常重要的。因此,局部色调映射算法被引入,尽管这类算法有着相对比较低的运算效率,但是处理结果却有着较高的对比度,局部细节信息也得到了很好保持。

在 1999 年, Tumblin 和 Turk 在他们的论文^[19]中提出了 LCIS (Low Curvature Image Simplifier) 算法。LCIS 算法将输入的高动态图像数据分隔成不同的层,对每个层分别进行压缩,然后再对压缩过后的每一层进行处理,最后结合这些处理结果,得到所需要显示的图像。LCIS 算法从高对比度图像中提取细节信息,因此避免了光晕(Halo)现象。采用与 Tumblin 和 Turk 类似的方法, Durand 和 Dorsey 在文献[20]中提出了一种更简单的基于分层显示高动态范围图像的方法。在文献[20]中,作者使用双边滤波将图像分成基层(Base Layer)和细节层(Detail Layer),仅在基层上压缩对比度,细节层保持不变,然后再把压缩后的基层和细节层结合起来组成需要显示的图像。作者之所以选择双边滤波,是因为双边滤波在模糊操作中能够很好地保持住图像的边缘信息,从而避免了在重组过程中的光晕现象。在文献[21]中,双边滤波也被 Li 等人采用,他们首先使用双边滤波把图像分解成基层和细节层,利用全局算法调整基层,之后再使用基层得到的增益图来增强细节层,然后再把这两层结合起来,得到最终的色调映射图像。然而,这种算法速度很慢,效率较为低下。

在文献[22]中,作者提出了一种多尺度图像处理算法来显示高动态图像。这个算法利用了一种对称的分析——综合滤波器簇,为多尺度子带图像计算出了光滑增益图,并且该算法成功地避免了光晕现象。同时, Reinhard 等在文献[18]中提出一种新颖的色调映射算法来显示高动态范围图像,在文献[18]中,作者的灵感来自于摄影中的遮挡和放相技术(Dodging and Burning),在论文中分别介绍了全局的和局部的色调映射算法。而 Fattal 等在文献[23]中则介绍了一种在对数空间上基于梯度的色调映射算法。他们在对数的亮度空间上计算梯度,并在与原始图像相关的区域检测对比度的级别,然后再用

比例因子来压缩梯度并且控制在高梯度上压缩比较多,而在低梯度上压缩比较少。然后再在改变梯度的区域通过指数运算恢复到亮度域。在文献[24]中,作者介绍了一种新颖的交互方法对图像的色调值以及其他视觉参数做局部的调整。在这种方法中,用户可以指定自己感兴趣的区域,算法将根据选择的区域自动调整亮度、对比度以及相关的其他参数值。而在文献[17]中,段江等提出了基于规则分块的局部色调映射算法。在文献[17]中,作者把把图像分成 $m \times n$ 块,然后在每一块上分别使用文献[14]中介绍的全局色调映射算法,之后使用双边加权的方法来去除边界效应和光晕现象,结果中细节信息得到了很好的保留,具有很高的对比度。下面将对文献[17]中所介绍的几种局部色调映射算法做简要的介绍。其中,将对文献[17]中的基于规则分块的色调映射算法做详细介绍。

本章前面介绍了一种快速的全局色调映射算法,而在文献[17]中,作者提出了一种新颖的基于规则分块的色调映射算法。作者首先把图像数据分成固定大小的块,比如每块大小为 32×24 个像素,然后分别在每一块上做 7.1.3 节中所介绍的快速全局色调映射算法 HALEQ,从而得到如图 7-10 所示的结果,可以看到有着明显的边界效应。

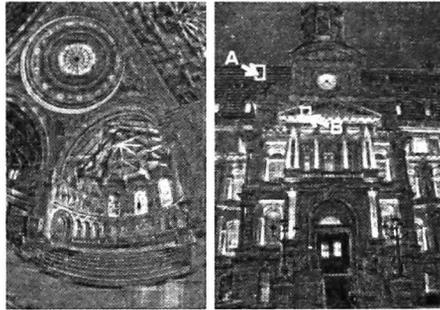


图 7-10 在每一块上面做 HALEQ 后所得到的结果^[17]

为了解决如图 7-10 所示的边界效应,作者介绍了一种基于距离权重的解决方法,如图 7-11 所示。对于图像中的每一个像素 $D(x, y)$,最终的映射像素值是根据式 (7.21),即 $\text{HALEQ}_1[D(x, y)]$, $\text{HALEQ}_2[D(x, y)]$, $\text{HALEQ}_3[D(x, y)]$ 等的权重均值计算出来的,为

$$d(x, y) = \frac{\sum_1^K \text{HALEQ}_n[D(x, y)]w_d(n)}{\sum_1^K w_d(n)} \quad (7.21)$$

式中, K 为计算权重函数时的所考虑的块的数量,比如作者在文献[29]中选择的是 5×5 的邻接块。其中权重函数

$$w_d(n) = e^{-(d_n/\sigma_d)} \quad (7.22)$$

式中, d_n 是当前像素到块 n 中心点的距离, σ_d 用来控制图像的平滑程度, σ_d 越大,在计算权重 $w_d(n)$ 时 d_n 的影响力就越小。这也就意味着,如果 σ_d 值比较大,则可以很好地去去除边界效应,但是却降低了局部对比度。而如果 σ_d 值比较小,可以保持好局部对比度,边界效应却有可能不能完全去除。为此,应该找一个合适的值,既能去除边界效应,又能保持好局部对比度,作者经过实验发现取 $\sigma_d = 20$ 可以能得到较好的效果,如图 7-12 所示。

在增强图像的对比度时,如果增强的强度太大,在图像变化比较平滑的区域则会引入噪声现象(见图 7-12 右图的天空部分)。在作者所介绍的方法中, α 参数控制了图像增强

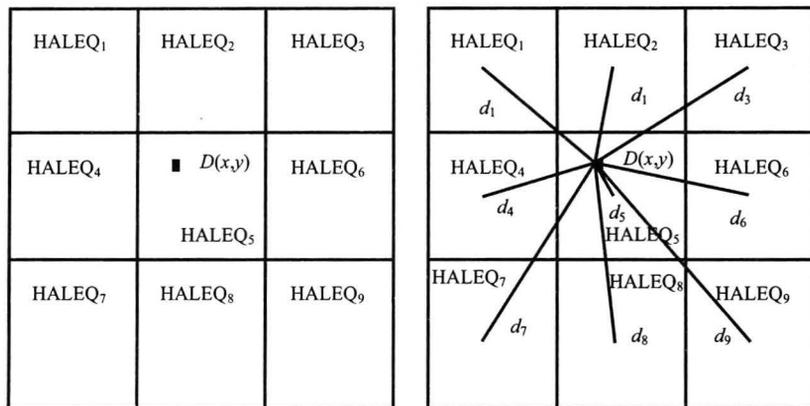


图 7-11 距离权重函数被用来去除边界效应 (为了描述方便, 只列出了 3×3 块^[17])



图 7-12 使用权重平均的方法去除边界效应后的结果^[17]

的强度, 为此, 需要控制 α 的取值来控制图像的增强强度, 而关键则是如何确认区域到底是不是平滑区域。正如作者所提到的, 图像在平滑区域做对数映射后的直方图有着比较窄的形状, 而非平滑区域的直方图形状则比较宽, 如图 7-13 所示。

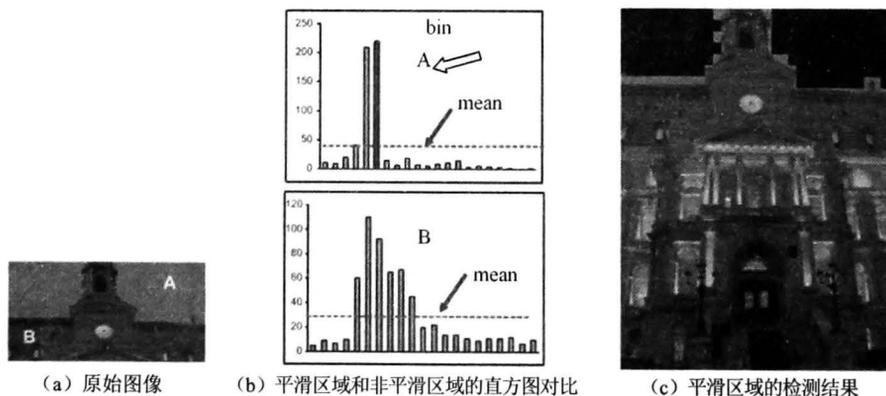


图 7-13 平滑区域和非平滑区域的效果对比^[17]

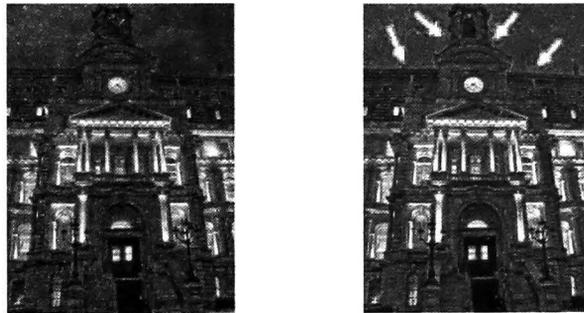
为了衡量平滑区域, 在直方图中, 每一个亮度值对应的柱状图为一个 bin, 如图 7-13 (b) 所示。因此定义:

$$SD_n = \frac{\sum_0^M |\text{Hist}[i] - \text{mean}_n|}{\text{binNumber}} \quad (7.23)$$

式中， M 是 bin 的数量， $\text{Hist}[i]$ 表示在第 i 个 bin 的像素总量，而 mean_n 表示块 n 中每一个 bin 的平均像素总量。 SD_n 越大，表明像素数量分布离 mean_n 也就越远，相应的区域也就越平滑。如果认为 $SD_n > \eta$ 时是平滑区域，则该区域增强程度要小， η 是设定的一个阈值，它决定什么样的区域属于平滑区域，在文献[17]中取 η 等于 17 能取得满意结果。因此，参数 α 为

$$\alpha = 0.6 \times [1 - e^{-[20 - SD_n]}] \quad (7.24)$$

式 (7.24) 意味着 SD_n 越大， α 越小，相应的区域增强程度也就越小。通过上面介绍的方法对参数 α 的控制，可以控制相应区域的增强程度，从而避免了噪声的产生，结果如图 7-14 所示。然而尽管如此，但是在仍然有着较为严重的光晕现象，如图 7-14 (b) 所示。



(a) 考虑平滑区域之前的效果 (b) 考虑平滑区域之后的效果

图 7-14 在考虑平滑区域之前和之后的对比^[17]

为了解决光晕现象，引入了双边权重策略来解决，即在做权重计算时，不但考虑像素与块中心的几何距离，还考虑块与块对应点在值上的相似度，这一原理与双边滤波相似，即

$$d(x, y) = \frac{\sum_1^N \text{HALEQ}_n[D(x, y)]w_d(n)w_s(n)}{\sum_1^N w_d(n)w_s(n)} \quad (7.25)$$

其中

$$w_d(n) = e^{-(d_n/\sigma_d)} \quad (7.26)$$

$$w_s(n) = e^{-(s_n/\sigma_s)} \quad (7.27)$$

$$s_n = \frac{|D(x, y) - D_{\text{mean}_n}|}{D_{\text{max}}} \quad (7.28)$$

式中， $w_d(n)$ 即前面介绍的距离权重函数； $w_s(n)$ 表示亮度级别上的相似权重函数； D_{max} 表示在 $D(x, y)$ 中的最大亮度值； s_n 表示当前像素值和块 n 的平均像素值之间的 (D_{mean_n}) 关系。尽管距离权重函数能够平滑块与块之间的边界效应，但当前块会被分配最大的权重值，因此也能保留着比较好的局部对比度。而相似权重函数却能够保证平滑区域的像素值能够被映射到相同的值上，尽管它们位于不同的块上面。 σ_s 越小，去除光晕现象的效果也就越好，但同时也就意味着减小了对比度，当 $\sigma_s = 0.1$ 能够得到比较满意的结果，

同时又能保证去除边界效应和光晕现象,如图 7-15 所示。

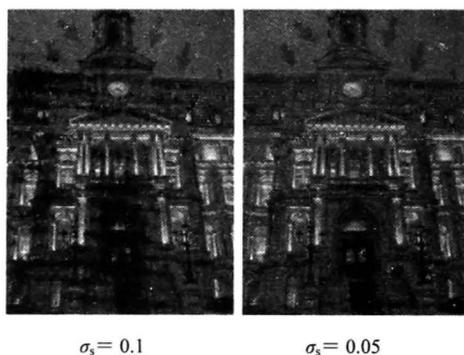


图 7-15 σ_s 取不同值对应的效果图^[17]

7.6 总结与展望

现实世界场景的动态范围达到了 10^{14} 数量级,人的视觉系统能够察觉到 10^5 数量级的动态范围,并且通过慢慢调整适应,最终可以察觉到高于 10^9 数量级的动态范围。为了能够记录影像场景更大的动态范围,一个有效的策略就是采用大于 8 位的图像数据格式来表达影像场景的信息。随着成像技术和计算机图形图像技术的进一步发展,目前的一些影像捕捉设备已经可以直接采集到 10 位、12 位、16 位或 32 位的亮度数据。采用高精度的图像数据格式便于人们更加准确地记录影像场景。一般情况下,把亮度值高于 8 位描述的图像统称为高动态范围图像。本章简要介绍了高动态范围成像技术中的各个环节和相应技术,重点介绍了色调映射技术。

由于色调映射算法涉及大量的运算,属于计算密集型的算法类型,单从算法层面的加速和优化是很难使局部色调映射算法的运算效率达到大规模的提升。幸运的是,图形处理单元(GPU)为大规模地加速局部色调映射算法提供了机会。现在的 GPU 已经发展成为具备强劲性能并且灵活易用的流式信息处理器。能够用 GPU 加速局部色调映射算法的一个重要的前提条件就是算法应该具备良好的并行运算的特点,即算法计算的各个部分应该相对地独立于其他的部分。对全局色调映射算法而言,只要计算出映射函数,其像素的映射就变成了简单的各自独立的对应关系,其并行性是很高的。然而目前绝大多数的局部色调映射算法由于要过多地考虑像素之间的关系,比较难满足这一特点。因此,设计出能产生高品质图片的局部色调映射算法,并且从根本上使其具有良好并行运算的特征,化解局部化的色调映射算法难以并行化这一矛盾是色调映射算法未来发展的主要方向。

色调映射技术的本质是如何将现实世界中高动态的亮度范围在低动态的显示设备中展现出来。不难发现,它的假设前提是显示设备不足以覆盖现实场景的动态范围。那么如果这项假设本身就不成立呢?也就是说,显示设备如果本身就是高动态的话,那么色调映射技术在这个意义上确实没有太大必要。现有的 HDR 显示设备主要针对的用户还是影片后期处理、卫星气象、医疗这些行业应用,在个人消费市场上还没有得到普及。

其主要原因还是受制于 LED 器件的工作效率以及生产成本等问题。最根本的原因是人们使用现在制造的白光 LED 虽然能产生 HDR 显示器所需要的超高亮度,但是它将电流转化为光能的效率还不高,相当大的能量转化为没用的热能,所以在现有的设备上还需要有功能强大的水冷和电风扇设计。因此, HDR 显示器也是未来高动态技术的研究和发展方向之一。

参 考 文 献

- [1] 母睿. 高动态范围图像色调映射算法及视频增强技术研究[D]. 成都: 西南财经政法大学, 2011.
- [2] Debevec P E, Malik J. Recovering high dynamic range radiance maps from photographs[C]//ACM SIGGRAPH 2008 classes. ACM, 2008: 31.
- [3] Mitsunaga T, Nayar S K. Radiometric self calibration[C]//Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. IEEE, 1999, 1.
- [4] Ward G. Real pixels[J]. Graphics Gems II, 1991: 80-83.
- [5] Larson G W. LogLuv encoding for full-gamut, high-dynamic range images[J]. Journal of Graphics Tools, 1998, 3(1): 15-31.
- [6] Larson G W. Overcoming gamut and dynamic range limitations in digital images[C]//Color and Imaging Conference. Society for Imaging Science and Technology, 1998, 1998(1): 214-219.
- [7] Tumblin J, Rushmeier H. Tone reproduction for realistic images[J]. Computer Graphics and Applications, IEEE, 1993, 13(6): 42-48.
- [8] Ward G. A contrast-based scalefactor for luminance display[J]. Graphics gems IV, 1994: 415-421.
- [9] Ferwerda J A, Pattanaik S N, Shirley P, et al. A model of visual adaptation for realistic image synthesis[C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, 1996: 249-258.
- [10] Larson G W, Rushmeier H, Piatko C. A visibility matching tone reproduction operator for high dynamic range scenes[J]. Visualization and Computer Graphics, IEEE Transactions on, 1997, 3(4): 291-306.
- [11] Drago F, Myszkowski K, Annen T, et al. Adaptive logarithmic mapping for displaying high contrast scenes[C]//Computer Graphics Forum. Blackwell Publishing, Inc, 2003, 22(3): 419-426.
- [12] Qiu G, Duan J, Finlayson G D. Learning to display high dynamic range images[J]. Pattern recognition, 2007, 40(10): 2641-2655.
- [13] Pardo A, Sapiro G. Visualization of high dynamic range images[J]. Image Processing, IEEE Transactions on, 2003, 12(6): 639-647.
- [14] Duan J, Qiu G. Fast tone mapping for high dynamic range images[C]//Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE, 2004, 2: 847-850.
- [15] Ferschin P, Tastl I, Purgathofer W. A Comparison of techniques for the Transformation of Radiosity Values to Monitor Colors [C]//Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference. IEEE, 1994, 3: 992-996.
- [16] Duan J, Qiu G, Chen M. Comprehensive fast tone mapping for high dynamic range image visualization[C]//Pacific Graphics. 2005.
- [17] Duan J, Bressan M, Dance C, et al. Tone-mapping high dynamic range images by novel histogram adjustment[J]. Pattern Recognition, 2010, 43(5): 1847-1862.
- [18] Reinhard E, Stark M, Shirley P, et al. Photographic tone reproduction for digital images[C]//ACM Transactions on Graphics (TOG). ACM, 2002, 21(3): 267-276.
- [19] Tumblin J, Turk G. LCIS: A boundary hierarchy for detail-preserving contrast reduction[C]//Proceedings of the 26th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., 1999: 83-90.
- [20] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images[J]. ACM Transactions on Graphics (TOG), 2002, 21(3): 257-266.

-
- [21] Li X, Lam K M, Shen L. An adaptive algorithm for the display of high-dynamic range images[J]. *Journal of Visual Communication and Image Representation*, 2007, 18(5): 397-405.
 - [22] Li Y, Sharan L, Adelson E H. Compressing and companding high dynamic range images with subband architectures[C]//*ACM Transactions on Graphics (TOG)*. ACM, 2005, 24(3): 836-844.
 - [23] Fattal R, Lischinski D, Werman M. Gradient domain high dynamic range compression[C]//*ACM Transactions on Graphics (TOG)*. ACM, 2002, 21(3): 249-256.
 - [24] Lischinski D, Farbman Z, Uyttendaele M, et al. Interactive local adjustment of tonal values[C]//*ACM Transactions on Graphics (TOG)*. ACM, 2006, 25(3): 646-653.

第 8 章 全景成像技术

8.1 概 述

随着图像传感器和数字图像处理技术的快速发展，数字成像设备取得了极大的进步，其应用也越来越广泛。就图像质量而言，从最开始仅仅只有几十万像素的低分辨率图像，到中期的 720P、1080P 等百万级像素的图像，再到目前的千万级像素的高分辨率图像，可以说，数字成像技术取得了跨越式的发展。

然而，随着应用的不断扩大以及用户需求的不断提高，传统数字成像设备的缺点也逐渐地暴露出来，其视场较小的缺点开始成为限制其发展的瓶颈。而目前，越来越多的场合都要求获得大视场的图像，以便从中提取感兴趣的信息。故而，获得超大视场图像成为了成像领域一个新的热点问题。

针对这一新出现的应用需求，全景成像技术应运而生。那么，什么叫全景成像呢？所谓全景成像，是指采用特殊的成像装置或者拍摄手段来获取超大视角的一种成像技术。基于这种成像技术，能够获得水平和垂直方向上 $180^\circ\sim 360^\circ$ 的超大视角的全景图像。全景成像技术能够一次性地提供所拍场景的全方位信息，使得人们有更多的时间用于后续的图像处理。目前，全景成像技术在安防监控领域已经取得了较大的发展，在机器人导航、军事侦察等领域也开始发挥作用。

8.1.1 全景图像

首先，要了解什么是全景图像。全景图像是可以表示水平方向 360° 、垂直方向大到半球以上视场的场景图像。全景图像按基准表面的映射模型一般可分为球表面、立方体表面和圆柱面等几种类型，它是通过将所得原始图像投影到这些简单表面上完成的。

目前，一般存在三种全景图像：一种是球面全景图像，一种是多面体全景图像，还有一种是最常用的柱面全景图像。如图 8-1 所示，它们分别把视点空间看成是球体、立方体与圆柱体。球面全景与多面体全景均可反映空间中任意方向的场景，但处理它们的难度比较大。柱面全景实际上是它们的简化形式，如图 8-2 所示。柱面全景没有顶盖与底盖两部分场景，限制了用户在垂直方向的观察角度，但在水平方向是 360° 视角，能满足大部分

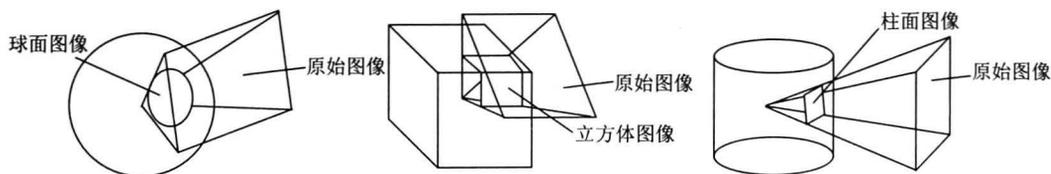


图 8-1 球面全景图像、多面体全景图像和柱面全景图像

应用的需要。柱面全景处理起来比球面全景和多面体全景简单得多,因而其应用比较广泛。

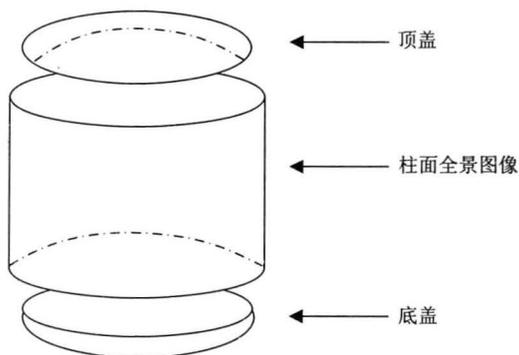


图 8-2 柱面全景图像概念图

8.1.2 全景成像方法

全景成像发展到现在,已成为了一项相对成熟的技术,就目前而言,全景成像方法主要有三种:拼接式全景成像、鱼眼全景成像以及折反射全景成像。

拼接式全景成像是全景成像技术刚刚提出时所采用的方法,通过用传统的相机拍摄得到场景的多幅图像,然后基于拼接、融合技术,合成得到场景的全景图像。拼接式全景成像的具体方法有两种,一种是采用一个相机,通过旋转相机对不同方向进行拍摄,从而获得多幅场景图像。这种方法简单易行,能够获得高分辨率的全景图像,但是由于成像速度缓慢,不具有实时性。另一种是采用多个相机,同时对多个方向同时进行拍摄,从而获得多幅场景图像。这种方法能够满足实时性的要求,也能获得高分辨率的全景图像,但是整个系统较为复杂,而且由于使用了多个相机,其成本也比较高昂。不管是采用一个相机进行旋转拍摄,还是采用多个相机同时拍摄,其目的都是为了采集得到包含整个场景信息的多幅图像,最后再合成得到全景图像。

鱼眼全景成像利用超广角物镜来实现全景成像。超广角物镜的焦距非常短,通常为6mm到16mm,因此能够获得接近180°视场角的半球面场景。而且,由于它采用凝视方式进行拍摄,不需要旋转移动,因此具有较好的隐蔽性。然而,这种全景成像方法存在一个缺点,它拍摄获得的图像存在极大的桶形畸变,即图像的中心区域景物保持不变,而其他本应该水平和垂直的景物则沿各个方向从中心向外辐射,形成鱼眼效果。因此,想要对拍摄得到的鱼眼全景图像进行显示观察,需要事先进行畸变校正,将之校正为符合人眼观察习惯的全景图像。此外,由于这类系统的结构复杂,设计难度大,想要再增大其拍摄视场已经十分困难。目前,鱼眼全景成像技术已经比较成熟,大量的鱼眼全景成像设备出现在市场。

折反射全景成像是将常规成像设备和反射光学元件相结合的成像技术,利用反射光学元件的反射特性来扩大成像设备的视场,从而能够获得水平方向360°和垂直方向一定角度的全景图像。这里的反射光学元件通常指的是各种类型的反射镜,如球面反射镜、圆锥面反射镜、椭球面反射镜、抛物面反射镜以及双曲面反射镜等。根据投影模型的不同,折反射全景成像系统可分为基于透视投影成像模型的系统 and 基于平行投影成像模型

的系统。根据是否满足单视点成像约束，又可分为单视点成像系统和非单视点成像系统。由于折反射全景成像能够获得大于半球空间的大视场，同时系统结构较为简单，设计较为灵活，成本也比较低，近年来得到了越来越广泛的研究和应用，已经成为了当前全景成像技术研究的主流方向。

本章从现有的全景成像技术出发，详细地介绍了目前较为流行的几类全景成像方法，以及其相对应的全景图像的生成方法，使读者对全景成像的整个过程有个基本的认识。

8.2 拼接式全景成像

拼接式全景成像主要包含两种方法。一种是采用单个相机，通过旋转拍摄不同方向的场景；另一种是采用多个相机同时拍摄不同方向的场景。采用这两种方法都能够得到场景的一系列局部图像，但是，由于这些局部图像分别是在不同方向上拍下的图像，它们的投影平面存在一定的夹角。因此，需要首先将它们统一投影到某一曲面上（如球面、柱面或多面体），使得现实世界中相同的景物在不同的局部图像中是相同的。然后，再采用拼接融合技术生成场景的全景图像，如图 8-3 所示，这时得到的全景图像是没有畸变的。具体的拼接、融合方法在第 9 章和第 10 章进行详细介绍，这里不再赘述。本节主要简单介绍上述两种拼接式的全景成像方法。

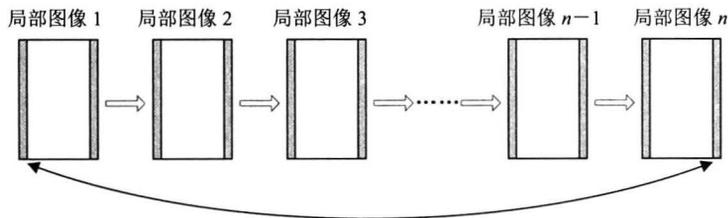


图 8-3 n 幅局部图像拼接为全景图像

8.2.1 基于单相机旋转拍摄的全景成像

首先，将相机固定在可水平旋转的支架上，相机的镜头需要位于支架的中心点。转动相机一周并隔一定角度拍照，拍摄得到一系列的场景图像，系统结构如图 8-4 所示。

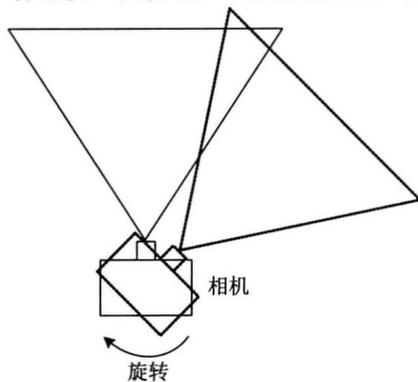


图 8-4 单相机旋转拍摄示意

其中，驱使相机转动的支架必须十分精确，通常使用全景云台控制，如图 8-5 所示。

由于要进行后续的拼接融合，对于云台的旋转要进行一定的限制，要使得拍摄得到的相邻的两幅图像之间有足够多的重合部分，以便进行拼接融合。同时，为了保证拼接融合的效果，应该使拍摄得到的图像视差尽量小，这就要求相机在旋转过程中要保持光心不动。当然，在一定的误差范围内还是可以实现最后全景图像的生成。图 8-6 所示为基于云台控制，采用单相机旋转拍摄得到三幅水平视

角为 60° 的普通图像后，经过拼接融合生成水平视角为 180° 的全景图像。

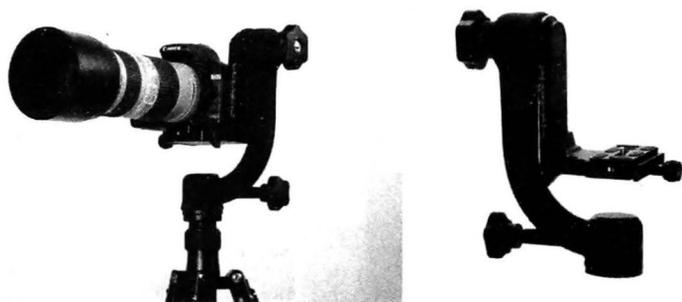


图 8-5 全景拍摄控制云台

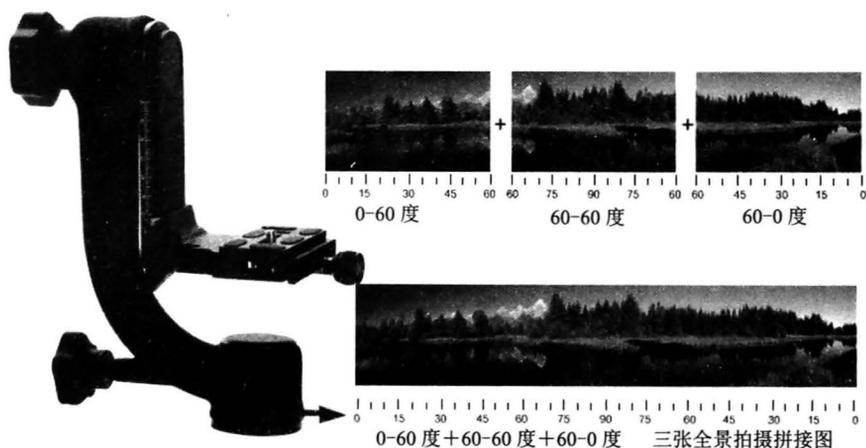


图 8-6 全景图像的生成

8.2.2 基于多相机同时拍摄的全景成像

基于多相机同时拍摄的全景成像将多个相机朝不同方向固定在同一水平面上，拍摄得到一系列的场景区图像，系统结构如图 8-7 所示。

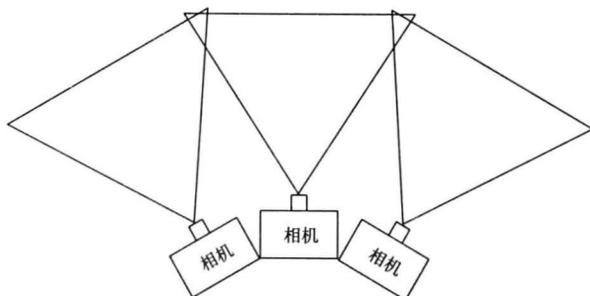


图 8-7 多相机同时拍摄示意

同样地，将多个相机同时拍摄得到的图像进行拼接，就能得到人工合成的全景图像。拍摄得到的相邻的两幅图像之间也要求有足够多的重合部分来进行拼接融合。同时，为了使拍摄得到的图像之间视差尽量地小，应该使相机光心之间的距离尽量地小。拍摄场景的

景深越小,对相机光心之间的距离要求就越高。图 8-8 所示为系统实物图,其中图 8-8 (a) 由三个镜头组成,8-8 (b) 由七个镜头组成。图 8-9 (a) ~图 8-9 (c) 为如图 8-8 (a) 所示系统拍摄得到的三幅图像,图 8-9 (d) 所示为拼接合成后的全景图像。

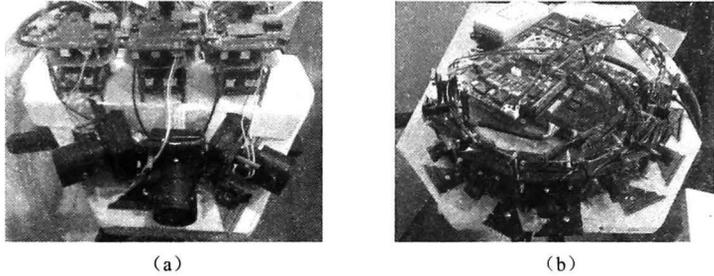


图 8-8 系统实物图

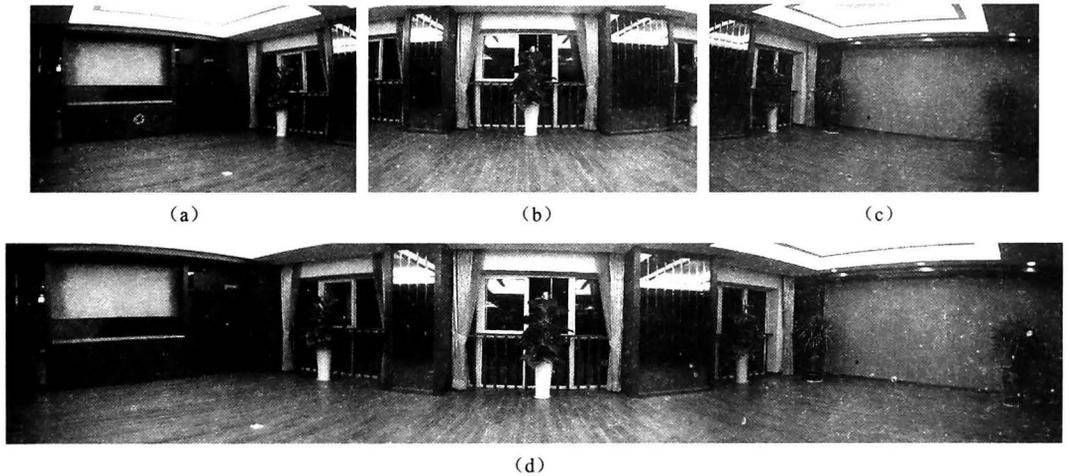


图 8-9 全景图像的生成

8.2.3 投影算法

不管是采用单个相机进行旋转拍摄,还是采用多个相机同时拍摄,它们最终获得的场景的序列图像由于是在不同方向上拍摄得到的,其投影平面存在一定的夹角。因此,需要首先将它们统一投影到某一曲面上(如球面、柱面或多面体),使得现实世界中相同的景物在不同的局部图像中是相同的。然后,再采用拼接融合技术生成场景的全景图像。

1. 柱面投影算法

1) 柱面正投影算法

实景图像的柱面正投影算法是为了将多张实景图像投影到一个圆柱面上,以柱面全景图像的形式存储。这样一方面消除了实景图像之间可能存在的重复景物信息,同时也得到了每张实景图像上的像素点在视点空间中的方位信息。在本小节中,将推导出实景图像的柱面正投影算法,并将证明该算法的一个性质,它是柱面全景图像拼接的理论基础。

如图 8-10 所示, 数码相机拍摄的一张实景图像 I , $P(x, y)$ 是实景图像 I 上的任意一个像素点。

如前所述, P 在相机坐标系下的坐标为 $\left(x - \frac{W}{2}, y - \frac{H}{2}, -f\right)$, 其中, W 和 H 分别是实景图像 I 的宽度和高度。把圆柱面的中心设为相机坐标系的原点, 并以数码相机的像素焦距 f 作为圆柱面的半径, 下面求出像素点 P 在圆柱面上的投影点 Q 在柱面全景图像中的坐标 (x', y') 。

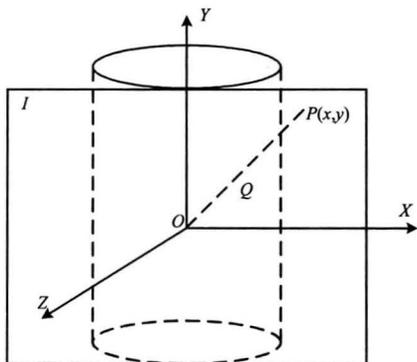


图 8-10 柱面投影示意

照相机坐标系原点与像素点 P 的直线方程可以表示为参数方程的形式:

$$\begin{cases} u = t \left(x - \frac{W}{2} \right) \\ v = t \left(y - \frac{H}{2} \right) \\ w = t(-f) \end{cases} \quad (8.1)$$

式中, t 是参数。圆柱面的方程可以表示为

$$u^2 + w^2 = f^2 \quad (8.2)$$

联立式 (8.1) 和式 (8.2) 可得

$$\begin{cases} t = \frac{f}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \\ u = \frac{f \left(x - \frac{W}{2}\right)}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \\ v = \frac{f \left(y - \frac{H}{2}\right)}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \\ w = \frac{f^2}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} \end{cases} \quad (8.3)$$

(u, v, w) 是像素点 (x, y) 在圆柱面上的投影点 Q 的参数坐标, 把所有这样的投影点组合在一起便得到全景图像。但是, 参数坐标是三维的, 需要把它们转变为二维的图像坐标才便于存储。这里, 选用下面的公式把三维的参数坐标转化为二维的图像坐标:

$$\begin{cases} x = f \arctan\left(\frac{u}{w}\right) + f\theta \\ y = v + \frac{H}{2} \end{cases} \quad (8.4)$$

式中, $\theta = \frac{h_{fov}}{2} = \arctan\left(\frac{W}{2f}\right)$ 。

联立式 (8.3) 和式 (8.4) 可得

$$\begin{cases} x' = f \arctan\left(\frac{x - \frac{W}{2}}{f}\right) + f \arctan\left(\frac{W}{2f}\right) \\ y' = \frac{f\left(x - \frac{W}{2}\right)}{\sqrt{\left(x - \frac{W}{2}\right)^2 + f^2}} + \frac{H}{2} \end{cases} \quad (8.5)$$

式 (8.5) 是将实景图像 I 上的任意一个像素点 $P(x, y)$ 正投影到柱面全景图像上一个像素点 $Q(x', y')$ 的投影公式。下面将证明该投影算法具有使景物在垂直方向上不发生形变的性质。

假设实景图像 I 上有竖直形状的景物, 取该景物在同一条垂直线上的两个像素点 (x, y) 和 (x', y') , 由式 (8.5) 可知这两个像素点在柱面全景图像上仍将有相同的横坐标。

实景图像的柱面正投影算法具有上述性质, 可以分别对每一张实景图像进行柱面投影变换, 然后通过图像拼接得到完整的柱面全景图像。

图 8-11 是实景图像柱面投影示意图, 图 8-11 (a) 和图 8-11 (c) 是实景图像, 图 8-11 (b) 和图 8-11 (d) 是对应的柱面投影变形图。

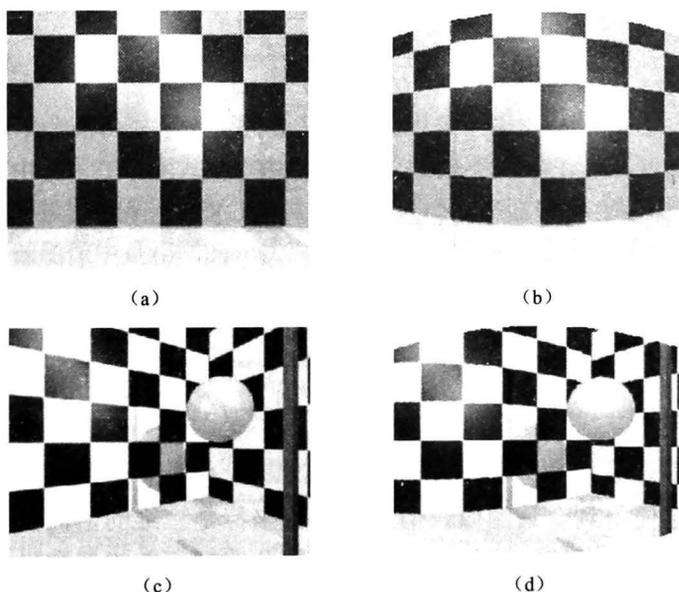


图 8-11 实景图像柱面投影示意

2) 柱面全景图像反投影算法

在得到柱面全景图后,可通过成像模型进一步从柱面全景图像重新构造出柱面视点空间每一个视线方向所对应的透视图像。变换投影算法如图 8-12 所示。设图像 C 为柱面全景图像,其高和宽分别为 H 、 W , $Q(x', y')$ 是柱面全景图像 C 上任意一个像素点, J 是生成的透视图像,假设 Q 在 J 上对应的点为 $P(x, y)$ 。若观察点位于圆柱中心 O ,即照相机坐标系 $OXYZ$ 的原点,下面将推导出 (x', y') 与 (x, y) 的关系。

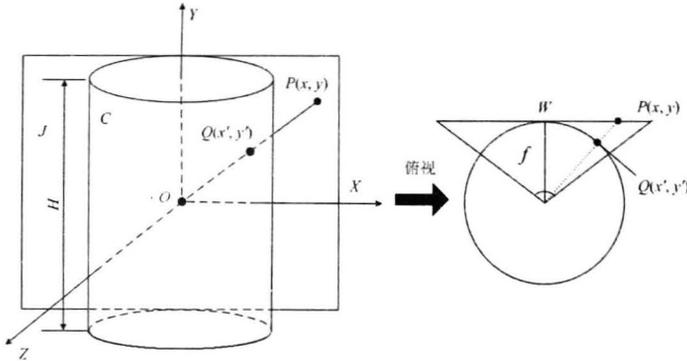


图 8-12 柱面全景图像透视展开示意

由前面的介绍可以求出 (x', y') 在圆柱面上的对应点 $R(u, v, w)$ 为

$$\begin{cases} u = f \sin \left(\frac{x' - f \arctan \left(\frac{W}{2f} \right)}{f} \right) \\ v = y' - \frac{H}{2} \\ w = f \cos \left(\frac{x' - f \arctan \left(\frac{W}{2f} \right)}{f} \right) \end{cases} \quad (8.6)$$

式 (8.6) 没有考虑照相机当前的方向,假设当前照相机相对于初始方向旋转了 θ ,则有

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad (8.7)$$

连接视点 O 与点 R 的直线的参数方程可表示为

$$\begin{cases} r = tu' \\ s = tv' \\ q = tw' \end{cases} \quad (8.8)$$

在照相机坐标系 XYZ 下,视平面的方程为

$$q = -f \quad (8.9)$$

联立式 (8.8) 和式 (8.9),得

$$t = -\frac{f}{w'} \quad (8.10)$$

令

$$\begin{cases} x = r + \frac{W}{2} \\ y = s + \frac{H}{2} \end{cases} \quad (8.11)$$

联立上述诸式可得

$$\begin{cases} x = -f \tan \left(\frac{x' - f \arctan \left(\frac{W}{2f} \right)}{f} - \theta \right) + \frac{H}{2} \\ y = -\frac{y' - \frac{H}{2}}{\cos \left(\frac{x' - f \arctan \left(\frac{W}{2f} \right)}{f} - \theta \right)} + \frac{H}{2} \end{cases} \quad (8.12)$$

而其逆关系为

$$\begin{cases} x' = f \arctan \left(\frac{x - \frac{W}{2}}{f} \right) + f \arctan \left(\frac{W}{2f} \right) \\ y' = -\frac{f \left(y - \frac{H}{2} \right)}{\sqrt{\left(x - \frac{W}{2} \right)^2 + f^2}} + \frac{H}{2} \end{cases} \quad (8.13)$$

基于上式建立了柱面全景图像上任意一个像素点 $Q(x', y')$ 与视线方向 θ 处反投影图像上的对应点 $P(x, y)$ 的函数关系。图 8-13 和图 8-14 是柱面全景图像反投影算法实验结果图。

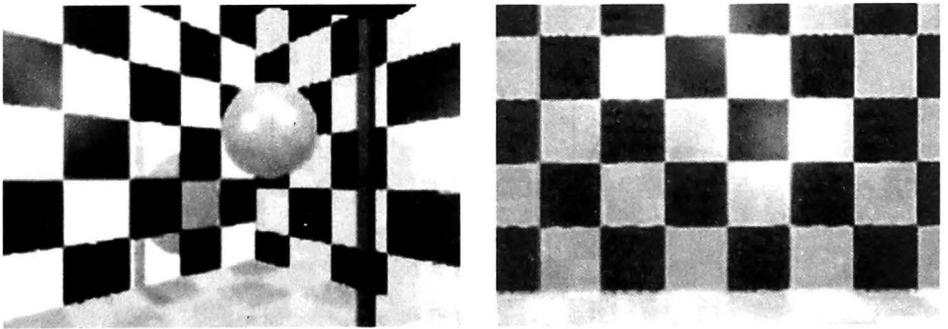


图 8-13 两幅柱面反投影图像 (一)



图 8-14 两幅柱面反投影图像 (二)

2. 球面投影算法

1) 球面正投影算法

实景图像的球面正投影算法是为了将多张实景图像投影到一个球面上,以球面全景图像的形式存储,这样做一方面消除了实景图像之间可能存在的重复信息,同时也得到了每张实景图像上的像素点在视点空间中的方位信息。在下面的内容中,将推导出实景图像的球面正投影算法。

如图 8-15 所示,图中世界坐标系为 XYZ , 照相机坐标系为 xyz , 假设照相机坐标系 xyz 是由世界坐标系 XYZ 绕世界坐标系中的 X 轴旋转,再绕世界坐标系中的 Y 轴旋转 β 得到的。实景图像 I 上任意一个像素点 P 的图像坐标设为 (x, y) , 设它在球面全景图像上的对应点 Q 在球面全景图像上的图像坐标为 (x', y') , 推导出 (x, y) 与 (x', y') 的关系。

像素点 $P(x, y)$ 在照相机坐标系 xyz 下的坐标为 $(x - \frac{W}{2}, y - \frac{H}{2}, -f)$, 则它在世界坐标系 XYZ 下的坐标 (u, v, w) 为

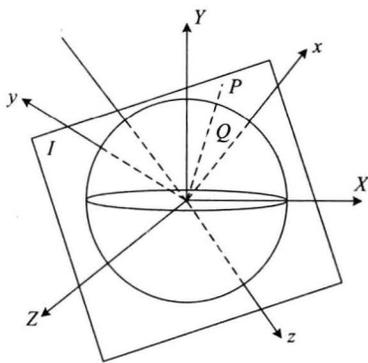


图 8-15 实景图像球面投影示意

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x - \frac{W}{2} \\ y - \frac{H}{2} \\ -f \end{bmatrix} \quad (8.14)$$

过点 P 的直线参数方程为

$$\begin{cases} u' = tu \\ v' = tv \\ w' = tw \end{cases} \quad (8.15)$$

图 8-15 中的球面方程可以表示为

$$u'^2 + v'^2 + w'^2 = f \quad (8.16)$$

联立式 (8.15) 和式 (8.16) 得

$$t = \frac{f}{\sqrt{u^2 + v^2 + w^2}} \quad (8.17)$$

与柱面全景图像生成算法一样, (u', v', w') 表述球面全景图像的三维参数坐标, 需要把它们转化为二维的图像坐标, 以便于存储。选择下面的方法实现这种转化。

当 $w' \geq 0$ 时, 令

$$\begin{cases} x' = f \cdot \arccos\left(\frac{u'}{\sqrt{u'^2 + v'^2}}\right) \\ y' = f \cdot \left(\frac{\pi}{2} + \arctan\left(\frac{v'}{u'^2 + v'^2}\right)\right) \end{cases} \quad (8.18)$$

否则令

$$\begin{cases} x' = f \cdot \left(2\pi - \arccos\left(\frac{u'}{\sqrt{u'^2 + v'^2}}\right)\right) \\ y' = f \cdot \left(\frac{\pi}{2} + \arctan\left(\frac{v'}{u'^2 + v'^2}\right)\right) \end{cases} \quad (8.19)$$

令

$$\Delta = \left(y - \frac{H}{2}\right) \sin \alpha \cos \beta - \left(x - \frac{W}{2}\right) \sin \beta - f \cos \alpha \cos \beta \quad (8.20)$$

由以上诸式可知, 当 $\Delta \geq 0$ 时, 有

$$\begin{cases} x' = f \cdot \arccos\left(\frac{(x - W/2) \cos \beta + (y - H/2) \sin \alpha \sin \beta - f \cos \alpha \sin \beta}{\sqrt{(x - W/2)^2 + ((y - H/2) \sin \alpha - f \cos \alpha)^2}}\right) \\ y' = f \cdot \left(\pi/2 + \arctan\left(\frac{(y - H/2) \cos \alpha + f \sin \alpha}{\sqrt{(x - W/2)^2 + ((y - H/2) \sin \alpha - f \cos \alpha)^2}}\right)\right) \end{cases} \quad (8.21)$$

当 $\Delta < 0$ 时, 有

$$\begin{cases} x' = f \cdot \left(2\pi - \arccos\left(\frac{(x - W/2) \cos \beta + (y - H/2) \sin \alpha \sin \beta - f \cos \alpha \sin \beta}{\sqrt{(x - W/2)^2 + ((y - H/2) \sin \alpha - f \cos \alpha)^2}}\right)\right) \\ y' = f \cdot \left(\pi/2 + \arctan\left(\frac{(y - H/2) \cos \alpha + f \sin \alpha}{\sqrt{(x - W/2)^2 + ((y - H/2) \sin \alpha - f \cos \alpha)^2}}\right)\right) \end{cases} \quad (8.22)$$

2) 球面全景图像反投影算法

在得到球面全景图后, 可通过成像模型进一步从球面全景图像重新构造出球面视点空间每一个视线方向所对应的透视图像。变换投影算法如图 8-16 所示。图像 S 是一张拼接完毕的球面全景图像, 为了推导方便, 用一个球面来表示, Q 是球面全景图像 S 上任意一个像素点, 它的图像坐标为 (x', y') , J 是需要生成的视图, 假设 Q 在 J 上对应的点为 P , P 在 J 上的图像坐标为 (x, y) 。图中的坐标系 XYZ 是世界坐标系, 坐标系 xyz 是照相机坐标系, 假设照相机坐标系 xyz 是由世界坐标系 XYZ 绕世界坐标系中的

X 轴旋转 α ，再绕世界坐标系中的 Y 轴旋转 β 得到的。下面，推导出 (x, y) 与 (x', y') 的关系。

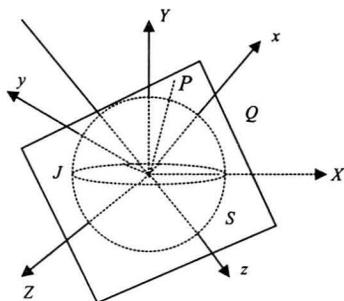


图 8-16 球面全景图像反投影示意

点 $Q(x', y')$ 对应于球面上的点 $Q'(u, v, w)$ ，有

$$\begin{cases} u = f \cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \cos\left(\frac{x'}{f}\right) \\ v = f \sin\left(\frac{\pi}{2} - \frac{y'}{f}\right) \\ w = f \cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \sin\left(\frac{x'}{f}\right) \end{cases} \quad (8.23)$$

将点 Q' 在世界坐标系 XYZ 下的坐标 (u, v, w) 变换为照相机坐标系 xyz 下的坐标 (u', v', w') ，则有

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad (8.24)$$

过点 Q' 的直线在照相机坐标系 xyz 下的参数方程为

$$\begin{cases} r = tu' \\ s = tv' \\ q = tw' \end{cases} \quad (8.25)$$

视平面在照相机坐标系 xyz 下的参数方程为

$$q = -f \quad (8.26)$$

联立上面诸式可得

$$t = -\frac{f}{w'} \quad (8.27)$$

令

$$\begin{cases} x = r + \frac{W}{2} \\ y = s + \frac{H}{2} \end{cases} \quad (8.28)$$

联立以上诸式得

$$\begin{cases} x = \frac{W}{2} \frac{f \cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \cos\left(\frac{x'}{y} + \beta\right)}{\cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \cos \alpha \sin\left(\beta + \frac{x'}{f}\right) - \sin\left(\frac{\pi}{2} - \frac{y'}{f}\right) \sin \alpha} \\ y = \frac{H}{2} \frac{f \sin\left(\frac{\pi}{2} - \frac{y'}{f}\right) \cos \alpha + f \cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \sin \alpha \sin\left(\frac{x'}{f} + \beta\right)}{\cos\left(\frac{\pi}{2} - \frac{y'}{f}\right) \cos \alpha \sin\left(\beta + \frac{x'}{f}\right) - \sin\left(\frac{\pi}{2} - \frac{y'}{f}\right) \sin \alpha} \end{cases} \quad (8.29)$$

上式建立了球面全景图像上任意一点 $Q(x', y')$ 与球面空间中视线为 (θ, φ) 的图像上一点 $P(x, y)$ 的函数对应关系。

下面两幅图像（见图 8-17 和图 8-18）是球面全景图像反投影算法的实验结果图。

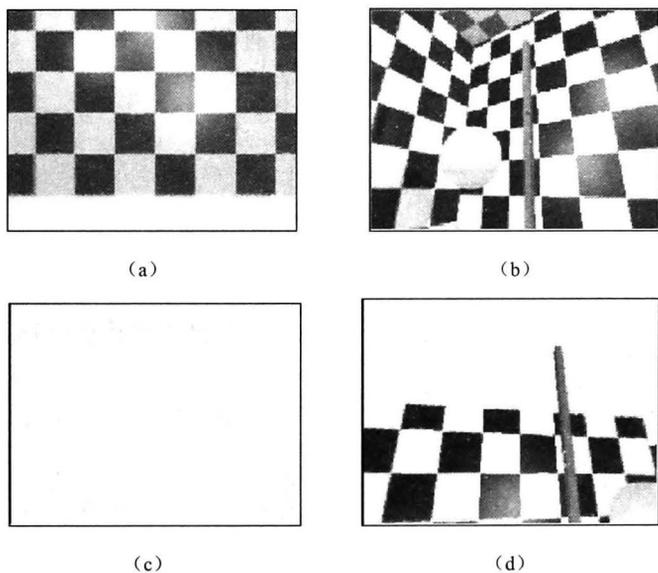


图 8-17 球面全景图像反投影图像（一）

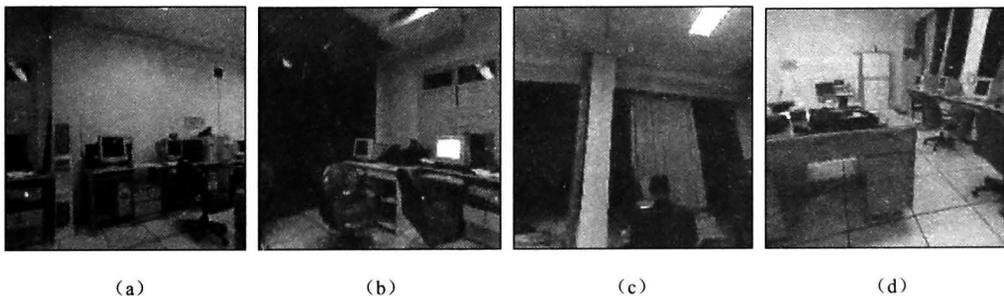


图 8-18 球面全景图像反投影图像（二）

8.3 鱼眼全景成像

鱼眼全景成像是利用超广角镜头来实现全景成像的一种成像技术，这里的超广角镜

头通常指的就是鱼眼镜头。那么，究竟什么是鱼眼镜头呢？为什么叫做鱼眼镜头呢？所谓的鱼眼又是什么呢？它跟人眼又有什么区别呢？下面将一一进行介绍，使读者对鱼眼镜头有一个初步的认识，进而对鱼眼全景成像有个大概的了解。

8.3.1 鱼眼镜头

鱼眼镜头其实是一种焦距极短并且视角接近或等于 180° 的超广角镜头，之所以称为“鱼眼镜头”，是因为它的镜片结构与真实鱼眼的结构颇为相似。为使镜头达到最大的摄影视角，鱼眼镜头的前镜片直径很短且呈抛物状向镜头前部凸出，这恰好与鱼眼结构相似，图 8-19 所示为尼康在 1962 年设计的 8mm 鱼眼镜头的结构（9 片 5 组），这款镜头直接奠定了尼康在这个领域的领导地位，使之成为了圆形鱼眼镜头生产的王者。图 8-20 展示了目前市场上主流的尼康、佳能等品牌的鱼眼镜头。

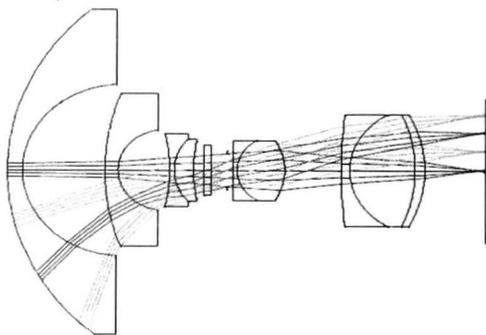


图 8-19 尼康的 9 片 5 组圆形鱼眼镜头设计（镜头长度为 88mm）



图 8-20 鱼眼镜头

鱼眼镜头的焦距很短，通常为 $6\sim 16\text{mm}$ ，这么短的焦距使得鱼眼镜头的视场范围很大，这也是其优势所在。通常，鱼眼镜头的视角可达 220° 或 230° ，这就为近距离拍摄大范围景物创造了条件。然而，众所周知，焦距越短，视角越大，因光学原理产生的形变也就越强烈。鱼眼镜头的这种设计使得拍摄得到的图像存在很大的桶形畸变，其结果是除了画面中心的景物保持不变，其他本应水平或垂直的景物都发生了相应的变化，这就使得拍摄得到的图像中的景物与现实中的景物存在很大的差别。但是，这种与现实的差异却能够产生意想不到的艺术效果。鱼眼镜头在接近被摄物拍摄时能造成非常强烈的透视效果，强调被摄物近大远小的对比，使所摄画面具有一种震撼人心的感染力。鱼眼镜

头具有相当长的景深,有利于表现照片的长景深效果。鱼眼镜头的成像有两种,一种像其他镜头一样,成像充满画面,如图 8-21 (a) 和图 8-21 (b) 所示。另一种成像为圆形,如图 8-21 (c) 和图 8-21 (d) 所示。无论哪种成像,用鱼镜头所摄的像,形变都相当厉害,透视汇聚感强烈。

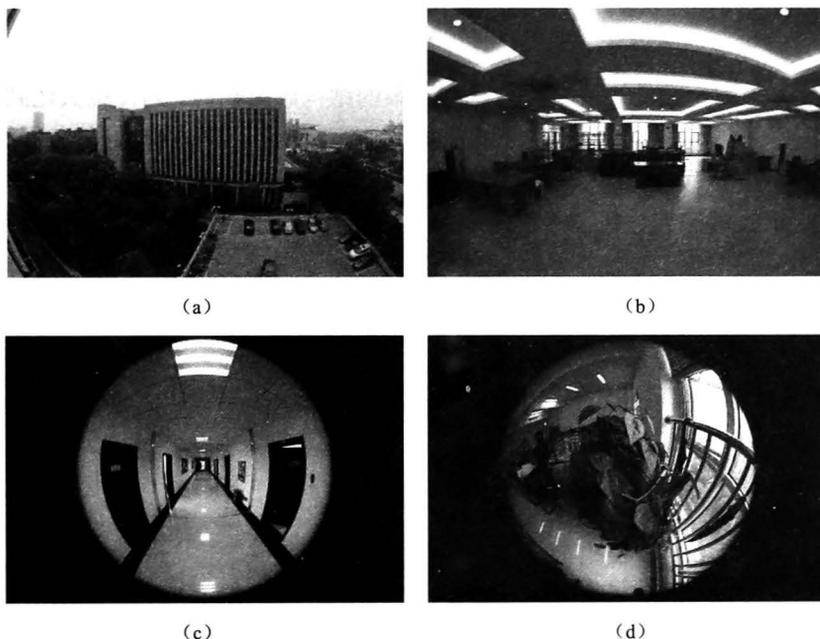


图 8-21 鱼镜头拍摄的图像

8.3.2 成像投影模型

通常,普通的光学镜头都遵守“相似”成像准则,即像与物总是相似的,理想的像高公式为

$$y' = f \tan w \quad (8.30)$$

式中, f 为光学镜头的焦距; w 为物方半视场角。对于鱼眼成像而言,其目的就是要将半球空间中的景物成像在有限的像面上,即 w 趋近于 90° , 此时, y' 将趋近于无穷大,即对于接近半球面的物面,像面将变得无穷大,这是不可能实现的,而且,将半球面成像为平面本身也违反了“相似”成像准则。因此,“相似”成像准则不适用于鱼镜头。

为了实现鱼眼成像,即将半球物面成像在有限像面上,需要选择适当的理想成像公式来满足鱼镜头的成像要求,一般有如下几种投影公式可以将半球视场或大于半球视场的空域投影在有限的像面上:

$$y' = 2f \tan(w/2) \quad (8.31)$$

$$y' = fw \quad (8.32)$$

$$y' = 2f \sin(w/2) \quad (8.33)$$

其中,式 (8.31) 描述的投影方式称为体式投影,式 (8.32) 描述的投影方式称为等距投影,式 (8.33) 描述的投影方式称为等立体角投影。上述四种投影方式所表示的像高

与视场角的关系曲线如图 8-22 所示。

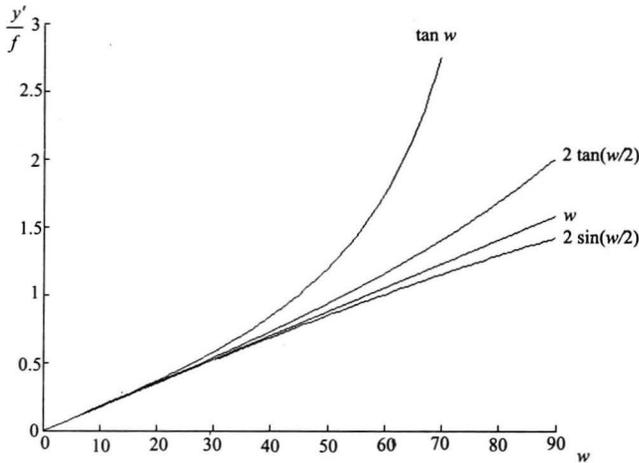


图 8-22 四种投影曲线

上述几种投影方式的主要作用就是为了引进变形，即通过“变形压缩”将半球视场或大于半球视场的空域投影在有限的像面上，从光学上说，这种“变形压缩”即是通过引入大量的桶形畸变来实现，但这种畸变不会对成像的清晰度产生影响，而是会产生一些特殊的效果，正如前文所述那样，这也是鱼镜头区别于其他镜头的特点之一。如图 8-23 所示，其他三种适用于鱼镜头的成像曲线与遵守“相似”成像准则的曲线的差值就代表了各自所能引入的桶形畸变量的大小。目前，市场上的大部分鱼镜头采用的是式 (8.32) 所描述的等距投影模型，因此，将重点介绍这一投影模型。

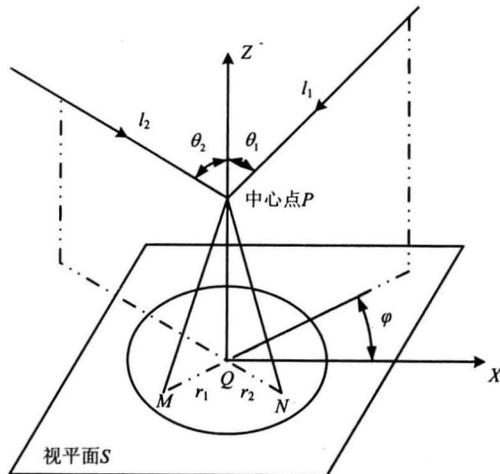


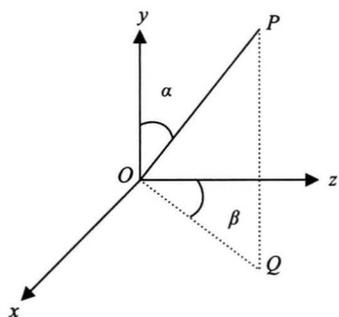
图 8-23 等距模型示意

图 8-23 所示为等距模型的示意图。图中的平面是视平面 S ，设鱼镜头的中心点为 P ，镜头中心 P 在视平面 S 上的成像点为 Q （位于视平面的中心），等距模型所描述的物理规律可以表述为：设过鱼镜头中心 P 的两条空间直线为 l_1 和 l_2 ，它们与照相机坐标系 Z 轴的夹角分别为 θ_1 和 θ_2 ，它们在视平面 S 上的成像点分别为 M 和 N ， Q 点与 M 点

的距离为 r_1 ， Q 点与 N 点的距离为 r_2 ，若 $\theta_1 = \theta_2$ ，则 $r_1 = r_2$ ，即空间中过鱼镜头中心 P 并且与世界坐标系 Z 具有相同夹角 θ 的直线在视平面 S 上的所有成像点共圆，直线的成像点在圆周上的位置取决于直线与照相机坐标系 X 轴的夹角 φ 。设空间中一条过鱼镜头中心 P 的直线与 Z 轴的夹角为 θ ，与 X 轴的夹角为 φ ，则该直线上所有点在视平面 S 上的成像点的坐标为

$$\begin{cases} x = c\theta \cos \varphi \\ y = c\theta \sin \varphi \end{cases} \quad (8.34)$$

称式 (8.34) 为鱼眼成像公式，其中 c 为常数，由照相机胶片的长度比率和鱼镜头的焦距决定。



8.3.3 鱼眼图像的校正算法

鱼镜头拍摄的鱼眼图像不能直接用于生成球面全景图像，需要对它先进行一定的校正和变换。在拍摄中使用了 Nikon FC-E8 的鱼镜头，它的水平视域达到 183° ，两张旋转 180° 的鱼眼图像就能覆盖整个视点空间，并能保证两张鱼眼图像的边缘部分存在一定程度的重叠。

在讨论鱼眼图像的校正方法之前，先证明理想鱼镜头的一个性质。如图 8-24 所示，在照相机坐标系 xyz 中，以视点为球心，以照相机焦距 f 为球半径的球面上有一个点 $P(u, v, w)$ ，点 P 在 xz 平面上的投影点为 $Q(u, 0, w)$ ，设

$$\begin{cases} \angle POY = \alpha \\ \angle ZOQ = \beta \end{cases} \quad (8.35)$$

则有

$$\begin{cases} \alpha = \arctan \left(\frac{\sqrt{u^2 + w^2}}{v} \right) \\ \beta = \arccos \left(\frac{w}{\sqrt{u^2 + w^2}} \right) \end{cases} \quad (8.36)$$

把球面参数按下式变换 (x, y) ：

$$\begin{cases} x = f \cdot \left(\arccos \left(\frac{w}{\sqrt{u^2 + w^2}} \right) \right) \\ y = f \cdot \left(\frac{\pi}{2} - \arctan \left(\frac{\sqrt{u^2 + w^2}}{v} \right) \right) \end{cases} \quad (8.37)$$

假设得到变换后的图像，注意式 (8.37) 中 x 的计算与 v 无关，这表明式 (8.37) 中的变换具有性质：照相机坐标系 xyz 中垂直直线（与 y 轴平行）在按式 (8.37) 变换生成的图像中仍是垂直直线（与图像的横轴垂直）。这一性质是对鱼眼图像进行校正的理论依据，它也使得可以对变换校正后的两张鱼眼图像进行拼接，得到一张完整的球面

全景图像。

首先, 定义三个矩阵算子 $R_x(\alpha)$ 、 $R_y(\alpha)$ 、 $R_z(\alpha)$, 它们是绕三个坐标轴的旋转矩阵, 有

$$\begin{cases} R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \\ R_y(\alpha) = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{pmatrix} \\ R_z(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{cases} \quad (8.38)$$

假设按照式 (8.37) 校正后的图像为 J , Q 是 J 上任意一个像素点, 它的图像坐标为 (x', y') , 由式 (8.37) 可求出 $Q(x', y')$ 对应的两个球面参数 α 、 β 为

$$\begin{cases} \alpha = \frac{\pi}{2} - \frac{\left(y' - f \cdot \frac{\pi}{2}\right)}{f} \\ \beta = \frac{x' - f \cdot \frac{\pi}{2}}{f} \end{cases} \quad (8.39)$$

鱼镜头等距模型的两个参数 θ 、 φ 分别是光线与 z 轴、 x 轴的夹角, 需要由 α 、 β 求出 θ 、 φ 。由图 8-25 可知, 点 P 所对应的向量 \vec{p} 为

$$\vec{p} = (\sin\alpha \sin\beta, \cos\alpha, \sin\alpha \cos\beta) \quad (8.40)$$

z 轴对应的向量 \vec{z} 为

$$\vec{z} = (0, 0, 1) \quad (8.41)$$

则有

$$\cos\theta = \frac{\vec{p} \cdot \vec{z}}{|\vec{p}| \cdot |\vec{z}|} = \sin\alpha \cos\beta \quad (8.42)$$

从而有

$$\theta = \arctan 2 \left(\sqrt{\cos^2\alpha + \sin^2\alpha \sin^2\beta}, \sin\alpha \cos\beta \right) \quad (8.43)$$

注: $\text{atan2}(m, n)$ 返回 m/n 的反余弦值。 atan2 返回 $-\pi$ 到 π 之间的值, 由 m 、 n 的符号决定返回值的象限。若 x 为 0, atan2 返回 0; 若 m 、 n 均为 0, 返回值也为 0。

点 P 在 xy 平面内的投影点 P' 对应的向量 \vec{u} 为

$$\vec{u} = (\sin\alpha \sin\beta, \cos\alpha, 0) \quad (8.44)$$

x 轴对应的向量 \vec{x} 为: $\vec{x} = (1, 0, 0)$, 则有

$$\begin{cases} \cos \varphi = \frac{\bar{u} \cdot \bar{x}}{|\bar{u}| \cdot |\bar{x}|} = \frac{\sin \alpha \sin \beta}{\sqrt{\sin^2 \alpha \sin^2 \beta + \cos^2 \alpha}} \\ \sin \varphi = \frac{\cos \alpha}{\sqrt{\sin^2 \alpha \sin^2 \beta + \cos^2 \alpha}} \end{cases} \quad (8.45)$$

设 (θ, φ) 表示球面上一点 $R(u, v, w)$, 则有

$$\begin{cases} u = f \cdot \sin \theta \cdot \cos \varphi \\ v = f \cdot \sin \theta \cdot \sin \varphi \\ w = f \cdot \cos \theta \end{cases} \quad (8.46)$$

设鱼眼图像的偏差主要是由绕三个坐标轴微小的旋转偏差 $\Delta\theta$ 、 $\Delta\varphi$ 、 $\Delta\gamma$ 造成的, (θ, φ) 精确对应于球面上一点 $R'(u', v', w')$, 则

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = R_x(\Delta\theta) \cdot R_y(\Delta\varphi) \cdot R_z(\Delta\gamma) \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad (8.47)$$

令 $\begin{cases} \theta' = a \tan 2(\sqrt{u'^2 + v'^2}, w') \\ \varphi' = a \tan 2(v', u') \end{cases}$, 由鱼镜头的等距模型, 并加上一个常数偏量(这是为了使 x, y 的值始终大于0), 得如下公式:

$$\begin{cases} x = f \cdot \theta' \cdot \cos \varphi' + \frac{W}{2} \\ y = f \cdot \theta' \cdot \sin \varphi' + \frac{H}{2} \end{cases} \quad (8.48)$$

这样即可计算出变换图像 J 上的每一个像素点 $Q(x', y')$ 在原始的鱼眼图像上的对应点 (x, y) , 根据重构的图像 J , 选择 $\Delta\theta$ 、 $\Delta\varphi$ 、 $\Delta\gamma$ 值, 使得图像 J 中的景物具有垂直方向不发生弯曲变形的性质。

图 8-25 是鱼镜头拍摄的一张鱼眼图像, 图 8-26 是变换校正后的图像(其中, $\Delta\theta = -3.0$ 、 $\Delta\varphi = 0.0$ 、 $\Delta\gamma = 0.0$), 图 8-27 是两张校正后的鱼眼图像拼接的实验结果。



图 8-25 鱼眼图像



图 8-26 校正后的图像



图 8-27 球面全景图像

8.4 折反射全景成像技术

1970年, Rees 在申请的专利中提出使用双曲面反射镜实现折反射全景成像^[1]。1996年, Pegard 等采用圆锥面作为反射镜设计全景成像系统用于移动机器人导航^[2]。1997年美国哥伦比亚大学的 Nayar 等设计并实现了一个采用抛物面反射镜和正交投影镜头构成的全景成像系统^[3]。自此, 折反射全景成像技术受到研究者的重视, 在折反射全景成像研究的推动下, IEEE 从 2000 年开始连续召开了国际全景成像专题研讨会, 其中, 折反射全景成像的文章占绝大部分, 成为计算机视觉研究的一个热点问题。

8.4.1 折反射全景成像原理

折反射全景成像技术利用曲面反射镜(如抛物面、双曲面、圆锥面等)把水平方向 360° 范围内物体的光线反射到成像传感器, 从而在一次性拍摄中获得远大于普通照相机视野范围内的景象。通常, 折反射全景成像系统拍摄获取的原始图像称为全向图像, 其包含了水平 360° 范围内的所有场景。与拼接全景和鱼眼全景成像技术相比, 折反射全景成像技术能一次性获取 360° 的全向图像, 具有大于半球空间的视场, 成像装置设计简单, 成本低等优点。

如图 8-28 所示, 折反射全景成像系统主要由三部分组成: ①感光器件, 如 CCD 或 CMOS; ②成像透镜, 如常规成像透镜或远心透镜; ③凸面反射镜, 其面形可以为二次曲面, 如球面、圆锥面、双曲面和抛物面, 或高次旋转对称非球面。

由作者所在实验室自行研制的抛物面反射全景成像系统 PROIS (Paraboloid Reflective Omni-directional Imaging System) 为例, 如图 8-29 所示, 说明折反射全景成像的原理与过程: ①来自场景中物体的入射光被凸面反射镜反射; ②反射光进入光学成像部件, 在其感光部件(如数码照相机的 CCD、CMOS)上形成关于周围场景 360° 视野范围的全向图像。

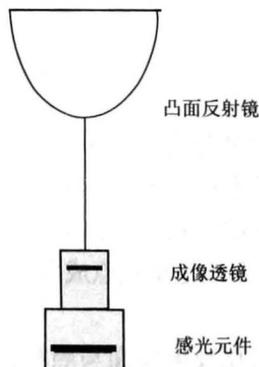


图 8-28 折反射全景成像系统构成

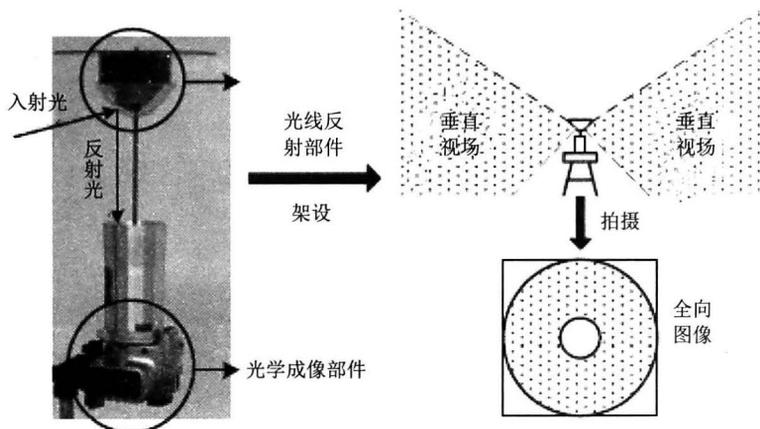


图 8-29 折反射全景成像过程

根据透镜成像模型的不同，折反射全景成像系统可以分为基于透视投影成像模型的成像系统和基于平行投影成像模型的成像系统，如图 8-30 所示。基于透视投影成像模型的系统采用常规的透视成像透镜，由于这类成像透镜较容易选购，价格又较低，因此实用较强。基于平行投影成像模型的系统采用远心成像透镜，这类透镜较难选购，且较昂贵，但由于反射镜与透镜的距离可以很小，所以系统较为紧凑。

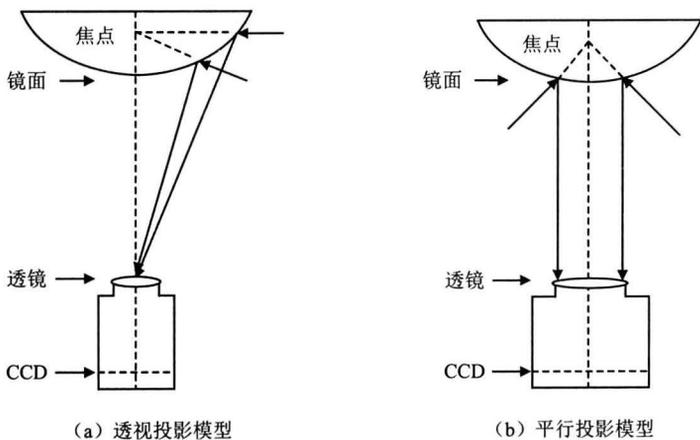


图 8-30 透视投影模型与平行投影模型示意

根据是否满足单视点约束，折反射全景成像系统又可分为单视点成像系统和非单视点成像系统，如图 8-31 所示。单视点成像是被反射镜面反射后通过摄像机的每一条入射光线的延长线都相交于同一点，即在单一的有效视点对所有方向成像，如图 8-31 (a) 所示。不满足单视点成像要求的系统称为非单视点（也称多视点）成像系统，如图 8-31 (b) 所示。

根据反射镜面与照相机类型以及它们间组合方式的不同，折反射全景成像系统具有不同的结构、不同的成像模型，以及不同的用途。

图 8-32 (a) 所示为 Nayar 设计的采用抛物面反射镜和正交投影镜头构成的全向镜头。基于抛物面和双曲面反射镜的全向摄像机是单视点成像系统，此类系统满足透视投

影成像模型条件，能用针孔成像模型进行成像分析，容易进行系统标定、图像分析和处理。但基于单视点折反射全向成像系统获得透视全向图像需要对采集的实景图像进行逆投影，计算量较大。图 8-32 (b) 所示为 R. A. Hicks 等提出的一种非单视点折反射全向成像系统，该系统可直接获取近似透视全向图像，使场景中的物点坐标与成像平面内像点坐标呈线性关系，实现全向成像的空间无畸变。

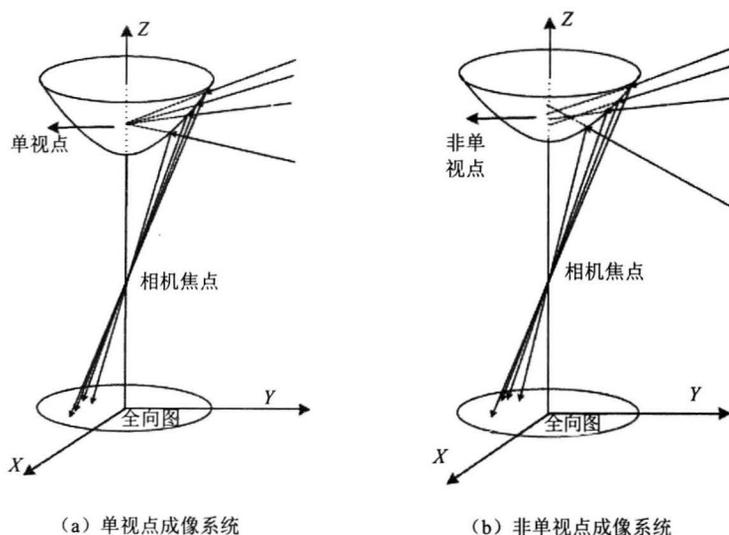


图 8-31 单视点成像系统和非单视点成像系统示意

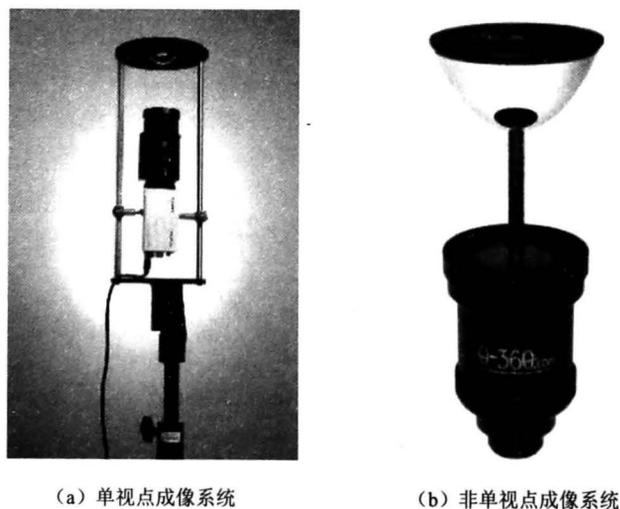


图 8-32 折反射全景成像系统实物

8.4.2 单视点折反射全景成像

通常，计算机视觉系统大多采用单视点成像模型，现有的计算机视觉理论、算法一般适用于单视点成像的折反射全景成像系统。因此，一般要求系统设计时应满足单视点约束。

在单视点的条件下，图像中的每一个像素都对应着经过视点的一条入射光线。若已知折反射系统的光路，就可以反推出每个像素对应的入射光线的方向。因此，可以将全景图像的像素点投影到离视点任意距离的某个平面上形成平面投影图像。由此可见，单视点的重要性在于它允许将全景图变换成平面投影图，看起来与普通摄像机得到的图像并无差别，而观察视角更大，从而可以使用一般的处理投影图的方法进行图像分析和处理。

图 8-33 所示为单视点全景成像示意图，下面根据此图推导满足单视点限制的折反射曲面面形公式。

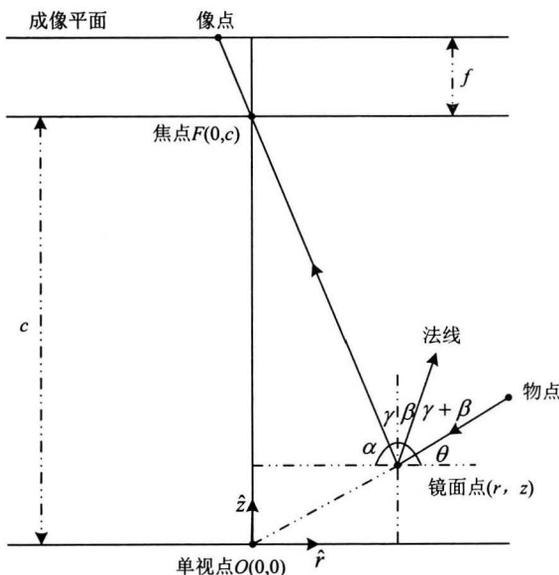


图 8-33 单视点全景成像示意

设镜面点坐标为 (r, z) ，角度关系满足：

$$\begin{cases} \gamma = 90 - \alpha \\ \alpha + \theta + 2\gamma + 2\beta = 180 \end{cases} \quad (8.49)$$

根据上述公式可以得到

$$2\beta = \alpha - \theta \quad (8.50)$$

进而能够得到

$$\frac{2 \tan \beta}{1 - \tan^2 \beta} = \frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta} \quad (8.51)$$

根据几何推导可知

$$\tan \alpha = \frac{c - z}{r} \quad (8.52)$$

$$\tan \beta = -\frac{dz}{dr} \quad (8.53)$$

$$\tan \theta = \frac{z}{r} \quad (8.54)$$

将式 (8.52) ~ 式 (8.54) 代入式 (8.51)，可得

$$\frac{-2 \frac{dz}{dr}}{1 - \left(\frac{dz}{dr}\right)^2} = \frac{(c-2z)r}{r^2 + cz - z^2} \quad (8.55)$$

变化上述公式可得

$$r(c-2z) \left(\frac{dz}{dr}\right)^2 - 2(r^2 + cz - z^2) \frac{dz}{dr} + r(2z-c) = 0 \quad (8.56)$$

这个方程的通用解为

$$\left(z - \frac{c}{2}\right)^2 - r^2 \left(\frac{k}{2} - 1\right) = \frac{c^2}{4} \left(\frac{k-2}{k}\right) \quad k \geq 2 \quad (8.57)$$

$$\left(z - \frac{c}{2}\right)^2 + r^2 \left(1 + \frac{c^2}{2k}\right) = \frac{2k + c^2}{4} \quad k > 0 \quad (8.58)$$

这两个通用解定义了所有满足单视点限制的折反射镜面曲线, 选择不同的 c 、 k 值可以得到很多理论上的解。但是, 这里的 c 和 k 具有特定的物理意义, 必须根据应用的具体要求来选取。 c 表示摄像机焦点和视点之间的距离, 必须是大于零的, 同时为了保证系统的紧凑性又不能取得过大。选定了 c 之后, 不同的 k 就决定了镜面的形状和曲率, 应根据系统的垂直视角来设计。当选择的 c 、 k 值不满足 $c > 0$ 、 $k > 0$ 的限制条件时, 方程的解是退化的, 不能用来构成实际的单视点全景视觉系统。

8.4.3 全向图像的展开

对于单视点折反射全向图像, 由于图像中的每一个像素都对应该经过视点的一条入射光线, 因此可以根据全向成像系统的参数, 将全向图像的像素点投影到离视点任意距离的某个圆柱面上形成柱面全景图。在得到柱面全景图后, 可通过成像模型进一步从柱面全景图像重新构造出柱面视点空间每一个视线方向所对应的透视图像, 透视展开的方法在 8.2.3 节中已有详细的描述。

柱面全景图像是将场景空间投影到一个指定半径的圆柱面上所获得的图像。将圆柱面沿径向切开并平铺, 可以得到一个二维的、矩形的柱面全景图, 如图 8-34 所示。柱面全景图可表达水平方向 360° 的空间信息, 是一种有效的全景表示方式, 同时也便于计算机处理。



图 8-34 柱面全景图

对于单视点全景成像系统, 由于所有入射光线的延长线均交于同一视点, 因此可以利用光路跟踪原理, 对成像点进行光路跟踪和投影变换, 建立单视点全向图像到柱面全景图像像素点之间的坐标映射关系。实际应用中, 根据全景成像系统所用反射镜面类型

和参数的不同,推算的像素坐标映射关系也不尽相同。此外,像素坐标映射关系还与成像系统焦距、位置以及柱面全景图像大小(即分辨率)等因素有关。

图 8-35 为描述全景成像系统的光路跟踪示意图。其中, A 为有效视点, O 为三维坐标原点。 C 为成像系统焦点, 焦距为 f 。反射面对应的方程为 $y = ax^2$, 半径为 r 。记三维坐标系原点 O 到圆柱底面的距离为 k , 反射镜面高 $BO = h$, 焦点到镜面顶点的长度 $CO = l$, 成像底片宽为 W , 高为 H 。把图 8-35 (a) 中虚拟圆柱面沿与 x 轴正方向相交的一条母线展开, 并以其左下角顶点为原点建立二维直角坐标系, 可得如图 8-35 (c) 所示的柱面全景图像。对于该柱面全景图像上的任意点 $P'_0(X'_0, Y'_0)$, 它与全向图像[见图 8-35 (b)]上对应像素点 $P_0(X_0, Y_0)$ 的坐标映射关系可用以下公式表示:

$$\begin{cases} X_0 = \frac{W}{2} - \frac{(aM + N)rf \cos \frac{X'_0}{r}}{aM^2 + MN + 2r^2(h+l)} \\ Y_0 = \frac{H}{2} + \frac{(aM + N)rf \sin \frac{X'_0}{r}}{aM^2 + MN + 2r^2(h+l)} \end{cases} \quad (8.59)$$

式中, $M = Y_0 - k - h$, $N = \sqrt{a^2 M^2 + 4ahr^2}$ 。

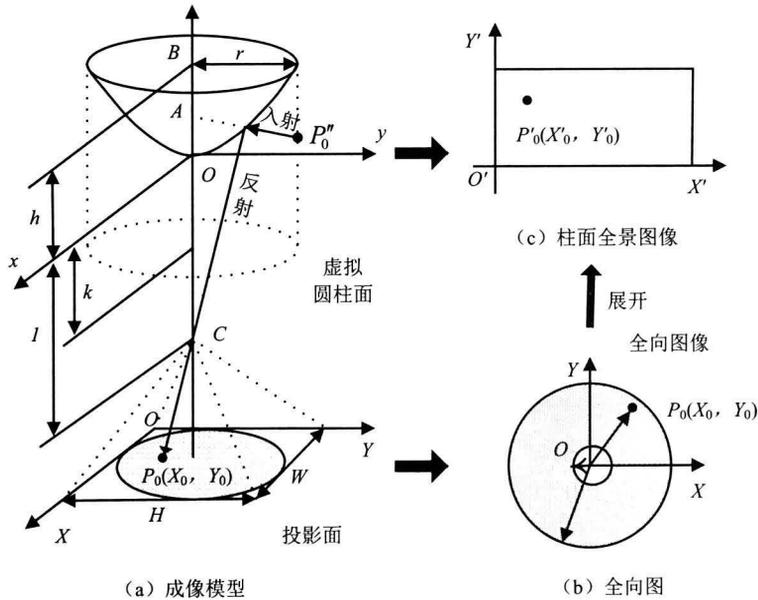


图 8-35 全向图像转换为柱面全景图像原理示意

8.4.4 互补结构折反射全景成像系统

虽然折反射全景成像系统能够一次性拍摄获取水平 360° 范围内的全景图像, 但是, 随着研究的不断深入, 折反射全景成像技术固有的两大缺陷已成为严重限制其推广和深入应用的瓶颈。

1) 空间分辨率低

在成像设备物理分辨率固定的情况下, 以照相机光轴垂直于地面放置为例, 由于折

反射全向成像具有水平 360° 宽广视野，而传统照相机成像的水平张角一般在 90° 以下，势必导致同一空间目标在成像平面上对应的像素分辨率相对来说明显偏低。换句话说，在成像分辨率固定的情况下，“看得清”和“看得广”之间的矛盾似乎很难避免。

2) 分辨率分布不均匀

由于全向图的内环像素数目明显低于外环，在同样的径向跨度中像素分辨率密度沿径向由内向外逐渐增大，导致全向成像的分辨率分布严重不均匀，即同样的物体若在全向图外环成像将比在内环成像要更为清晰。由于折反射全向成像自身特性所限，这个问题似乎也不可避免。

为了解决折反射全向成像分辨率低且分布不均匀的问题，在传统折反射全向成像原理的基础上，尝试从成像系统的结构设计上进行原创性创新，提出了一种“圆锥面镜+环形平面镜+双曲面镜”的互补结构折反射全向成像系统，使得同一空间点在多个反射镜面的共同作用下，分别经两个不同的光路同时在全向图的内环和外环互补二次成像，通过融合展开的两幅全景图像进行超分辨率全景空间重建。

如图 8-36 所示，根据小孔成像原理及光学反射定律，通过几何推导得出该结构折反射全向成像系统生成的全向图与相应的柱面全景图像之间的正向映射关系和逆向映射关系。

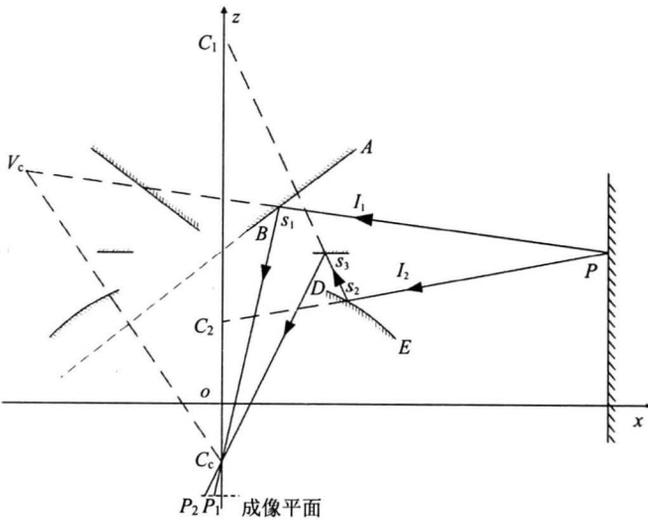


图 8-36 “圆锥面镜+环形平面镜+双曲面镜”的互补折反射全向成像示意

假设投影圆柱面的半径为 R_c ，圆柱面上一点 $p(R_c, h)$ 在全向图的内环和外环的成像点分别为 $P_1(-r_1, 0)$ 、 $P_2(-r_2, 0)$ 。全向图与柱面全景图像之间的正向映射关系即为 $P_1 = f_1(p)$ 及 $P_2 = f_2(p)$ ，而逆向映射关系即为 $p = f_1^{-1}(P_1)$ 及 $p = f_2^{-1}(P_2)$ 。

在纵截面，即 xoz 平面内进行分析，首先令照相机的成像平面方程为 $z = 0$ ，假定三个反射面的方程分别如下。

(1) 上方的圆锥面为

$$z = k \cdot x + m \tag{8.60}$$

(2) 中间的环形平面镜为

$$z = z_{pm} \tag{8.61}$$

(3) 下方的双曲面为

$$(z-L)^2/a^2 - x^2/b^2 = 1 \quad (8.62)$$

根据单视点约束, 要求相机的小孔中心 C_c 中间的环形平面镜中所成的像恰好位于双曲面的上焦点 C_1 处, 由于 C_1 点坐标 $(0, L+c)$ 已知, 可求得小孔中心 C_c 的纵坐标 $z_c = 2z_{pm} - (L+c)$ 。

1. 内环对应的圆锥面折反射成像

对于上方的圆锥面折反射成像, 点 p 发出的一条入射光线 ps_1 经双曲面上一点 s_1 反射后, 进入照相机小孔中心 C_c 并与成像面交于点 P_1 。如图 8-36 所示, 点 V_c 为照相机小孔中心 C_c 在圆锥面反射镜中所成的像, 根据圆锥面和平面镜反射特性可知, 所有的入射光线 ps_1 的延长线必经过点 V_c , 即点 V_c 相当于上方的圆锥面折反射成像的一个虚拟视点。对于整个全向图内环成像来说, 虚拟视点的轨迹即为点 V_c 绕 z 轴旋转一周所得的圆。

根据式 (8.60) 的圆锥面右纵截面所在直线方程, 可求出相机小孔中心 C_c 的像点, 即虚拟视点 V_c 的坐标 (x_v, z_v) 。由于点 s_1 即为直线 pV_c 与圆锥面 AB 的交点, 可根据如下方程组求得点 s_1 的坐标 (x_1, z_1) :

$$\begin{cases} z_1 = k \cdot x_1 + m \\ \frac{z_1 - h}{x_1 - R_c} = \frac{z_v - h}{x_v - R_c} \end{cases} \quad (8.63)$$

点 P_1 即为直线 s_1C_c 与成像平面 $z=0$ 的交点, 因此可求得点 P_1 的坐标为

$$r_1 = \frac{z_c x_1}{z_1 - z_c} \quad (8.64)$$

将式 (8.63) 解得的点 s_1 坐标 (x_1, z_1) 代入式 (8.64), 即可求得柱面全景图像上一点 p 成像到全向图内环上一点 P_1 的正向映射关系 $P_1 = f_1(p)$ 。

2. 外环对应的“双曲面+环形平面镜”折反射成像

对于“双曲面+环形平面镜”组合对应于外环的折反射成像, 点 p 发出的一条入射光线 ps_2 经双曲面上一点 s_2 反射后, 在中间的环形平面镜上一点 s_3 二次反射, 再进入照相机小孔中心 C_c 并与成像面交于点 P_2 。

根据单视点约束及双曲面特性, 所有的入射光线都射向双曲面的内焦点, 即虚拟单视点 C_2 。因此, 点 C_2 必定在入射光线 ps_2 的延长线上, 从而可根据如下方程组求得点 s_2 的坐标 (x_2, z_2) :

$$\begin{cases} (z_2 - L)^2/a^2 - x_2^2/b^2 = 1 \\ z_2 = \frac{h - (L - c)}{R_c} \cdot x_2 + (L - c) \end{cases} \quad (8.65)$$

由于点 C_1 坐标 $(0, L+c)$ 已知, 可通过如下方程组求得直线 s_2C_1 与环形平面镜的交点 s_3 坐标 (x_3, z_3) :

$$\begin{cases} z_3 = z_{pm} \\ z_3 = \frac{z_2 - (L + c)}{x_2} \cdot x_3 + (L + c) \end{cases} \quad (8.66)$$

点 P_2 即为直线 s_3C_c 与成像平面 $z=0$ 的交点, 因此可求得点 P_2 的坐标为

$$r_2 = \frac{z_c x_3}{z_3 - z_c} \quad (8.67)$$

将式 (8.66) 解得的点 s_3 坐标 (x_3, z_3) 代入式 (8.67), 即可求得柱面全景图像上一点 p 成像到全向图外环上一点 P_2 的正向映射关系 $P_2 = f_2(p)$ 。

类似地, 从全向图上一点投影到柱面全景空间的逆向映射关系 $p = f_1^{-1}(P_1)$ 及 $p = f_2^{-1}(P_2)$, 推导过程更为简单, 且由于 $P_1 = f_1(p)$ 及 $P_2 = f_2(p)$ 的简单性, 亦可直接求逆得到, 此处不再赘述。

图 8-37 (a) 所示为上述的实物原型系统, 整套成像装置由支撑部件、该支撑部件支撑固定且上下依次布置的圆锥面镜、环形平面镜和双曲面镜, 以及由高分辨率成像器件和光学镜头构成的成像设备组成。

其中, 三个反射镜面均是采用超精密单点金刚石车床加工而成, 成像设备选用德国 Basler 公司的 500 万像素 Pilot piA2400-17gc 型数字摄影机, 配套的成像镜头选用焦距为 5mm 的日本 Kowa 公司的 LM5JCM 的百万像素镜头。

在一幢大楼的楼顶进行实景成像实验, 得到内外环互补的全向图, 如图 8-37 (b) 所示。

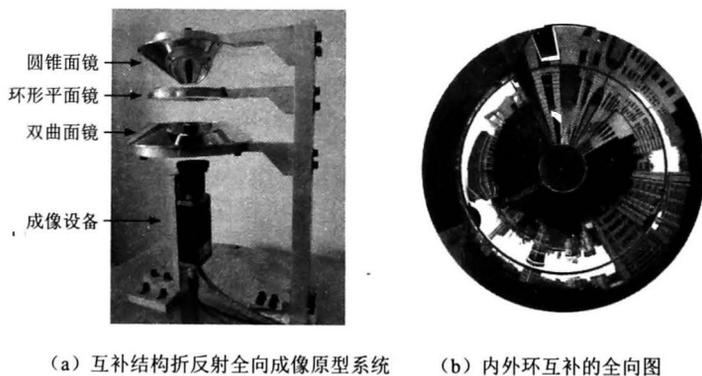


图 8-37 全向成像原型系统和全向图

参 考 文 献

- [1] REES D. Panorama television viewing system: US, 3 505 465[P]. 1970-04-07.
- [2] PEGARD C, MOUADDIB E M. A mobile robot using a panorama view: proceedings of IEEE International Conference on Robotics and Automation, Minneapolis, 1996[C]. New York: IEEE, 1996.
- [3] NAYAR S K. Catadioptric omnidirectional camera: proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, 1997[C]. New York: IEEE, 1997.

第9章 图像拼接技术

9.1 概 述

图像拼接技术，顾名思义，就是将不同的图像拼成一幅大视场、高分辨率的无缝图像的技术。这些源图像可能拍摄于不同的时间，有着不同的拍摄视角，或者拍摄自不同的传感器等。但是这些图像必须满足一个条件，拥有公共部分。更加通俗地说，就是有一部分场景内容，在相邻两幅待拼接的图像中都能找到。

如图 9-1 所示，图像拼接技术的流程主要包括以下五个步骤。



图 9-1 图像拼接流程

1) 图像预处理

图像预处理包括数字图像处理的基本操作，如去噪、边缘提取、直方图处理等。建立图像的匹配模板以及对图像进行变换，如傅里叶变换、小波变换等操作。

2) 图像配准

图像配准就是采用一定的匹配策略，找出待拼接图像中的模板或特征点在参考图像中对应的位置，进而确定两幅图像之间的变换关系。

3) 建立变换模型

根据模板或者图像特征之间的对应关系，计算出数学模型中的各参数值，从而建立两幅图像的数学变换模型。

4) 统一坐标变换

根据建立的数学变换模型，将待拼接图像转换到参考图像的坐标系中，完成统一坐标变换。

5) 图像融合

将待拼接图像的重合区域进行融合，得到大视场、高分辨率的无缝图像。

在上述步骤中，图像配准与图像融合是两个关键技术。图像配准，是图像拼接的基础，一言以蔽之，就是确定待拼接图像的相对位置。图像融合，则是通过特殊手段，消除不同图像之间的差异，包括曝光、色差、运动物体等，使拼接后的图像达到视觉上的一致性。图像融合包括两个步骤：第一步是查找到相邻两幅图像的最佳缝合线。第二步是对两幅图像进行融合，使之达到无缝、无差异的视觉效果。图像配准与图像融合中的最佳缝合线查找技术在本章介绍，图像融合中的其他内容在第 10 章中介绍。

9.2 图像配准

图像拼接的关键是精确找出相邻两张图像中重叠部分的位置,然后确定两张图像的变换关系,即图像配准。由于视角、拍摄时间、分辨率、光照强度、传感器类型等的差异,待拼接的图像往往存在平移、旋转、尺度变化、透视形变、色差、扭曲、运动目标遮挡等差别,配准的目的就是找出一种最能描述待拼接图像之间映射关系的变换模型。目前常用的空间变换模型有平移变换、仿射变换以及投影变换等^[1],如图9-2所示。

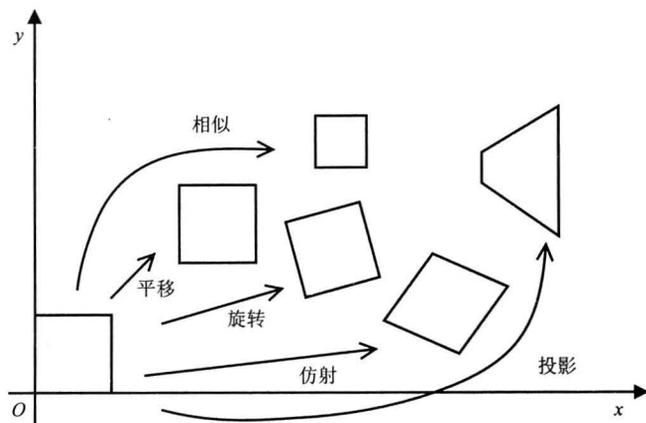


图9-2 变换模型示意

上述关系模型可以用8参数模型来描述,如式(9.1)所示。

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (9.1)$$

其中,参数矩阵 M 中各参数的作用如表9-1所示。

表9-1 投影变换矩阵 M 中各个参数作用

参 数	作 用	参 数	作 用
m_2	水平方向位移	m_0, m_1, m_3, m_4	尺度和旋转量
m_5	垂直方向位移	m_6, m_7	水平与垂直方向的变形量

根据各参数的意义及不同变换模型的特点,对矩阵 M 作相应简化就可以得到各个变换模型参数矩阵。

以航拍拼图为例,航拍拼图是将有一定重合区域的航拍图像拼接成视野范围较大的地面全景图像的过程,它能够弥补航拍分辨率低、视野范围小的缺陷,有助于人们对航拍内容有更全面、更直观的了解。在遥感遥测、军事侦察、空中监测预警、目标跟踪与无人机导航以及虚拟环境生成等方面具有重要的应用价值。

航空拍摄地面景物属于远距离拍摄。与拍摄距离比较起来,地表的起伏或地物的深度均可以忽略,可以将拍摄目标(地面场景)近似看成平面。一般而言,同一平面场景

在两个不同视点上的图像之间，满足 8 参数的关系。

然而对于航拍摄像机来说，一方面，相邻两帧图像间的变化发生在极短的时间内（典型情况是 1/30s），可以近似认为摄像机沿 x 轴和 y 轴方向的旋转角度很小；另一方面，地表的起伏可以忽略不计。基于此，提出邻近两帧之间的全局运动近似满足 4 参数相似变换模型。

对于平面场景中任意一点 P ，世界坐标系中的坐标为 (X, Y, Z) ，在两幅图像中的坐标分别为 (x_1, y_1) 和 (x_2, y_2) ，对应的齐次坐标为 $(x_1, y_1, 1)$ 和 $(x_2, y_2, 1)$ ，则有

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & -\beta & t_x \\ \beta & \alpha & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (9.2)$$

式中， α 为图像尺度变化因子； t_x 和 t_y 分别为图像在 x 轴方向和 y 轴方向上的位移。用 θ 表示图像在像平面内的旋转角度， $\theta = \beta/\alpha$ 。

9.2.1 基于特征的图像配准

基于特征的图像配准是利用图像的明显特征来估计图像之间的变换，而不是利用图像全部信息。这些明显特征，如图像的特征点（角点或关键点）、轮廓和一些不变矩等。

1) 控制点配准算法

根据 8 参数模型，图像的配准问题可以归结为求解对应点集。在待配准的图像中选取一些特征点，并且对准这些特征点，两幅图像也就配准了。控制点方法往往要借助人工选取初始匹配点，这大大降低了算法的速度和适用范围。因此有必要采用一些数学方法自动实现图像间对应控制点的选取。

2) 自动角点检测配准算法

这类算法通过检测图像中的角点，然后对两幅图像的角点按照一定的配准原则进行配准，最后剔除误匹配对，得到正确的配准结果。

角点没有明确的数学定义。一般认为角点是二维图像亮度变化剧烈的点或图像边缘曲线上曲率极大值的点。这些点在保留图像图形重要特征的同时，可以有效地减少信息的数据量，使其信息的含量很高，有效地提高了计算的速度和配准的可靠性，使得实时处理成为可能。图像的角点检测方法可概括为两类。第一类方法先将图像分割为区域，用链码表示目标边界，然后通过方向变化确定角点。这种方法的主要缺点是角点检测的结果依赖于前面的图像分割结果。第二类方法直接对图像灰度级进行操作，这些方法主要利用梯度和曲率度量检测角点。经典的第二类角点检测算法包括 Harris 角点检测算法^[2]、SUSAN（Smallest Univalued Segment Assimilating Nucleus，最小同值分割吸收核）角点检测算法以及 DoG（Difference of Gaussian，双高斯差）算子角点检测算法。

具体配准步骤如下。

- (1) 利用角点检测算子检测图像中的角点。
- (2) 利用控制点匹配算法对检测到的角点进行匹配，找出角点匹配对。

(3) 剔除伪匹配对, 得到正确匹配对, 根据这些匹配对计算出变换参数。

(4) 进行拼接融合得到全景图像。

基于特征点的配准算法计算量较小, 配准精度高, 缺点是角点少的图像、大旋转和大尺度缩放的图像和多光谱图像不能很好地进行配准。

3) 基于轮廓特征的配准算法

基于轮廓特征的配准算法首先对图像进行直方图均衡和去噪处理, 再对图像进行轮廓提取, 然后对提取的轮廓进行配准, 进而确定重叠区域。轮廓的匹配准则可以选取链码相关或者其他一些相似性准则, 如主轴和不变矩等。

基于轮廓特征的配准算法适用于光照不一致、存在尺度较大变化的图像。该方法需要准确提取出明显的轮廓特征, 对于数据的缺失比较敏感, 要求两幅图像对应轮廓要比较完整, 对于轮廓特征不明显或者噪声干扰较大的图像不适用。

4) 基于 SIFT 的配准算法

SIFT (Scale Invariant Feature Transform, 尺度不变特征变换) 配准算法由 Lowe 于 1999 年提出, 2004 年完善总结^[3]。该算法利用图像关键点的 SIFT 特征向量进行匹配, 是目前国内外特征点配准领域的研究热点。

基于 SIFT 的配准算法主要步骤如下。

(1) 检测尺度空间极值点, 初步确定关键点的位置和所在尺度。

图像 $I(x, y)$ 在不同尺度下的尺度空间可表示为图像与高斯核的卷积:

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (9.3)$$

式中, $G(x, y, \sigma) = 1/2\pi\sigma^2 e^{-(x^2+y^2)/2\sigma^2}$ 是尺度可变高斯函数; (x, y) 是图像像素位置; σ 表示尺度空间因子。

为了有效地在尺度空间检测到稳定的关键点, Lowe 在图像平面空间和 DoG 尺度空间同时检测局部极值以作为特征点。DoG 算子定义为两个不同尺度的高斯核的差分:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] \otimes I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (9.4)$$

(2) 精确确定关键点的位置和尺度, 同时剔除低对比度的关键点和不稳定的边缘相应点。

(3) 分配关键点方向。利用关键点领域像素的梯度方向分布特性为每个关键点指定方向参数, 保证 SIFT 算子的旋转不变性。关键点所在尺度的梯度模值和方向为

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (9.5)$$

$$\theta(x, y) = \tan^{-1} \{ [L(x+1, y) - L(x-1, y)] / [L(x, y+1) - L(x, y-1)] \} \quad (9.6)$$

(4) 生成关键点描述子。将坐标轴旋转为关键点的方向, 然后一关键点为中心区 8×8 的窗口, 计算每个 4×4 的小块上 8 个方向的梯度方向直方图, 每个梯度方向的累加值形成一个种子点。实际计算过程中, 为了增强匹配的稳健性, Lowe 建议对每个关键点使用 4×4 共 16 个种子点来描述。这样对于一个关键点就可以产生 128 个数据, 形成 128 维的 SIFT 特征向量。

(5) 生成两幅图像的 SIFT 特征向量后, 采用关键点特征向量的欧式距离作为两幅图像中关键点的相似性判定准则, 得到满足准则的 SIFT 匹配点对。

(6) 根据得到的 SIFT 匹配点对计算出图像的变换参数。

SIFT 特征是图像的局部特征，对旋转、尺度缩放、亮度变化保持不变性，对视角变化、放射变换、噪声也有一定的健壮性。

9.2.2 基于区域的图像配准

1. 相关法

相关法是指对于存在平移、旋转和尺度缩放的图像，利用图像间相似性最大化的原理实现配准，即通过优化相似性准则计算图像间的变换参数。相似性准则包括灰度差的平方、相关函数和归一化相关函数等。特征块匹配算法就是一种利用相关性准则的图像配准法。由于相关法是一种寻优的全搜索算法，计算量相当大。Kuglin 等人提出了相位相关法，大大降低了运算复杂度。

2. 对数极坐标变换方法

相位相关法给出了两幅图像之间的平移检测方法。后续的研究都致力于将图像的旋转和尺度关系转换为平移关系后利用相位相关法进行平移检测。Reddy 等人利用对数极坐标变换提出了一种频域相位相关的方法。

对数极坐标变换的关键点就是将存在旋转和尺度关系的图像通过对数极坐标变换变为该坐标系下的平移关系。其中极坐标用来解决旋转问题，对数变换用来解决尺度关系问题。借助对数极坐标变换和相位相关法能有效解决平移、旋转和尺度缩放问题，但是要求带配准的图像间有较高的重合度。

3. 最大互信息配准法

互信息是最常用的多模态图像相似性测度，由 Viola 和 Collignon 提出。该方法不需要对两种成像模式中图像强度之间的关系性质作任何假设，也不需要任何预处理，所以应用广泛。互信息用熵来定义，常用的是 Shannon 熵互信息。这里不作详细介绍。

9.3 最佳缝合线查找

完美配准以及曝光一致的图像之间的融合是一件非常惬意的事情，但是在实际情况中待融合的图像往往存在曝光差异、配准差异以及图像中存在运动物体，这些因素导致了图像间存在缝隙。为了更好地减轻图像融合的压力，需要在重叠区域中找到一条最佳的分界线。

缝合线是指图像经过配准之后，在图像重叠区域确定的一条分割线，拼接后的图像，在此分割线两侧的像素分别来源于左、右不同图像。如图 9-3 所示， Ω 是一条缝合线，拼接后图像在其左侧像素取值来源于 I_1 ，在其右侧像素取值则来源于 I_2 。缝合线查找就是在图像重叠区域确定这条分割线，使得这种分割最有利于后续的图像融合。显然，最优的缝合线就是重叠区域中两幅图像之间差异最小的一条切割线。

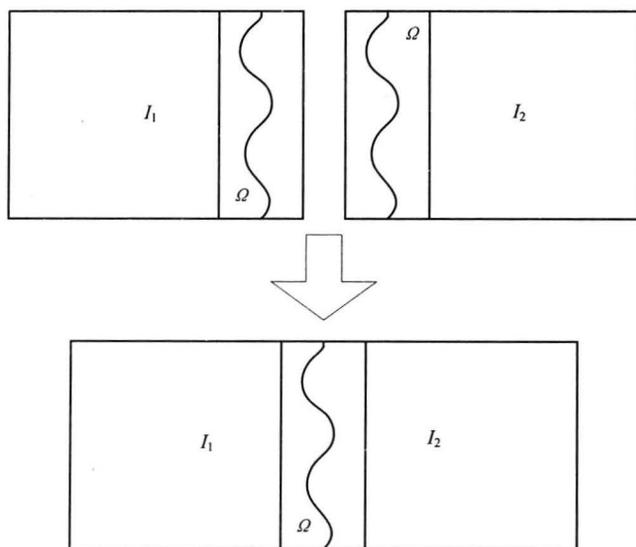


图 9-3 缝合线示意图

9.3.1 静态图像的最佳缝合线查找

静态图像拼接中缝合线的查找算法比较成熟，主要分为两大类，基于优化能量函数的算法以及基于计算最小差异的算法。

基于优化能量函数的算法。这类算法通过优化马尔科夫随机场 (Markov Random Field) 能量函数来得到最优缝合线。对于马尔科夫随机场优化问题，其求解过程是一个 NP 问题，近年来许多研究人员提出一系列近似优化方法来解决这个优化问题，包括 Graph Cuts 算法^[4~6]，模拟退火算法^[7]，Loopy Belief Propagation 算法^[8]。其中，Graph Cuts 是缝合线查找比较成熟，并且效果较好的算法。这个算法通过计算图像中最小分割来最小化能量函数项。

Graph Cuts，一般译为“图割”。在计算机视觉领域普遍应用于图像分割 (Image Segmentation)、立体视觉 (Stereo Vision)、抠图 (Image Matting) 等。Graph Cuts 理论最早是出现在网络优化领域的，比如说水管网络、通信传输网络、城市车流网络等。Graph Cuts 算法是指采用最大流-最小分割来优化的图模型。

先介绍 Graph Cuts 中的 Graph 和 Cuts 的意思。这里的 Graph 和普通的 Graph 稍有不同。普通的图由顶点和边构成，如果边有方向，这样的图则被称为有向图，否则为无向图，且边是有权值的，不同的边可以有不同权值，分别代表不同的物理意义。而 Graph Cuts 图是在普通图的基础上多了两个顶点，这两个顶点分别用符号“S”和“T”表示，统称为终端顶点。其他所有的顶点都必须和这两个顶点相连形成边集合中的一部分。所有 Graph Cuts 中的边也分为两种，一种是普通顶点与相邻点或者与之有关联的点之间的连线，一种是普通顶点和两个终端顶点之间的连线。Graph Cuts 中的 Cuts 是指这样一个边的集合，该集合中边的断开会导致残留“S”和“T”图分开，且这些边的集合中所有权值之和是最小的，因此也称为最小割。

一张图像可以看成是一个矩阵，图中的每个像素就是矩阵中的一个位置，这个像素

可以看做是图中的一个顶点。然后把相邻顶点之间用短线（有向或无向都行）连接起来，这就形成了边，现在这幅图像有顶点和边了，也就构成一幅 Graph 了。图像中边的权值一般由两部分构成，平滑项和数据项，这两部分的比重用一个系数来平衡，其中平滑项主要体现顶点像素和其相邻区域像素之间值的变化强度，如果变化剧烈说明这两者很有可能出于边缘部分，则被分割开的可能性比较大，而按照最小割的分割原理，这两者的平滑项权值应该较小。而数据项部分则表示对应顶点属于前景或者背景的惩罚项。基本公式如下：

$$E(A) = \lambda \cdot R(A) + B(A) \quad (9.7)$$

其中

$$R(A) = \sum_{p \in P} R_p(A_p) \quad (9.8)$$

$$B(A) = \sum B_{\{p,q\}} \cdot \delta(A_p, A_q) \quad (9.9)$$

$$\delta(A_p, A_q) = \begin{cases} 1 & \text{当 } A_p \neq A_q \text{ 时} \\ 0 & \text{其他} \end{cases} \quad (9.10)$$

式中， $R(A)$ 表示的是区域数据项； $B(A)$ 表示的是边界平滑项； $E(A)$ 表示的是权值，即损失函数，也叫能量函数，图割的目标就是优化能量函数使其值达到最小。

在图像重叠区域的最优缝合线查找，就是将重叠区域的图像分割成两个部分，其中的分割线就是图像拼接中的缝合线。对于 Graph Cuts 算法，有两个终端节点“S”和“T”，可以分别表示需要分割的两个区域，接下来根据缝合线查找的具体目标函数定义能量函数和约束项，通过 Graph Cuts 理论来不断迭代，实现对重叠区域进行分割，最终得到的“最小割”就是最优缝合线。

需要指出的是，图割是一种需要耗费大量计算时间同时需要较大的内存空间的算法，虽然许多研究工作都致力于提高图割算法的效率，例如将算法放到 GPU 上运行，并行运算或者分布式处理。这些加速虽然从一定程度上加快了算法的运行，但是由于算法本身具有随机读取数据，需要开辟大的内存空间存储数据等特点，决定了其不能达到实时计算缝合线的要求，也不能很好地在硬件上实现。其较高的计算耗时以及内存需求成为其向视频拼接发展的瓶颈。

9.3.2 视频拼接的最佳缝合线查找

利用多摄像机之间的图像拼接技术，来获取大视场和超高分辨率视频是一个研究热点。与传统的静态图像拼接类似，视频拼接技术也包括配准、缝合线查找以及融合等步骤。

与静态图像拼接中的缝合线查找不同，视频拼接中缝合线的查找需要考虑算法实时性和帧间画面一致性约束。实时性是指缝合线查找算法计算速度要快，能够满足视频实时播放的要求。帧间画面一致性是指为了保证相邻帧画面一致性，帧间缝合线不应剧烈变化。这就要求在查找最佳缝合线时，除了将差异最小作为优化准则之外，还需加入帧间缝合线差异作为约束条件。

这里提出一种基于前后帧约束的视频缝合线实时查找方法。该方法首先在梯度域下，计算重叠区域中梯度光滑性与梯度相似性，并用这两个参数来描述两幅图像重叠部分的差异大小，作为缝合线优化准则。然后，提出一种改进的动态规划算法来查找最优缝合线，并在查找时加入帧间画面一致性约束条件。

1) 差异描述子

采用计算两幅图像重叠区域梯度域下的光滑性与相似性,来描述图像之间的差异^[9]。

梯度光滑性,描述图像中的梯度光滑程度,也就是图像的梯度变化程度,梯度光滑性函数 S_m 定义如下:

$$S_m(p, q) = \|\nabla I_1(p)\| + \|\nabla I_1(q)\| + \|\nabla I_2(p)\| + \|\nabla I_2(q)\| \quad (9.11)$$

式中, $\|\nabla * \|\$ 表示图像像素的梯度; I_1 和 I_2 表示两幅图像; p, q 表示图像像素坐标。

S_m 将图像的梯度平滑程度引入到代价函数的计算中,在缝合线查找过程中,可以有效避免缝合线经过物体的边缘,保持重叠区域中物体的完整性。

梯度相似性,描述两幅图像重叠区域对应像素点梯度的相似性,梯度相似性函数 S_d 的定义如下:

$$S_d(p, q) = \|\nabla_x I_1(p) - \nabla_x I_2(p)\| + \|\nabla_x I_1(q) - \nabla_x I_2(q)\| + \|\nabla_y I_1(p) - \nabla_y I_2(p)\| + \|\nabla_y I_1(q) - \nabla_y I_2(q)\| \quad (9.12)$$

式中, S_d 表示重叠区域 Ω 中,图像 I_1 与图像 I_2 对应像素点在梯度域上的差异; p, q 表示图像像素坐标; $\|\nabla_x * \|\$ 表示 x 方向梯度; $\|\nabla_y * \|\$ 表示 y 方向梯度。

梯度相似性函数的引入,目的是让缝合线计算过程中,经过两幅图像梯度相似度高的地方,也就是两幅图像重叠区域中差异小的区域,使得缝合线两端的图像能够最平滑地过渡,从而有利于后续的图像融合。

然后利用加权和将两个权值函数项结合起来,定义梯度代价函数 $S(p, q)$,表示 RGB 三个颜色通道的梯度值之和:

$$S(p, q) = \sum_{\text{RGB}} ((1 - \beta)S_m + \beta S_d) \quad (9.13)$$

式中, β 是平衡两个度量函数相对影响力的因子; S_m 表示梯度光滑性函数; S_d 表示梯度相似性函数。在最优缝合线计算的计算过程中,首要考虑的因素是经过两幅图像中相似度高的地方,其次考虑缝合线尽量不要经过物体的边缘。

2) 改进动态规划算法查找最优缝合线

通过计算图像重叠区域的梯度代价函数可以得到差异矩阵,最优缝合线查找等价于在差异矩阵中寻找一条代价最小的分割线,如图 9-4 所示。

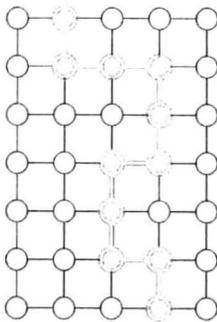


图 9-4 最优缝合线查找算法的图模型(粗实线表示计算得到的最优缝合线)

显然,上述问题是一个图像分割问题,图像分割有许多经典的方法,包括模拟退火、Graph Cuts 等,也有将上述问题抽象成寻找最优路径的过程,运用 Dijkstra, 动态规划算

法等寻找最优路径算法来实现最优缝合线的查找。

动态规划算法是一个分阶段求解优化问题的方法，在求最优缝合线的过程中，计算速度快，可以达到实时处理的要求，虽然动态规划算法求解得到的是局部最优解，但是缝合线的局部最优解也可以得到不错的拼接效果。特别是，动态规划的算法结构适于硬件实现。

虽然动态规划在计算时间上有优势，但是由于传统的动态规划算法在寻找最优缝合线的过程中，搜索方向受到限制，导致在某些情况下动态规划不能找到令人满意的缝合线。传统的动态规划算法寻找最优缝合线的过程中，每次决策只搜索下一行与当前位置相邻的三个位置，其递推关系如式 (9.14) 所示，搜索方向如图 9-5 所示。

$$E_{i,j} = e_{i,j} + \min(E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1}) \tag{9.14}$$

式中， $E_{i,j}$ 表示从起点到位置 (i, j) 所经过的位置的差异值之代数和； $e_{i,j}$ 表示位置 (i, j) 处的差异值。

搜索方向的限制，寻找的最优缝合线可能会经过物体的边缘，这极大影响了图像拼接效果，是需要极力避免的。为了克服这类问题，提出一种改进的动态规划算法，改进的动态规划算法，如图 9-5 所示，不仅可以比较搜索下一行与当前位置相邻的三个点，还可以搜索同一行与当前位置相邻的两个点，扩大了搜索范围。其递推关系如下：

$$E_{i,j} = e_{i,j} + \min(E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1}, E_{i,j+1}) \tag{9.15}$$

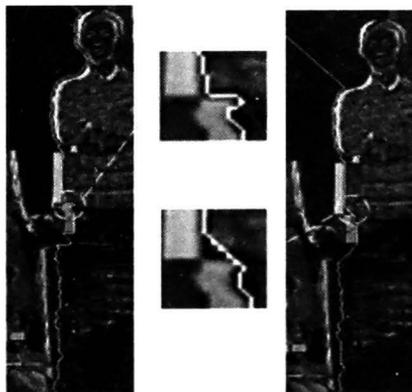
式中， $E_{i,j}$ 表示从起点到位置 (i, j) 所经过的位置的差异值之代数和； $e_{i,j}$ 表示位置 (i, j) 处的差异值。



(a) 传统的动态规划算法下阶段搜索方向 (b) 改进的动态规划算法下阶段搜索方向

图 9-5 传统动态规划算法与改进动态规划算法搜索方向对比

改进的动态规划算法可以让缝合线绕过差异较大的大片区域，一定程度上提高了拼接效果，如图 9-6 所示。



(a) 改进的动态规划算法 (b) 传统的动态规划算法

图 9-6 改进的动态规划算法与传统动态规划算法的效果对比

3) 帧间画面一致性

帧间画面一致性保证了视频拼接的实时性,视频图像帧之间的一致性 是视频拼接需要关注的另外一个问题。视频帧间一致性是指视频相邻图像帧之间差异小,不存在剧烈变化,不出现画面跳变。静态图像拼接由于处理静态图像,不需要考虑图像前后连贯,与图像拼接缝合线查找算法相比,视频拼接中的缝合线查找算法则需要考虑一致性的问题。如果视频拼接中缝合线查找算法不引入一致性约束,那么视频前后帧图像中查找到的最优缝合线可能出现位置上的跳变,造成拼接后的视频前后图像帧出现剧烈变化,视频播放不流畅,影响视频拼接质量,如图 9-7 所示。



(a) 前一帧视频拼接图像(左边为计算得到的缝合线,右边为拼接得到的图像) (b) 后一帧视频拼接图像(左边为计算得到的缝合线,右边为拼接得到的图像)

图 9-7 视频图像中前后两帧图像缝合线与图像拼接效果对比

针对上述问题,提出一种利用帧间信息的约束方法,来实现缝合线查找的一致性约束。该方法的基本思想是:利用前一帧计算到的最优缝合线信息,生成一个一致性约束矩阵,结合当前帧所计算得到的梯度差异,计算当前帧的最优缝合线。因此一致性约束矩阵的设计,是该方法的关键。

缝合线的一致性,要求前后帧图像的缝合线位置变化不大,这也是保证视频图像前后帧一致性的内在要求。采用计算帧间缝合线偏差,利用与缝合线水平偏移距离作为约束矩阵的生成标准,来实现视频拼接的一致性约束。

按照一致性约束的要求,不希望前后两帧视频图像的缝合线偏离太远,也就是说一致性约束将极力避免前后两帧图像的缝合线偏离太远。这里引入一致性矩阵,来约束缝合线的位置。前后帧缝合线位置偏移越大,惩罚因子就要越大,因此约束矩阵中惩罚因子与前后帧缝合线的水平偏移距离是一个正相关的关系,而且距离缝合线越远的点,约束矩阵的值应该越大。对指数、线性以及幂函数三种正相关的约束进行了实验比较,如图 9-8 所示。其中图 (a) 所示为前一帧图像,计算得到的缝合线;图 (b) 所示为当前帧图像,利用指数函数关系约束,计算得到的缝合线;图 (c) 所示为当前帧图像,利用线性函数关系约束,计算得到的缝合线;图 (d) 所示为当前帧图像,利用幂函数函数关系约束,计算得到的缝合线。从实验结果比较中,采用指数函数关系,来实现缝合线一致性约束。

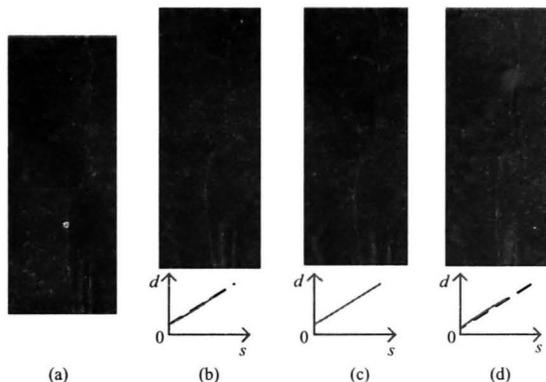


图 9-8 三种约束关系下查找得到的最优缝合线

定义重叠区域中 Ω ，每个像素距离缝合线的约束矩阵 $D(p, q)$ ，其中 p 和 q 表示像素坐标，约束矩阵 D 的维数与重叠区域 Ω 的大小一致。

一致性约束矩阵 D 的计算定义为

$$D(p, q) = Ae^{bd} \tag{9.16}$$

A 与 b 为参数, d 表示约束矩阵中像素点 (p, q) 距离前一帧图像缝合线的水平距离, 定义为

$$d = \|(p - p_0) | q\| \tag{9.17}$$

式中, (p_0, q) 表示前一帧图像缝合线的坐标位置。

计算出一致性约束矩阵之后, 利用这个约束矩阵来实现视频拼接的帧间一致性, 并且结合当前图像帧重叠区域, 计算最优缝合线。采用加权求和的方法, 将一致性约束矩阵与图 9-8 (a) 中计算的差异矩阵结合。

$$E = S(p, q) + \alpha D(p, q) \tag{9.18}$$

式中, α 为参数, 平衡图像差异因子和一致性约束因子两个权值的影响。

一致性矩阵更新了图像差异矩阵, 重新计算了查找最优缝合线的数据基础, 考虑了更加全面的因素, 完善了缝合线查找的约束, 可以找到合理的、适合视频拼接的缝合线。

4) 实验效果

采用三路视频拼接装置进行实验, 如图 9-9 所示。装置由三个相同的普通摄像头组成, 固定在一个水平支架上, 每个摄像头之间夹角为 60° , 每路摄像头采集的视频图像分辨率为 1280×720 像素。

在如图 9-9 所示的实验平台上, 采集三路原始视频图像。为了更好地说明问题, 针对其中两路视频的图像数据做分析, 分别用 Graph Cuts、模拟退火算法以及提出的算法进行缝合线查找。实验结果比较如图 9-10 所示。

从效果上看, 提出的算法同 Graph Cuts 以及模拟退火算法所查找的最优缝合线都可以有效地绕开差异较大的区域, 避免拼接过程中出现鬼影等影响。为了对实验数据进行进一步对比, 定义了缝合线最优性指标以及算法计算耗时指标来定量对实验结果进行说明。其中最优化指标用来评价缝合线的优劣程度, 计算耗时指标用来评价缝合线计算耗时。

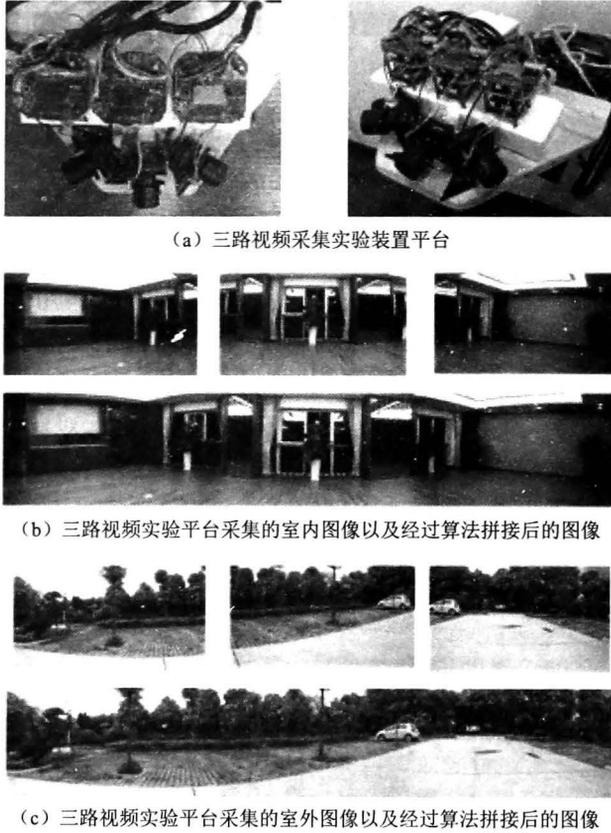


图 9-9 实验装置平台与拼接效果

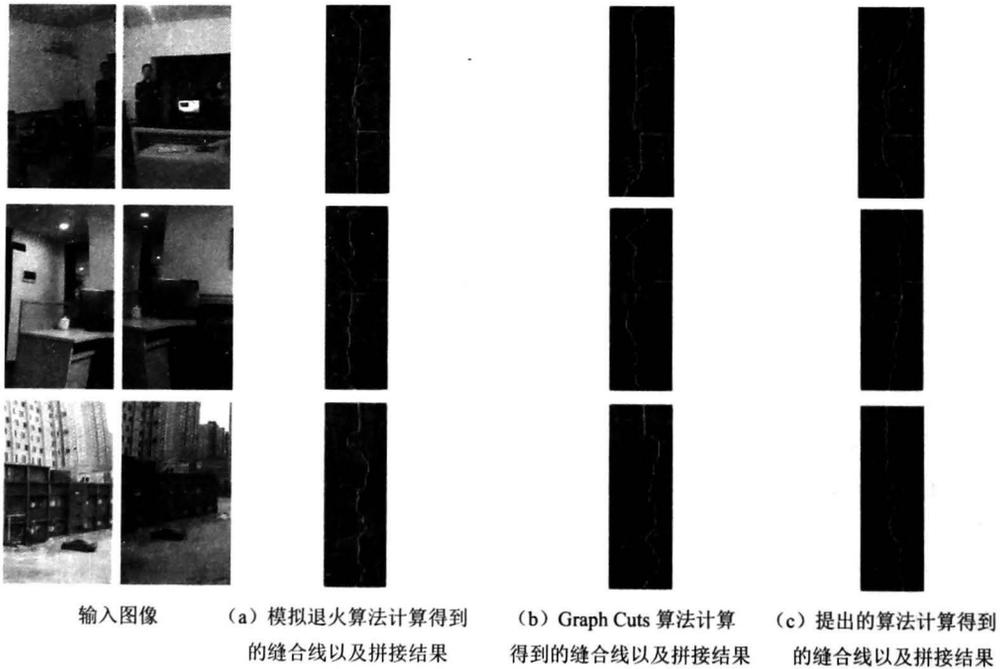


图 9-10 三种缝合线算法比较

定义缝合线的最优性指标 O 为缝合线经过像素的梯度差异之和:

$$O = \sum_{i \in \text{seam}} E(p_i, q_i) \tag{9.19}$$

式中, (p_i, q_i) 表示缝合线经过的第 i 个像素位置; E 表示当前位置的梯度差异。

定义缝合线计算耗时指标 T 为缝合线查找耗时:

$$T = t_{\text{cost}} \tag{9.20}$$

式中, t_{cost} 表示计算得到缝合线所消耗的时间。

对三个算法在 PC 端的计算耗时做比较, 在 Intel® Core i5-3470 的微处理器中, CPU 主频为 3.20GHz, 内存容量为 4GB 的计算机, 在 Matlab R2009a 环境上运行三种算法, 对采集到的 80 组图像实验数据进行处理比较, 每组数据包括三幅图像, 两个重叠区域。对实验的缝合线最优性指标 O 以及计算耗时 t 进行对比, 如图 9-11 所示。从图中可以看出, 三种算法的缝合线最优性的指标 O 相差不大, 三种算法的效果基本一样, 但是对于耗时 t , Graph Cuts 和模拟退火算法的耗时比提出的算法要多两个数量级。

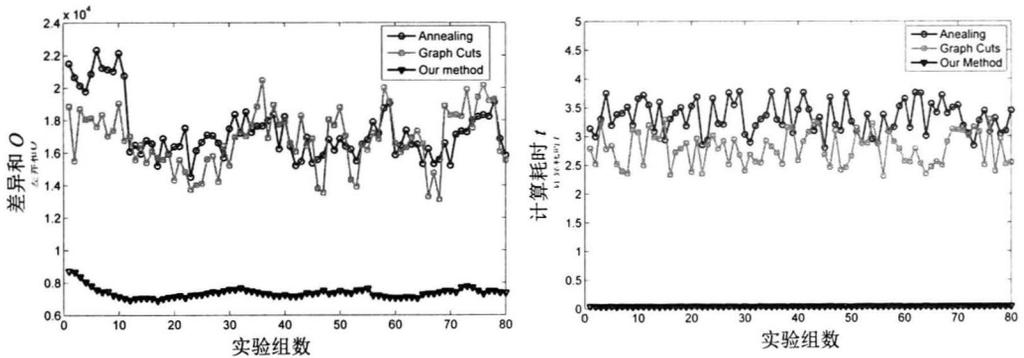


图 9-11 80 组图像的缝合线最优性指标 O 与计算耗时 t 数据统计图

对三种算法的耗时进行统计与分析, 从图 9-11 中可以看出, Graph Cuts 与模拟退火算法进行最优缝合线查找的平均计算耗时分别为 2.5s 和 3.1s, 这远远达不到视频实时处理的要求。因为 Graph Cuts 需要较大的空间开销和计算, 而模拟退火算法需要迭代运算, 导致这两个经典算法的计算耗时大。提出的算法平均耗时为 0.04s, 视频实时处理需要每秒处理 25~30 帧, 每帧计算时间为 40ms 左右, 因此提出的算法可以满足视频实时处理的要求。

对于视频数据, 相邻两帧图像不是独立的, 存在相互关联性。本实验平台采集了四组三路视频数据进行处理, 得到一个时间序列上的缝合线数据, 以及拼接后的视频图像, 如图 9-12 所示。图 9-12 显示了连续 40 帧视频图像帧拼接后的结果以及连续 40 帧缝合线结果。实验结果中, 缝合线在绕开重叠区域梯度较大的同时, 还保持了前后帧之间的关联。

为了定量说明视频图像帧的前后关联程度, 采用拼接后得到的视频图像前后帧的重叠区域的差异来描述关联程度, 差异小, 说明视频图像保持有较好的连续性。

定义拼接后得到的视频图像帧的重叠区域的差异值 D 为

$$D = \sum_{p, q} (\|I_{i+1}(p, q) - I_i(p, q)\|) \tag{9.21}$$

式中, I_i 表示第 i 帧视频图像; (p, q) 表示重叠区域像素坐标位置; $\|\bullet\|_1$ 表示 1 范数。

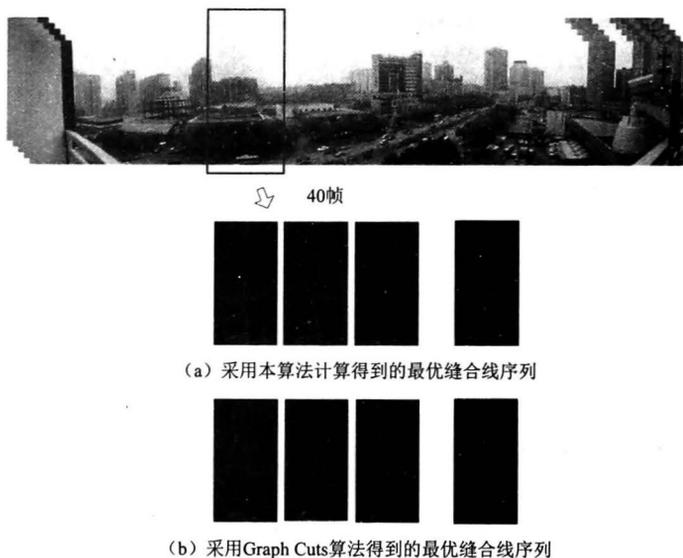


图 9-12 三路视频采集拼接平台采集、拼接后的连续 40 帧 3840×720 像素的全景图像以及对应 40 帧图像的重叠区域

采用差异指标 t , 对采集到的四路视频数据利用不同算法进行拼接, 并且对拼接得到的视频图像计算前后帧差异, 进行分析, 如图 9-13 所示。

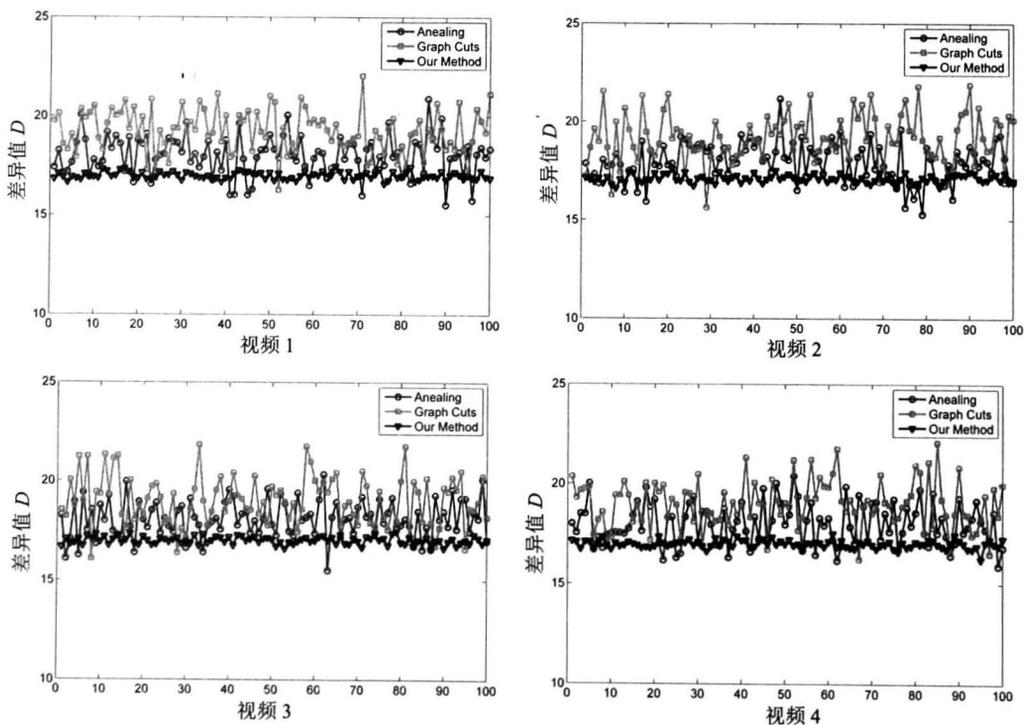


图 9-13 利用三种不同算法拼接四路视频数据, 得到的视频图像前后帧差异统计图

从图 9-13 中可以看出, 加入前后帧一致性约束算法后, 帧间差异曲线总体平缓, 没有剧烈跳变, 有效约束了视频图像前后帧, 保证了视频图像的一致性。而没有加入帧间约束方法的两个算法的帧间差异曲线变化比较剧烈, 视频没有较好的一致性。进一步, 对采用不同算法所得到的帧间差异值的数值特征进行计算分析, 如表 9-2 所示。

表 9-2 四路视频数据帧间差异值的数值分析

算法	Graph Cuts		Annealing		本书算法	
	均值	方差	均值	方差	均值	方差
视频 1	18.9942	1.0839	17.9202	1.0116	16.9427	0.1147
视频 2	18.0620	0.9449	18.6201	0.9972	17.0732	0.1051
视频 3	18.5650	0.9349	17.4325	0.7140	17.0023	0.1085
视频 4	17.3342	0.9827	17.8958	1.3430	16.7893	0.0979

从表 9-2 中的数据可以看出, 加入了帧间一致性约束算法之后, 视频相邻帧之间的差异值的均值和方差明显减小, 说明视频相邻帧之间的变化减小, 有效实现了视频帧间的一致性, 保证了视频播放的连续性。

参 考 文 献

- [1] Szeliski R, et al. Video mosaics for virtual environments[J]. Computer Graphics and Applications, 1996,16: 22-30.
- [2] Harris C, Stephens M. A combined corner and edge detector[C]// The Alvey Vision conference, April 4-7, 1988 Manchester. London: BMVC, 1988: 50-52.
- [3] Lowe D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60: 91-110.
- [4] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26:1124-1137.
- [5] Boykov Y, et al. Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23:1222-1239.
- [6] Boykov Y, Jolly M. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images[C]// 8th IEEE International Conference on Computer Vision, July7-14, 2001, Vancouver. New York: IEEE Computer Society, 2001: 105-112.
- [7] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984,12: 721-741.
- [8] Sun J, et al. Stereo matching using belief propagation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25: 787-800.
- [9] Jia J, Tang C. Eliminating structure and intensity misalignment in image stitching[C]// 10th IEEE International Conference on Computer Vision, October 17-20,2005, Beijing. New York: IEEE Computer Society, 2005: 1651-1658.

第 10 章 图像融合技术

10.1 概 述

图像融合 (Image Fusion) 是指综合多幅输入图像的信息, 以获得更高质量输出图像的过程。与融合前的多幅图像相比, 融合后的图像应更加适于进一步的观察或处理。图像融合属于数字图像处理领域, 在计算摄影学中, 用于拍摄所得图像的后期处理。

图像融合的作用主要包括: 扩展图像在空间域或时间域的覆盖范围, 减少信息冗余和不确定性, 压缩存储空间等^[1]。通常情况下, 输入图像由同一台照相机在不同时间或者不同照相机在同一时间采集, 输出图像可能反映真实的客观场景, 如图 10-1 所示, 也可能是利用客观场景中的素材构建的虚拟场景, 如图 10-2 所示。

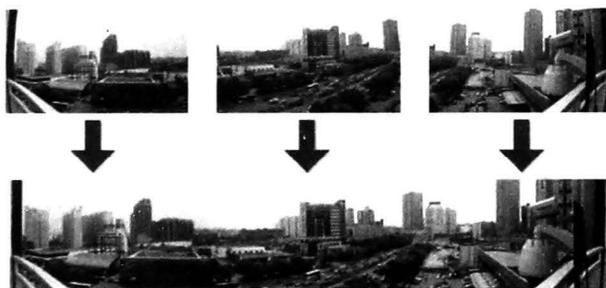


图 10-1 反映真实客观场景的图像融合



图 10-2 图像融合构建的虚拟场景

按照数据抽象的层次, 图像融合可分为三级, 即像素级融合、特征级融合和决策级融合^[2]。其中, 像素级融合是指直接对原始图像数据进行处理而获得融合图像的过程; 特征级融合对原始图像进行特征提取, 并在融合过程中利用和保留特征信息; 决策级融合考虑主观的要求, 并为基于图像内容的决策提供依据。

像素级融合是高层次图像融合的基础, 其最大的优点是能够保留尽量多的原始数据, 有利于图像的进一步处理。计算摄影中涉及的图像融合技术通常属于像素级融合, 主要应用领域包括全景图像拼接, 图像超分辨率, 超现实图像编辑等。本章所讲的图像融合主要针对全景图像拼接。

现有的图像融合算法主要有加权平均融合法、金字塔融合法、梯度域融合法、小波变换法以及结构变形等^[3]。本章将分节介绍几种典型的图像融合方法并对各种方法的特性和融合效果进行介绍。

10.2 加权平均融合法

对图像融合最直观的理解就是要对有重叠区域的输入图像进行平滑过渡, 即选取一

个合适的过渡权值函数对重叠区域的像素进行加权平均。

考虑两幅图像的融合，设输入图像为 I_1 、 I_2 ，融合图像 I_F 可由下式计算得到：

$$I_F(h, w) = w_1(h, w)I_1(h, w) + w_2(h, w)I_2(h, w) \quad (10.1)$$

式中， w_1 、 w_2 分别为 I_1 、 I_2 对应的权值函数，且有 $w_1 + w_2 = 1$ 。

权值函数的取值可以有多种形式，如线性权值、高斯分布函数、Sigmoid 函数等。对于如图 10-3 (a)、图 10-3 (b) 所示的两幅输入图像，分别采用上述各权值函数进行融合。各种权值函数的图像及其融合效果如图 10-4 所示。

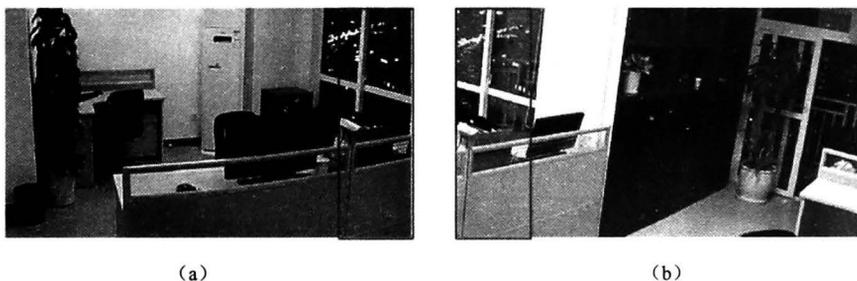


图 10-3 输入图像（红色线框内为其公共区域）

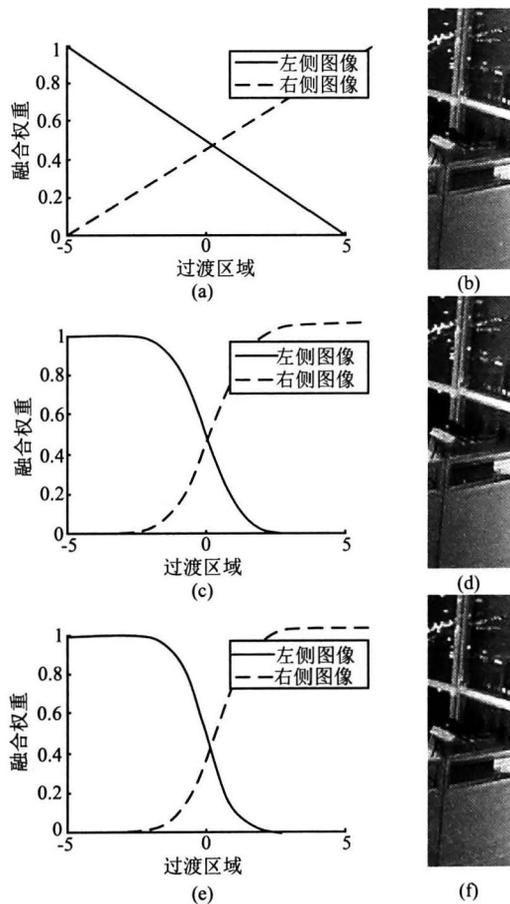


图 10-4 各种权值函数图像及其融合效果

在如图 10-4 所示融合效果中, 图 10-4 (d)、图 10-4 (f) 的融合效果相对较好, 因此在实际应用中常采用高斯分布函数和 Sigmoid 函数作为融合函数。高斯分布函数的计算涉及数值积分, 计算量较大, 通常采用查表的方式实现; Sigmoid 函数的具体形式为 $\lambda(x) = \frac{1}{1 + e^{\alpha x}}$, 相对高斯分布函数, 计算较为简单, 但需要实验确定形状参数 α 的具体取值。加权平均融合法作为早期且较初级的融合方法, 其优势是思路简单, 易于实现且速度较快, 但对于由视差和亮度差异造成的输入图像间的不一致, 该方法并不能很好地解决, 在融合图像中容易出现重影和错位等问题。

10.3 金字塔融合法

使用传统的融合算法, 在图像融合过程中, 最常见的两个问题分别是突变和重影, 其中, 突变对应大尺度特征的融合, 如图 10-5 (a) 所示, 当融合跨度不足以平滑背景色的差异时就会出现突变现象; 重影对应小尺度特征的融合, 如图 10-5 (b) 所示, 当融合跨度远大于错位边界的纹理时, 就会出现重影现象。

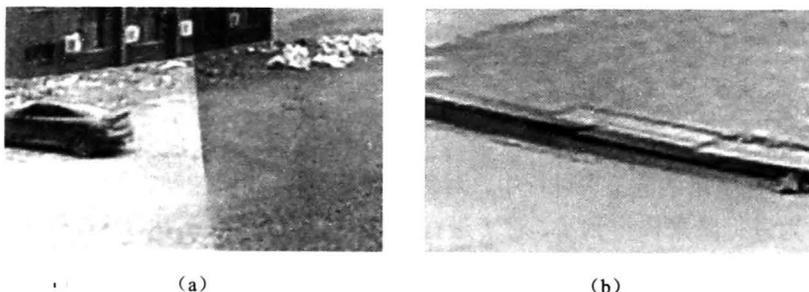


图 10-5 突变与重影

为解决上述问题, 对不同尺度的特征, 应该采用不同的平滑过渡宽度, 可以通过建立图像的拉普拉斯与高斯金字塔 (Laplace & Gaussian Pyramid) 实现^[4, 5]。

10.3.1 算法思想

Laplace 金字塔的每一层实际上是对图像进行带通滤波并降采样的结果, 表示原图像在不同尺度上的信息。底层图像的频率高、采样间隔小, 表示图像的小尺度特征; 高层图像的频率低, 采样间隔大, 表示图像的大尺度特征^[4, 5]。金字塔融合的算法思想主要包括以下几点。

- (1) 采用带通滤波对图像进行分频处理。
- (2) 对不同频率的图像分别进行加权平均融合, 融合函数为高斯分布函数。
- (3) 合成各频率的图像融合结果得到最终的融合图像。

金字塔融合的过程即对不同尺度下的特征采用不同的融合跨度, 以尽可能避免突变和重影的出现。

10.3.2 算法流程

1. 图像拉普拉斯与高斯金字塔的建立

通过对原始图像的低通滤波与降采样操作可以建立其高斯（Gaussian）金字塔，而拉普拉斯（Laplace）金字塔的每一层则由该层高斯金字塔图像与滤波后图像做差得到，如图 10-6 所示，各层滤波前的图像 f_0 、 f_1 、 f_2 组成原图的高斯金字塔，各层差值图像 h_0 、 h_1 与最低频图像 f_2 组成原图的拉普拉斯金字塔。

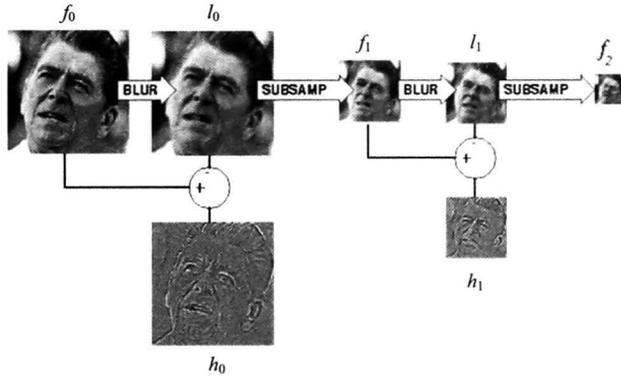


图 10-6 拉普拉斯金字塔的建立

拉普拉斯金字塔的建立^[6]过程中，对原图像 f_0 进行低通滤波得到 l_0 ， h_0 为 f_0 与 l_0 的差值，对 l_0 为进行降采样得到 f_1 ，对 f_1 进行低通滤波得到 l_1 ， h_1 为 f_1 与 l_1 的差值，对 l_1 为进行降采样得到最低频图像 f_2 。

2. 利用拉普拉斯和高斯金字塔对图像进行复原

图像复原是金字塔建立的反操作。如图 10-7 所示，由高层到底层分别对图像进行插值与滤波，并与下一层的拉普拉斯金字塔图像相加，最终可得到全频率的原图像 f_0 。图中，对最低频图像 f_2 进行插值，得到 l_1 ， f_1 为 l_1 与 h_1 的和，对 f_1 进行插值得到 l_0 ，原图像 f_0 为 l_0 与 h_0 的和。

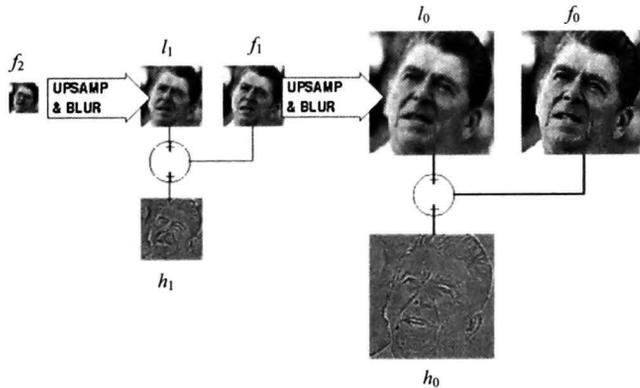


图 10-7 图像复原^[6]

上述金字塔建立于图像复原过程中,图像的低通滤波可以通过图像与高斯核的卷积操作实现。一个层数为 3 的拉普拉斯金字塔如图 10-8 所示。

对于已得到的最优拼接缝,利用 Flooding 算法得到左侧图像相对于融合图像的权值矩阵,右侧图像的权值矩阵可通过 1 减去左侧图像的权值矩阵得到。一个具体的权值矩阵图像如图 10-9 所示,其中,白色部分表示权值为 1,黑色部分表示权值为 0。

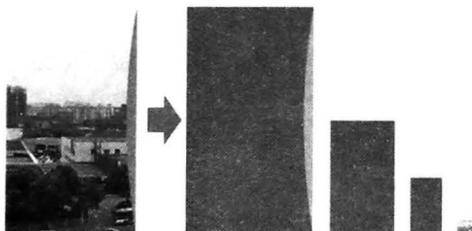


图 10-8 层数为 3 的拉普拉斯金字塔

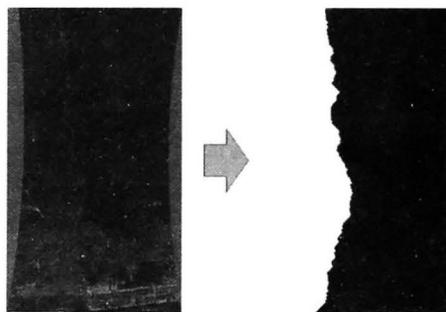


图 10-9 权值矩阵的生成

建立左侧图像权值矩阵的高斯金字塔,其各层数值对应于图像拉普拉斯金字塔各层的融合权值,由于建立金字塔时的高斯滤波操作,在高斯金字塔高层会出现非整数权值,其几何意义为拼接缝附近图像的平滑过渡。

融合后,图像的拉普拉斯金字塔通过对左右两侧图像拉普拉斯金字塔的加权平均得到,最后利用得到的拉普拉斯金字塔对图像进行复原,得到最终的融合图像。

10.3.3 融合结果

对于图 10-10 (a)、图 10-10 (b) 所示输入图像,采用金字塔融合所得公共区域图像如图 10-10 (c) 所示,整幅融合图像如图 10-10 (d) 所示。

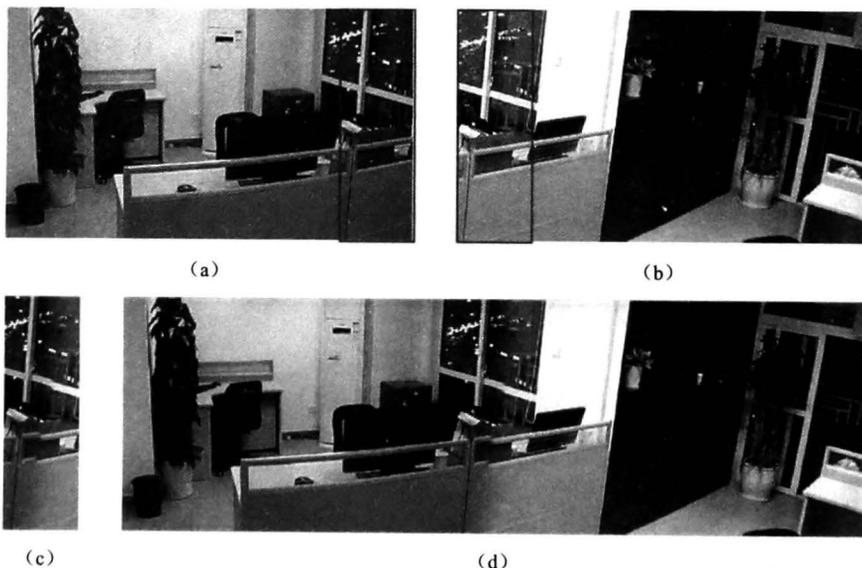


图 10-10 金字塔融合效果

金字塔融合基于多尺度分频处理的思想，在融合图像中尽量避免突变和重影的出现，其融合效果相对于加权平均法有较大的改善，但对于由结构偏差引起的断裂和错位，该方法并不能有效予以修正。

10.4 梯度域融合法

基于梯度域的计算作为一种图像处理方法已广泛应用于高动态范围图像压缩^[7]，图像修复^[8]、图像融合^[9~11]等领域。其中，梯度域融合方法能够在实现色彩平滑过渡的同时完整保留图像的结构信息，且不要求输入图像必须有重叠区域，相对于金字塔融合具有更广泛的应用前景。

10.4.1 算法思想

考虑两幅图像融合的情况，取其中一幅输入图像为目标图像，另一幅图像为源图像，如图 10-11 所示，将原图像 f 融合到目标图像 f^* 的某一区域，设该区域为 Ω ，梯度域融合基于一条已知的边界，即区域 Ω 的边界 $\partial\Omega$ 。

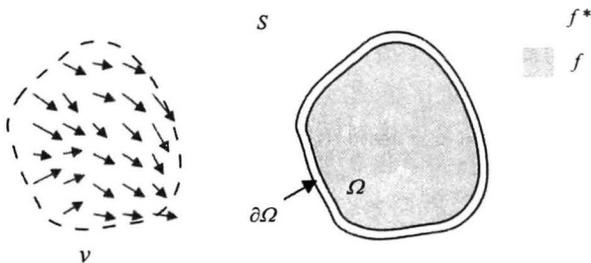


图 10-11 梯度域融合^[9]

设原图像在高度和宽度方向上的梯度图数据为 $v = (G_h, G_w)$ 。基于梯度域的融合就是根据 (G_h, G_w) 与已知的重叠区域边界 $\partial\Omega$ 求解整个 Ω 图像的过程。

10.4.2 算法流程

梯度域融合问题可抽象为下述 Poisson 方程的求解：

$$\begin{cases} \Delta f = \frac{\partial G_h}{\partial h} + \frac{\partial G_w}{\partial w} \\ f|_{\partial\Omega} = f^*|_{\partial\Omega} \end{cases} \quad (10.2)$$

式中， $f^*|_{\partial\Omega}$ 为目标图像在边界处的取值。式 (10.2) 的离散形式等价于如下最小化问题：

$$\min_f \sum_{\substack{p \in \Omega - \partial\Omega \\ q \in N_p}} (f(p) - f(q) - v_{pq})^2 \quad (10.3)$$

式中， f 为定义在 Ω 上任意可能的图像； p, q 为 Ω 内两点； N_p 表示 p 的四邻域； v_{pq} 表示从 p 到 q 的梯度，其具体数值可由下式计算：

$$v_{pq} = \begin{cases} \frac{G_w(p) + G_w(q)}{2} & h_p = h_q, w_p + 1 = w_q \\ \frac{G_w(p) + G_w(q)}{2} & h_p = h_q, w_p = w_q + 1 \\ \frac{G_h(p) + G_h(q)}{2} & h_p + 1 = h_q, w_p = w_q \\ \frac{G_h(p) + G_h(q)}{2} & h_p = h_q + 1, w_p = w_q \end{cases} \quad (10.4)$$

式 (10.2) 可进一步转化为如下线性方程组^[9]:

$$|N_p|f(p) - \sum_{q \in N_p \cap (\Omega - \partial\Omega)} f(q) = \sum_{q \in N_p \cap \partial\Omega} f(q) + \sum_{q \in N_p} v_{pq} \quad p \in \Omega - \partial\Omega \quad (10.5)$$

式中, $|N_p|$ 为与点 p 相邻的点数, 在四邻域情况下 $|N_p| = 4$ 。式 (10.5) 的系数矩阵是一个大型、稀疏、对称并且正定的方阵, 可选择收敛速度较快的序列超松弛迭代法^[12]对该问题进行求解。

10.4.3 融合结果

对于图 10-12 (a)、图 10-12 (b) 所示输入图像, 采用金字塔融合所得公共区域图像如图 10-12 (c) 所示, 整幅融合图像如图 10-12 (d) 所示。

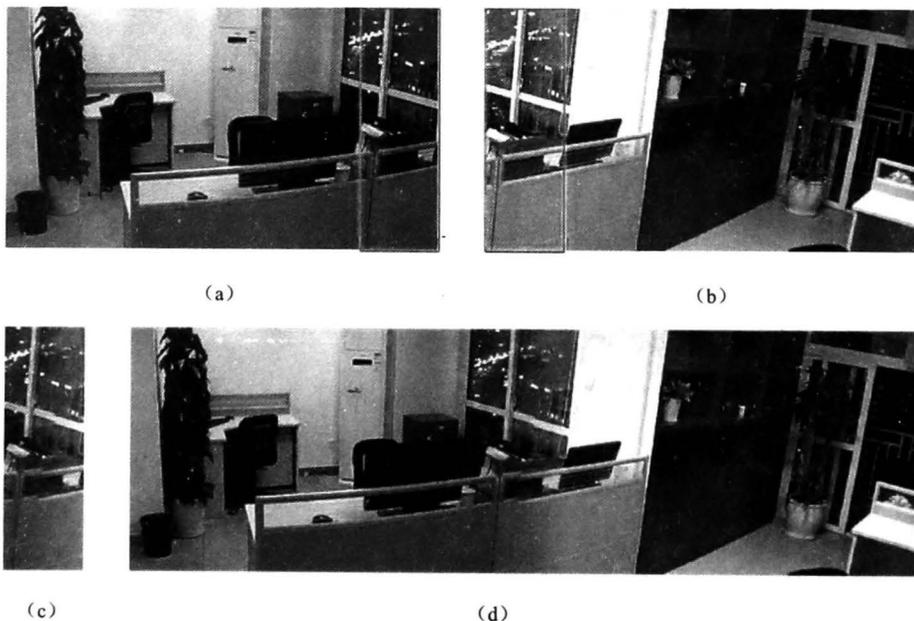


图 10-12 梯度域融合效果

与加权平均融合法和金字塔融合相比, 梯度域融合具有不需要输入图像重叠的优势, 但其局限是需要迭代求解, 选取合适的迭代初值可以有效减少收敛所需迭代次数, 实际操作中迭代初值可取为求解区域的源图像或其他容易获得的近似解。

与金字塔融合类似, 梯度域融合可以较好地修正输入图像间的亮度差异, 但对于由视差或配准误差造成的断裂和错位, 该方法依然不能有效解决。

10.5 结构变形

拍摄时照相机曝光参数的差异会造成输入图像间的亮度偏差，同样地，视差的存在也会影响输入图像间的一致性，拼缝两侧的物体边缘可能会出现断裂和错位的现象。为简便起见，称这种由视差引起的结构不一致现象为结构偏差。图 10-13 (a)、图 10-13 (b) 为两幅待融合的输入图像，若将其重叠区域简单粘贴，得到的效果如图 10-13 (c)、图 10-13 (d) 所示，其中亮度和结构的偏差都很明显。金字塔融合法和梯度域融合法可以较好地修正输入图像间的亮度差异，但对于输入图像在重叠区域的结构偏差，这些方法通常无能为力。

结构变形方法^[12, 13]通过在图像中引入结构约束，据此对图像进行与相机模型无关的空间变换，能够保证输入图像在拼缝附近的结构一致性，其处理效果如图 10-13 (e) 所示。

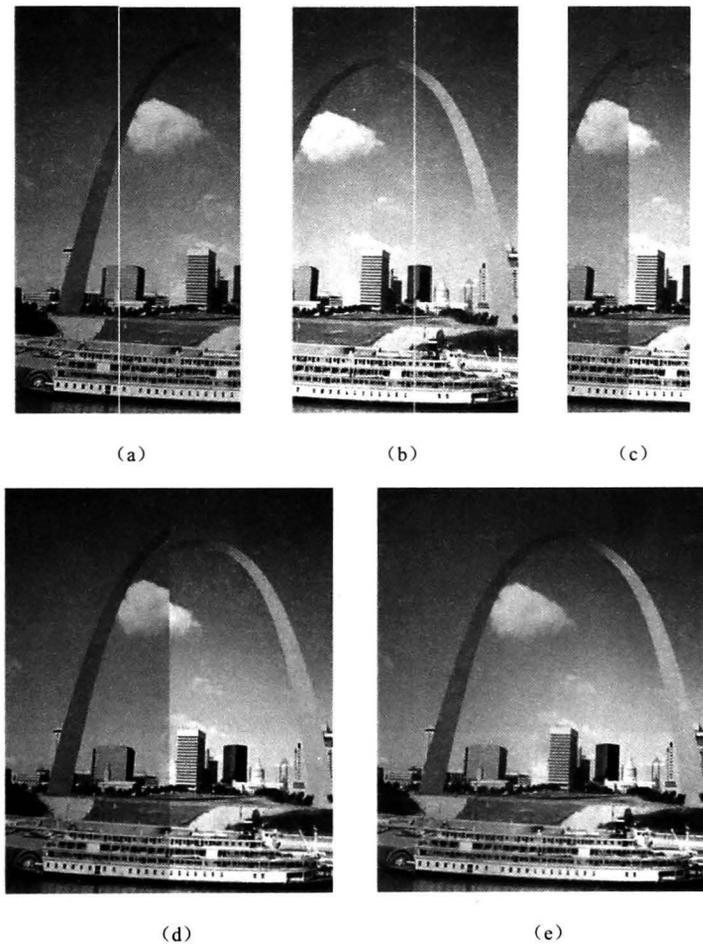


图 10-13 亮度和结构偏差以及结构变形处理效果

10.5.1 算法思想

结构变形根据图像结构信息对拼接缝附近的输入图像进行形变以解决传统方法无法解决的结构偏差问题。首先通过沿拼缝方向的一维特征点提取与匹配获得拼缝两侧断裂物体边缘的连接关系，并据此对整个重叠区域的图像作变形，使结构偏差得到修正，而输入图形间亮度偏差的修正则通过变形后图像数据的梯度域融合实现。具体可分为下述四个步骤。

(1) 最优拼缝查找。设定合适的目标函数和搜索策略，在重叠区域的两幅图像内查找最优拼缝。

(2) 一维特征点提取与匹配。定义适当的特征描述值，沿着拼缝方向进行一维特征点的提取，找到物体边界的连接关系。

(3) 形变的量化与传播。将拼缝上每一对匹配的特征点移动到重合位置，以形变向量的形式记录每一个特征点的位移量，并以一定的策略将形变传播到图像的其余部分，使图像的形变在视觉上是连续且平滑的。

(4) 梯度域融合。由图像每一像素点的形变向量得到拼接图像在重叠部分的梯度图，利用梯度域融合方法求解得到最终的拼接图像。

10.5.2 算法流程

在使用结构变形方法进行图像融合的四个步骤中，最优拼缝查找的方案和具体实现可参照第 9 章，梯度域融合过程参照 10.4 节的内容。本节主要介绍结构变形的核心步骤，即一维特征点提取与匹配和形变的量化与传播。

1) 一维特征点提取与匹配

眼睛对色彩的突变比渐变更为敏感，在观察自然图像时，相对于平滑区域内的色彩变化，人会更多地关注物体的边界。因此，以图像内拼缝与物体边界的交点为特征点，沿着拼缝方向进行一维特征点的提取与匹配，特征点匹配的结果能够反映拼缝两侧物体边缘的连接关系。

上述特征点的定义在梯度域体现为灰度值的局部最大点^[14]，特征点提取的过程转化为沿着拼缝方向对梯度域局部最大值的搜索。记两幅输入图像在重叠区域的梯度图为 $\|\nabla I_1\|$ 与 $\|\nabla I_2\|$ ，在沿拼缝高度 h 处的取值分别为 $C_L(h)$ 与 $C_R(h)$ 。由于自然图像中噪声和细节纹理的存在，曲线 C_L 、 C_R 的变化趋势并不能很好地反映物体边界的位置，如图 10-14 (b)、图 10-14 (d) 中黑色细实线所示。因此在特征点提取之前，首先对 C_L 、 C_R 进行适当的一维高斯平滑，得到平滑处理后的曲线 C_L^{smth} 、 C_R^{smth} ，如图 10-14 (b)、图 10-14 (d) 黑色粗实线所示。应用非最大值抑制法 (Non-Maximum Suppression)^[15]可快速求解得到特征点的位置，记为 $\{F_{L,1}, F_{L,2}, \dots, F_{L,m}\}$ 和 $\{F_{R,1}, F_{R,2}, \dots, F_{R,n}\}$ ，其中 m 和 n 分别为两幅输入图像在拼缝上的特征点数目。

综合特征点的灰度值和几何位置信息，定义反映特征点匹配程度的特征距离

$$d_F = \lambda_G d_G + \lambda_E d_E \quad (10.6)$$

式中， d_G 表示不在同一拼缝上的两个特征点在梯度图上的灰度值差异； d_E 表示其在重叠区域的几何距离； λ_G 、 λ_E 为对应的权重值。

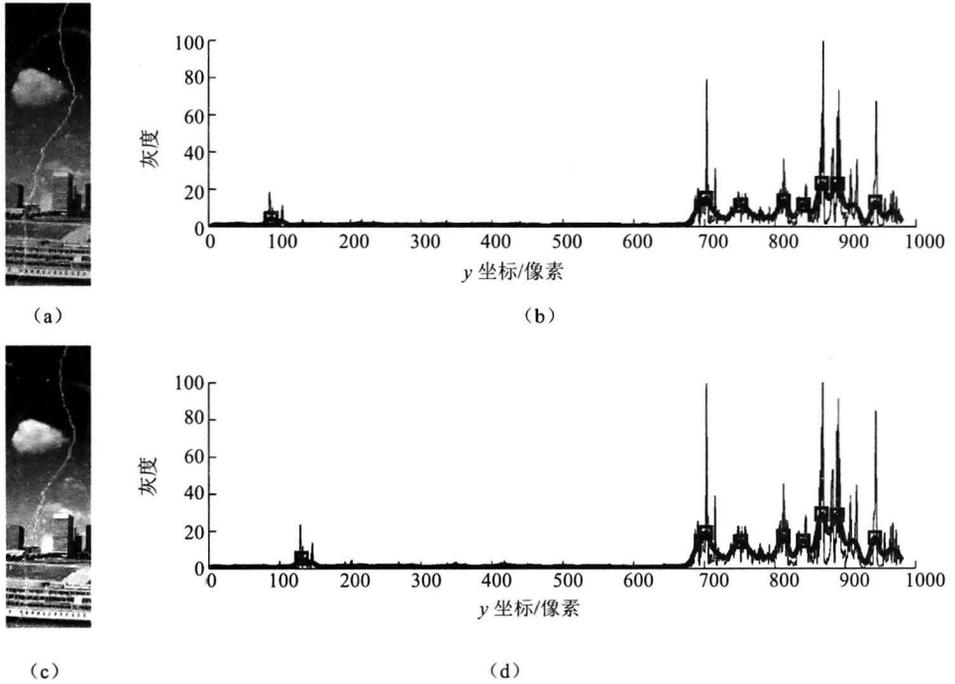


图 10-14 一维特征点提取

为消除噪声和细节纹理的影响， d_G 的计算过程应使用高斯平滑后的灰度值。任取一对特征点 $F_{L,i}(h_{L,i}, w_{L,i})$, $1 \leq i \leq m$ 与 $F_{R,j}(h_{R,j}, w_{R,j})$, $1 \leq j \leq n$, 有

$$d_G(F_{L,i}, F_{R,j}) = |C_L^{\text{smth}}(h_{L,i}) - C_R^{\text{smth}}(h_{R,i})| \quad (10.7)$$

$$d_E(F_{L,i}, F_{R,j}) = \sqrt{(h_{L,i} - h_{R,j})^2 + (w_{L,i} - w_{R,j})^2} \quad (10.8)$$

对于特征点集合 $\{F_{L,1}, F_{L,2}, \dots, F_{L,m}\}$ 和 $\{F_{R,1}, F_{R,2}, \dots, F_{R,n}\}$, 计算不同集合特征点两两之间的特征距离 d_F 。以 d_F 作为特征点相似程度的度量, 特征点的匹配可通过贪婪搜索^[13]或基于动态规划的搜索策略^[16]完成。

2) 形变的量化与传播

注意到一维特征点的匹配反映了拼缝两侧物体边界的连接关系, 为保证融合图像中物体边界的连续性, 匹配的特征点在融合图像上应该是重合的, 如图 10-15 所示。任取一组匹配的特征点 $F_{L,i}(h_{L,i}, w_{L,i})$, $1 \leq i \leq m$ 与 $F_{R,j}(h_{R,j}, w_{R,j})$, $1 \leq j \leq n$, 设形变之后二者在 S_{dst} 上重合于目标位置 $F_{dst,k}(h_{dst,k}, w_{dst,k})$, $1 \leq k \leq K$, 其中 k 为匹配的特征点对数, $F_{dst,k}$ 的图像坐标 $(h_{dst,k}, w_{dst,k}) = \left(\frac{h_{L,i} + h_{R,j}}{2}, S_{dst} \left(\frac{h_{L,i} + h_{R,j}}{2} \right) \right)$ 。

以形变向量的形式记录每一个特征点的位移量 $V(F) = (V_h(F), V_w(F))$, 其中 $V_h(F)$ 与 $V_w(F)$ 分别为 $V(F)$ 在高度和宽度方向上的分量:

$$F \in \{F_{L,1}, F_{L,2}, \dots, F_{L,m}\} \cup \{F_{R,1}, F_{R,2}, \dots, F_{R,n}\}$$

将形变向量的定义拓展到整幅图像, 形变传播即为将拼缝上一维特征点的形变向量 $V(F)$ 平滑地传播到图像其他部分的过程。定义 Ω_p 为形变传播的影响范围, 可以取 Ω_p 为包含拼缝的任意大小和形状的局部区域, 甚至是整幅图像。权衡形变的平滑度和计算量,

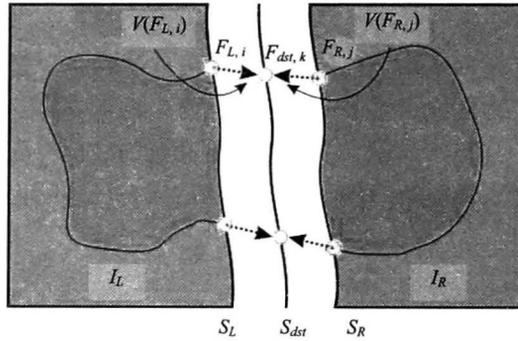


图 10-15 形变量化示意图

同时为了处理方便,取影响范围为重叠区域,即 $\Omega_p = \Omega$,融合图像的其他两部分 $I_L - \Omega_L$ 和 $I_R - \Omega_R$ 则不受形变传播的影响,有

$$V(p) = (0, 0) \quad \forall p \in \partial\Omega \cup (I_L - \Omega_L) \cup (I_R - \Omega_R) \quad (10.9)$$

形变向量的求解可描述为下述最小化问题:

$$V^* = \arg \min_V \iint_{p \in \Omega} \|\nabla V\|^2 dp \quad (10.10)$$

式 (10.10) 等价于拉普拉斯方程的求解:

$$\begin{cases} \Delta V = 0 \\ V|_{\partial\Omega} = V^*|_{\partial\Omega} \end{cases} \quad (10.11)$$

拉普拉斯方程为泊松方程的特例,可通过超松弛迭代求解。

将输入图像的梯度域图像按照形变向量 V 进行变形,得到融合图像在重叠区域的梯度域图像 (G_h, G_w) ,再使用梯度域融合即可求解最终的融合图像。

10.5.3 融合结果

对于图 10-16 (a)、图 10-16 (b) 所示输入图像,采用金字塔融合所得公共区域图像如图 10-16 (c) 所示,整幅融合图像如图 10-16 (d) 所示。

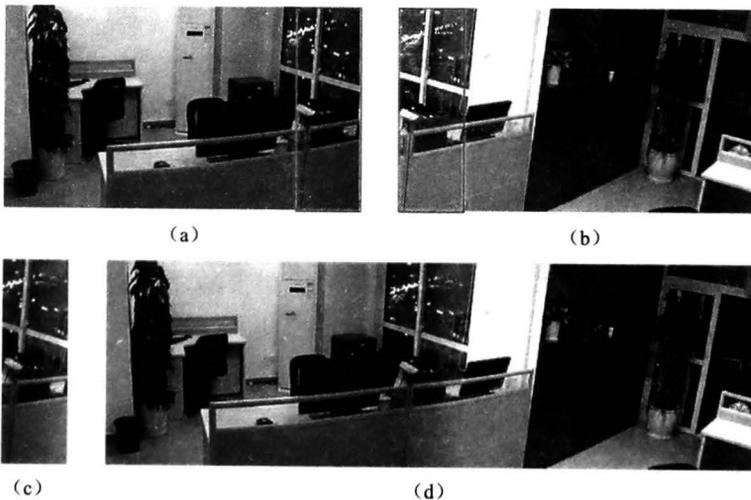


图 10-16 结构变形融合效果

与加权平均法、金字塔融合法以及梯度域融合法的融合效果相比,使用结构变形的的方法能够修正图像在亮度和结构上的偏差,获得更好的融合效果,代价是更高的算法复杂度。使用结构变形的融合方法要求多路拍摄时照相机光心的距离要远小于景深,否则输入图像将具有很大的视差,可能导致难以找到准确的结构对应关系,影响拼接效果。当场景中存在大量复杂且类似的结构时,也可能造成特征匹配上的困难。

参 考 文 献

- [1] MITCHELL H. Image fusion: theories, techniques and applications[M]. Burlin: Springer, 2010.
- [2] STATHAKI T. Image fusion: algorithms and applications[M]. Amsterdam: Elsevier, 2011.
- [3] SADJADI F. Comparative Image Fusion Analysis: Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, San Diego, June 20-26,2005[C], New York: IEEE, 2005.
- [4] BURT P, ADELSON E. A multiresolution spline with application to image mosaics[J]. ACM Trans. Graph., 1983(2). 217-236.
- [5] ADELSON E, ANDERSON C, BERGEN J, et al. Pyramid methods in image processing[J]. RCA engineer, 1984 (29), 33-41.
- [6] Rachel-Zhang. Laplacian Pyramid Blending[EB/OL]. 2012[2013-08-23]. <http://blog.csdn.net/abcjennifer/article/details/7628655>.
- [7] FATTAL R, LISCHINSKI D, WERMAN M. Gradient domain high dynamic range compression[J]. ACM Trans. Graph., 2002(21), 249-256.
- [8] BALLESTER C, BERTALMIO M, CASELLES V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. Image Processing, IEEE Transactions on, 2001(10), 1200-1211.
- [9] PEREZ P, GANGNET M, BLAKE A. Poisson image editing[J]. ACM Trans. Graph., 2003(22), 313-318.
- [10] LEVIN A, ZOMET A, PELEG A, et al. Seamless Image Stitching in the Gradient Domain: Computer Vision - ECCV 2004. 2004(3024), Prague, Czech Republic, May 11-14,2004[C]. Burlin: Springer, 2004.
- [11] ZOMET A, LEVIN A, PELEG A, WEISS Y. Seamless image stitching by minimizing false edges[J]. Image Processing, IEEE Transactions on, 2006(15), 969-977.
- [12] BOLZ J, FARMER I, GRINSPUN E, et al. Sparse matrix solvers on the GPU: conjugate gradients and multigrid[J] ACM Trans. Graph., 2003(22), 917-924.
- [13] JIA Jiaya, TANG Chi-Keung. Eliminating structure and intensity misalignment in image stitching: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Beijing, 2005[C]. New York: IEEE, 2005.
- [14] PERONA P, MALIK J. Detecting and localizing edges composed of steps, peaks and roofs[C]. Computer Vision, 1990. Proceedings, Third International Conference on, 1990: 52-57.
- [15] Neubeck A, Gool L. Efficient Non-Maximum Suppression: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Hong Kong, August 20-24, 2006[C], New York: IEEE, 2006.
- [16] JIA Jiaya, TANG Chi-Keung. Image Stitching Using Structure Deformation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008(30): 617-631.

第 11 章 光场成像技术

11.1 概 述

Michael Faraday 早在 1846 年就提出, 和磁场的性质类似, 光本身应该被定义成场的形式, 可以简单地称为光场 (Light Field)。概念上, 光场是一个可以描述空间中每个点、每个方向光线的方程。基于光场的图像采集和成像技术简称为光场成像技术。具体的, 光场成像技术是指通过特殊设计的光场照相机, 在成像过程中所捕捉的不再是一个普通相机采集到的光面, 而是记录了整个光场。也就是说, 光场照相机能捕捉一个场景中来自所有方向的光线, 能够让用户在拍完照片之后再借助相关算法对光线进行处理, 得到特殊的成像效果以及其他的应用。

光场成像技术采用新的成像装置, 属于光电技术与计算机技术紧密结合的交叉研究课题, 将获取全光信息的采集和处理作为研究重点。近年来随着其技术的不断发展, 在机载光学成像探测方面也有潜在的应用前景。通过光场成像技术, 拍照时只需要构图并按下快门即可, 对焦的步骤可以在拍照完后在计算机上完成。这彻底改变了现有拍照习惯, 甚至让自动对焦的功能和算法都变得多余。无论拍摄的照片模糊与否, 只要在照相机的焦距范围内, 对焦点可以在拍完之后随意选择, 因为其已经将拍摄场景内大部分光学信息都记录在内了。有人如此解释这一原理, 如果把一张照片记录的场景比作一场音乐会, 传统照相机记录的是乐团演奏的整体效果, 光场照相机记录的则是乐团每个成员的演奏效果, 用户可以选择放大不同乐器的声音。

进入 21 世纪以来, 光场成像技术无论在理论上还是应用上都得到了长足的发展。2011 年 Lytro 公司发布了世界首款消费级的光场照相机, 通过光场技术可以让用户拥有“先拍照后对焦”的功能。可以让任何景深的景物立刻成为拍摄焦点, 完全不去考虑景深问题。或者是改变观看照片的视角, 将一张照片在 2D 和 3D 模式之间来回切换。这款照相机之所以拥有这些功能, 就是因为其应用了光场成像技术, 可以收集进入照相机所有光线的强度和方向。Lytro 公司甚至都不用人们熟悉的像素指标来评价光场照相机的性能。取而代之的, 照相机的分辨率被称为 1100 万射线 (Ray), 也就是说可以捕捉到 1100 万束光线。另外, 这种照相机拍出来的照片是直接支持 3D 的, 用 3D 的显示器展示照片并戴上眼镜后, 可以清楚地看到 3D 照片, 而且在 3D 的环境下同样能够做到实时地数字变焦。

11.2 光场的数学定义

光场 (Light Field) 是光在空间中一个点在给定方向上的辐射度, 它是光线的一种五维辐射函数: $L(p_1, p_2, p_3, \theta, \phi)$, 其中 (p_1, p_2, p_3) 为该点空间三维坐标值, (θ, ϕ)

为该点处光线的俯仰角和方位角。

当光线沿直线传播时，可以假设其辐射度不会改变，因此在三维空间中，该五维的表述可以减到四维，从而减少了一维变量，简化了算法复杂度和数据存储空间。选择了四维光场这样的表述方法后，下面最主要的问题就是如何参数化表示这个空间里光线的信息。有个解决方案是 2PP (2 Planes Parameterization) 的方法，即通过光线与两个相互平行位面的交点来参数化该光线，这样，四维光场可以被解释成有向光线的空间之上的函数 $L(u, v, x, y)$ ，如图 11-1 所示。UV 平面和 XY 平面可以按需选取。

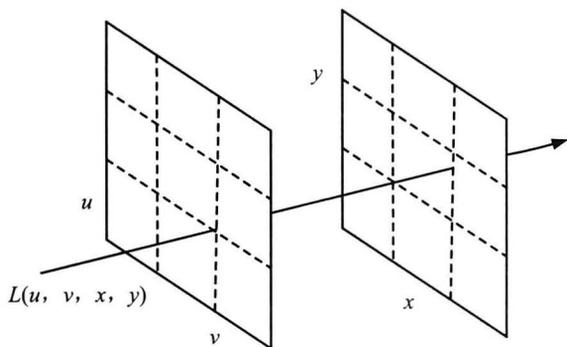


图 11-1 一个四维光场表示

目前，获取光场的手段主要有以下两种。

(1) 照相机阵列。它是指通过照相机在空间的一定排布来同时抓取一系列视角略有差别的图像，从而重构出光场数据的方法。比如斯坦福大学的 128 照相机阵列，采用大范围的空间排布，能够采集对场景多视角的图片，然后再对这些图片进行处理，实现数字重聚焦。

(2) 微透镜阵列。这也是常用的光场获取方式，实现方式也最简单。在普通成像系统的一次像面处插入一个微透镜阵列，每个微透镜元记录的光线对应相同位置不同视角的场景图像，从而得到一个四维光场。

除了目前研究比较多的这两种光场照相机之外，还有其他如植入掩膜的光场照相机，显微光场照相机等光场照相机。光场照相机本身并没有固有的设计模式和成像方法，只要能通过某种创新的光学结构重构出光场数据，都可以称为光场照相机。比如 Veeraraghavan 的光场照相机，通过在普通照相机光路中插入一个掩膜实现。其获取的图像看似与普通照相机类似，但经过变换到频域后发现，其频域呈规律性分布，与光场数据的频域特性类似，也能处理得到四维光场信息。它的优点在于掩膜是非折射元件，不管是从后期成像质量还是硬件方面，都比微透镜阵列更容易实现。又如可编程孔径照相机，插入的是一个特殊的可通过编程变化的遮光板，来提高图像的空间分辨率和景深。由于篇幅的关系，本节并没有覆盖光场照相机所有的细节，将通过重点和详细介绍照相机阵列和微透镜的光场照相机，来展示光场成像的原理和相关算法。

11.3 照相机阵列光场成像技术

照相机阵列是指通过照相机在空间的一定排布来同时抓取一系列视角略有差别的图像，从而重构出光场数据的方法。斯坦福大学的 128 照相机阵列，大尺度空间排布的照相机阵列主要用于合成孔径成像实现“透视”监测，或通过拼接实现大视角全景成像，而紧密排布型则主要用于获取高性能的动态场景。还有 Isaksen 的单照相机扫描系统，是通过照相机在场景中特定移动获取不同视角的图像，它构建的初衷在于研究光场数据的动态参量化。此外，比较成功的样机还有麻省理工学院的 64 照相机阵列，卡耐基-梅隆大学的 3D Room 等。

11.3.1 照相机阵列的结构和图像采集

下面以斯坦福大学的 128 照相机阵列合成孔径成像^[1]为例，讲述阵列照相机的基本原理。该阵列照相机如图 11-2 所示。每行有 16 个照相机，一共 8 行，总共 128 台照相机。每台照相机分辨率为 640×480 像素，30 帧/s，能同步采集图像信号。

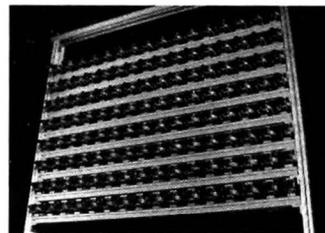


图 11-2 斯坦福大学阵列照相机^[1]

阵列照相机的合成孔径成像类似于合成孔径雷达。如果说只是单个照相机拍摄带有遮挡物的场景，那么拍摄主体可能会被挡住。在图 11-3 中，灌木丛背后的人大部分被遮挡。但是如果有多个照相机对同一场景进行拍摄，就能从不同的角度获得拍摄主体的影像，虽然每幅单独的图像，只能看见主体的部分，甚至很小一部分，但是当多幅图像融合在一起后，就能重组出完整的主体。这就类似于合成孔径雷达的透视效果（See Through），如图 11-4 所示。



图 11-3 单个照相机的成像^[2]



图 11-4 阵列照相机合成孔径成像效果^[2]

11.3.2 照相机阵列的数字重聚焦原理

现在要解决的问题是如何将这些图片融合在一起。如果融合效果不好，那么主体将会被模糊。另一个问题是，如果主体位于不同的深度位置，那么，能否对不同深度的主体聚焦。下面就来回答这些问题。

事实上，阵列照相机可以看成是一个超大光圈的单相机。如图 11-5 所示，单个的阵

列照相机可以看成是一个针孔照相机，景深非常大，大光圈的照相机景深却比较小。一般单个镜头的照相机光圈的直径是有限，而阵列照相机的等效光圈直径可达1~3m，这是单个照相机无法比拟的。因此阵列照相机的合成图像景深非常小，除了主体所在焦平面外，其他景物都已经模糊。例如，主体前灌木丛。阵列照相机的好处之一与大光圈相同，就是良好的信噪比。不同之处在于，大光圈照相机是通过光学的进光量大实现良好的噪声抑制性能，而阵列照相机是通过图像合成的方式。阵列照相机的超大光圈的另一好处就是前面提到的透视效应。

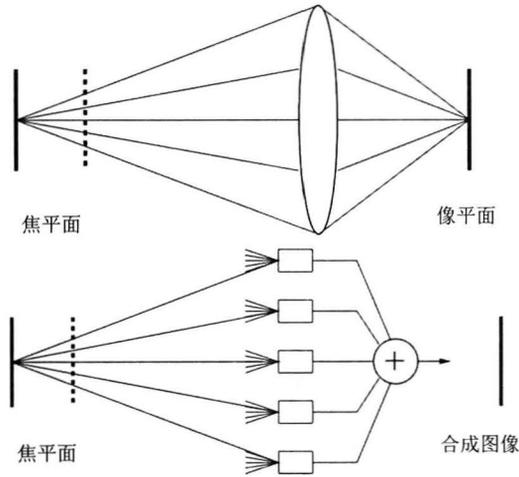


图 11-5 单镜头成像与阵列照相机成像

下面介绍一种基于校正板与视差 (plane parallax) 的阵列照相机配准校正合成方法^[3]。拍摄时，取景如图 11-6 所示。照相机到遮挡物 (灌木丛) 的距离为 33m，到目标主体的距离为 45m。

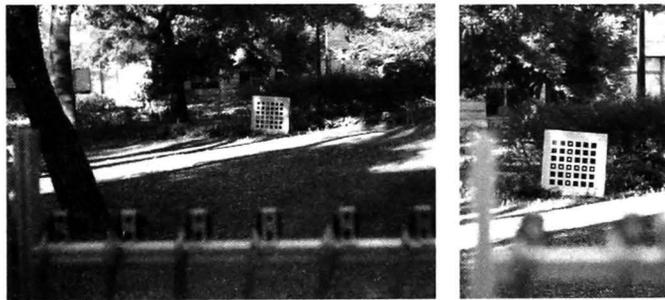


图 11-6 阵列照相机拍摄情况^[3]

通过拍摄，可以得到一系列同一场景不同视点的图像。为了配准目标主体，可以采用测量校正的方法。即通过计算照相机内参和外参：焦距、倾斜角、旋转角、平移量等，得到每幅图像间的配准关系，然后通过图像的稠密匹配，将主体配准合成。这种方法复杂，涉及非线性优化、初始值估计等，而且对于阵列照相机这种窄视角长焦距的成像，配准将会变得非常不稳定，误差较大。

在 plane parallax 方法中, 回避了照相机参数估计问题。而是使用标定板标定照相机位置。标定照相机时, 首先拍摄如图 11-7 (a) 和图 11-7 (b) 所示的一组图片。从图像中可以看到每幅图像中标定板都是有视差的。然后对图像进行变换, 将标定板变换到同一位置, 如图 11-7 (c) 和图 11-7 (d) 所示。将两幅图像叠加得到如图 11-7 (e) 所示的叠加图。图中可以看到标定板叠加清晰, 而背景出现重影。但是当多幅图像叠加时 [见图 11-7 (f)], 奇迹出现了, 背景出现了模糊效果, 这样相当于拍摄到了一张聚焦在标定板, 而背景模糊的图像。

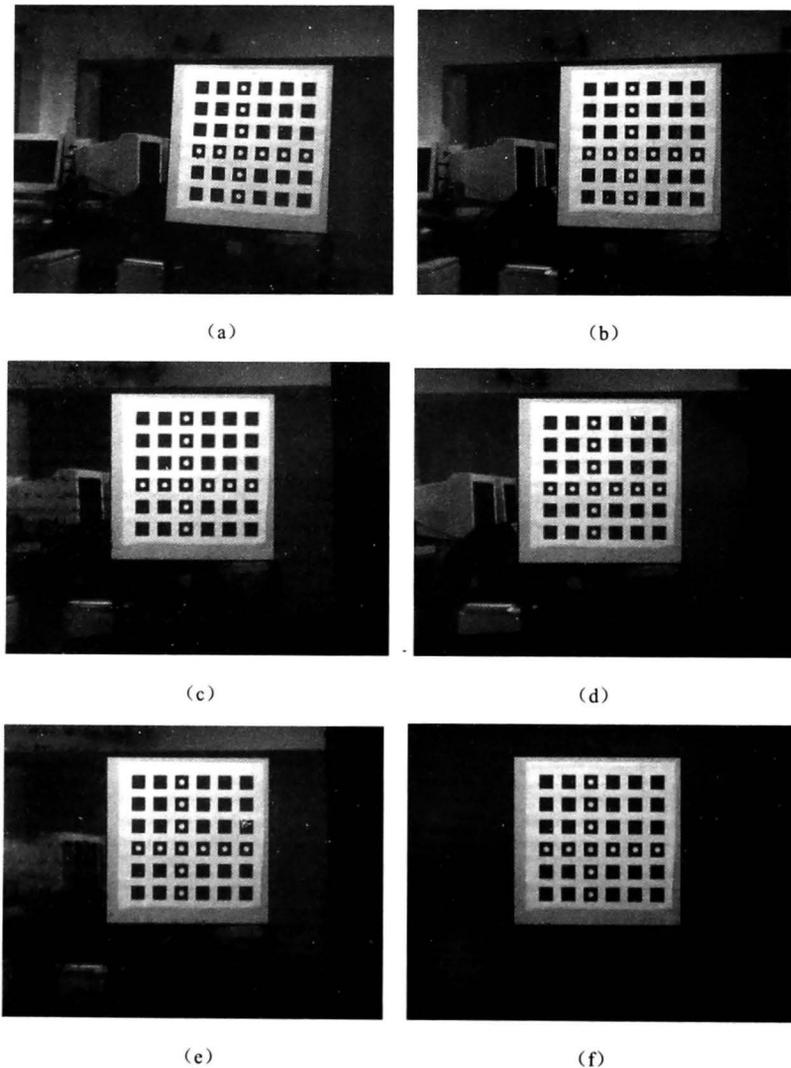


图 11-7 阵列照相机标定原理示意^[3]

那么, 如果要将焦点聚焦在其他焦平面怎么办呢? 那就要通过视图的平移。如图 11-8 (a) 和图 11-8 (b) 所示, 将右侧图像平移一段距离, 再将两幅图叠加, 就得到了如图 11-8 (c) 所示的效果。

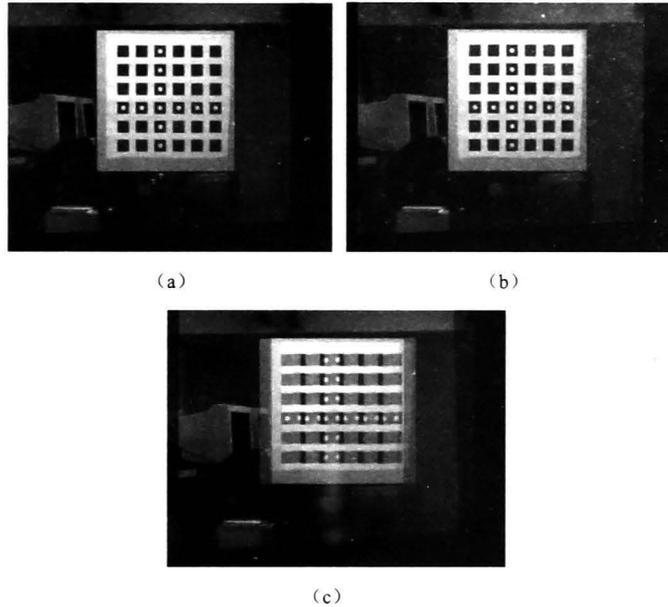


图 11-8 阵列照相机通过视图平移实现重聚焦示意^[3]

聚焦面到了标定板后的 CRT 显示器上，而前方的标定板出现了模糊。图 11-9 中的视差几何视图进一步说明了这一原理。

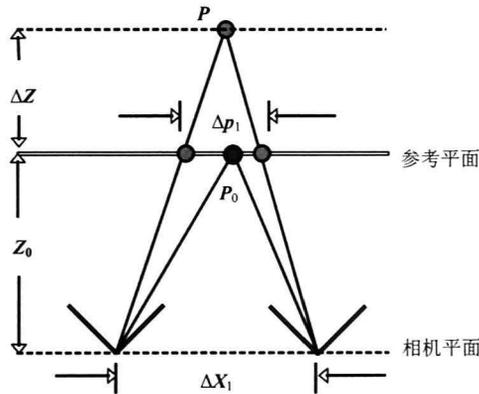


图 11-9 视差几何示意

假设当前标定板位于参考平面处，现在需要聚焦在参考平面。经过图像校正后，各照相机在参考平面深度上的物体都位于同一位置，即绿色的 P_0 点。叠加后， P_0 点聚焦。而不在参考平面深度上的橙色 P 点，在参考平面上将形成视差，视差的大小为 Δp_1 ，这一距离大小也是需要平移的大小。当需要聚焦在 P 点所在平面时，图像就需要平移一个视差 Δp_1 的距离。现在的问题是每次都手工进行平移，如何才能根据聚焦平面的选择，自动的计算出平移量的大小呢？从图 11-9 中可以得到平移量 Δp_1 与相机间距离 ΔX_1 的关系为

$$\Delta p_1 = \Delta X_1 \cdot \Delta z / (Z_0 + \Delta z) = \Delta X_1 \cdot d_p \quad (11.1)$$

式中， d_p 为相对深度大小。根据式 (11.1)，如果知道了照相机间距离 ΔX_1 ，当每次需要

聚焦到新的深度上（确定 d_p ）时，就可以自动计算出平移量 Δp_1 。现在问题转换为求取照相机间的距离。

当有多个照相机时，可以测量得到针对所有照相机的一组平移量 $[\Delta p_1 \ \Delta p_2 \ \cdots]^T$ ，如图 11-10 所示，而如果考虑多个视差点 P ，那么可以得到一个视差矩阵为

$$\begin{bmatrix} \Delta p_{11} & \cdots & \Delta p_{1n} \\ \vdots & \ddots & \vdots \\ \Delta p_{m1} & \cdots & \Delta p_{mn} \end{bmatrix} = \begin{bmatrix} \Delta X_1 \\ \vdots \\ \Delta X_m \end{bmatrix} [d_1 \ \cdots \ d_n] \quad (11.2)$$

式中， m 表示有 m 个照相机； n 表示有 n 个视点 $P: P_1, \cdots, P_n$ 。 $d_1 \ \cdots \ d_n$ 表示各视点的相对深度。式 (11.2) 可写为 $\Delta P = \Delta X \cdot D$ ，其中 ΔP 为秩为 1 的矩阵。根据这一特点可以通过奇异值分解 SVD，求解出照相机间距离 ΔX 。这样就可以对场景任意深度进行重聚焦了。实际应用时，通常以标定板所在位置为基础，记标定板到照相机平面的深度为 f 。在不同平面聚焦，意味着标定板深度的改变，即 f 的变化，那么图像则平移 $f \cdot \Delta X_i$ 。

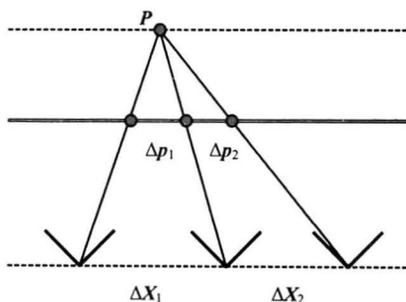


图 11-10 多相机视差示意图

这里给出阵列照相机合成孔径成像一个新的实验^[1]，如图 11-11 所示。图中，人躲在树丛背后，从单幅图像很难看到人脸。但经过处理后，树丛被模糊了，但人脸清晰可见。



图 11-11 阵列照相机合成孔径成像实验效果^[4]

除了合成孔径成像外，阵列照相机还可以利用帧插法做高速摄影，控制照相机的不同曝光做高动态范围成像，以及通过超分辨率处理提高相机性能等。限于篇幅，这里不再赘述。对于阵列相机而言，目前存在的问题主要是深度提取较为困难，导致重聚焦时，聚焦物体的模糊。未来的研究重点将是如何解决这一问题。

11.4 微透镜光场成像技术

能够对一次曝光后获得的照片进行数字重聚焦是光场成像技术很重要的一个应用。11.3 节介绍的阵列照相机的数字重聚焦技术主要实现的是对遮挡物的透视功能，通过模糊前景来清晰化被遮挡的物体，或者虚化背景来数字变焦到前景。基于这样的原理，不可能实现前景、后景的同时聚焦，也就是没有办法实现全聚焦清晰图像。而且对于消费级的光场照相机来说，阵列照相机显然不是一个现实的选择。因此，小型化的微透镜照相机成了许多研究者关注的重点，微透镜照相机对不同深度图像的重聚焦，通过图像融合技术可以生产全聚焦的整幅清晰图像。

11.4.1 微透镜光场照相机计算成像原理

微透镜光场照相机的光学成像系统由三部分构成：主透镜、微透镜阵列和数字感光芯片，如图 11-12 所示。微透镜光场照相机的成像原理为：物体上一点经过主透镜系统后聚焦于微透镜阵列平面，而后经单个微透镜分散出强度和方向分量，到达探测器的不同面元，此即该物点的光场信息。如果把微透镜看作成像器件，主透镜的主面与探测器阵列满足物像共轭关系。所以，主透镜系统的光瞳经过单个微透镜所成的像正好覆盖光电探测器的若干像素点，如图 11-13 所示。为达到最大的角分辨率，需保证主透镜系统的像方数 F 与微透镜阵列的数一致，否则会造成光场方向信息和成像目标空间分布的混叠^[2, 5]。

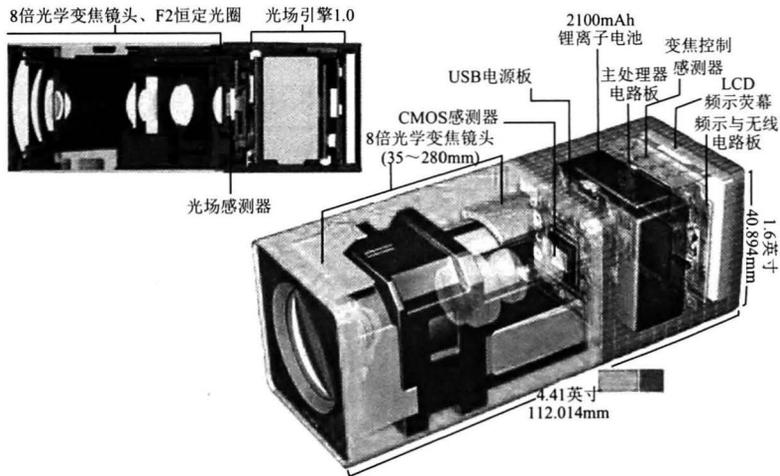


图 11-12 Lytro 微透镜光场照相机结构示意图^[4]

光场照相机的光学系统抽象成四维光场，其中， u, v 面是光学系统的主面， x, y 面是探测器所在平面， $L_F(x, y, u, v)$ 代表给定光线的光辐射量，下标 F 代表两平面之间的距离，像面上接收到的辐射量可表示为

$$E_F(x, y) = \frac{1}{F^2} \iint L_F(x, y, u, v) A(u, v) \cos^4 \theta \, du \, dv \tag{11.3}$$

式中， θ 为光线与 u, v 面法线的夹角； $A(u, v)$ 为光瞳函数。

图 11-13 微透镜光场照相机采集到的原始图片^[5]

假定 x, y 和 u, v 面无穷大, 光瞳之外的光线为 0, 引入近轴近似, 将公式简化为

$$E_F(x, y) = \frac{1}{F^2} \iint L_F(x, y, u, v) du dv \quad (11.4)$$

通过数值积分近似求解, 便可得到一幅数字图像。

11.4.2 微透镜光场照相机的数字重聚焦原理

如图 11-14 所示, $x-y$ 面为重聚焦目标物所在的空间像平面, 其光辐射量可表示为

$$\begin{aligned} & L\alpha_F(x, y, u, v) \\ &= L_F\left(u\left(1-\frac{1}{\alpha}\right)+\frac{x}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y}{\alpha}, u, v\right) \\ &= L_F\left\{B_\alpha[x, y, u, v]^T\right\} \end{aligned} \quad (11.5)$$

代入式 (11.4), 定义切片算子 $\beta[f](x) = f(\beta x)$, 重聚焦后的图像可表示为

$$E_F(x, y) = \frac{1}{\alpha^2 F^2} \iint B_\alpha[L_F] du dv \quad (11.6)$$

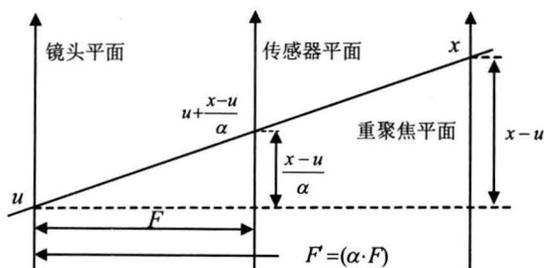


图 11-14 数字重聚焦原理

根据傅里叶切片定理, x, y 面上得到的光辐射量是光场的一个切片的投影积分。也就是说, 通过一次曝光得到的 4D 光场信息, 可用于重建不同焦距处的图像。当对一幅图像进行多次重聚焦处理时, 频域的方法比空域更加简便。将四维光场变换到频域为

$$G_F(f_x, f_y, f_u, f_v) = \zeta\{L_F(x, y, u, v)\} \quad (11.7)$$

经过切片处理后得到图像频谱为

$$G_F(f_x, f_y) = \frac{1}{F^2} G_F\left(B_\alpha^{-T} \cdot [f_x, f_y, f_u, f_v]^T\right) \quad (11.8)$$

逆变换后得到图像表达式为

$$E_F(x, y) = \zeta^2 \{G_F(f_x, f_y)\} \tag{11.9}$$

图 11-15 为空域和频域两种算法的复杂度示意图。由于 a 的连续变化，空域和频域的计算复杂度分别集中在投影积分和二维切片阶段。频域相比空域的计算，速度有很大的提升，这也是为什么在频域进行傅里叶切片的原因。

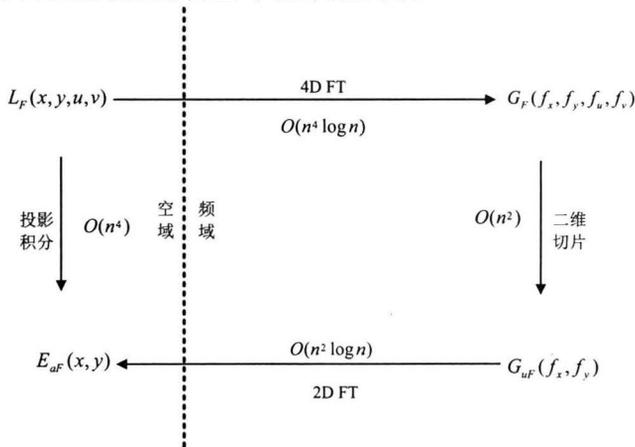


图 11-15 空域和频域两种重聚焦算法的计算复杂度比较示意图

图 11-16 所示是一次曝光获得的光场照片经过重聚焦处理的结果，如图 11-16 (a) 所示的焦面在后面的被拍摄者脸上，如图 11-16 (b) 所示的焦面在前面被拍摄者脸上。

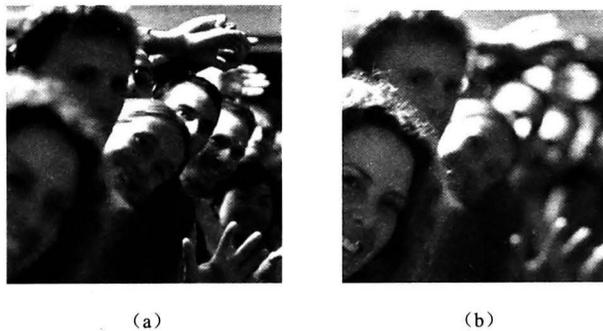


图 11-16 微透镜光场照相机实现数字重聚焦效果^[4]

11.5 总结与展望

传统照相机在成像的同时限制了图像本身的重塑性，而光场成像则保留了这种可能，因为它记录的是包含位置和方向信息的四维光场数据。正因如此，图像的拍摄者能对视角、远近景，甚至是光线本身进行拍摄后的操控，使得图像信息处理实现灵活化和多元化。可以通过数字重聚焦技术摆脱拍照时失焦的困扰，增加对图片处理的灵活性。在安防监控领域，通过阵列光场照相机可以实现透视监控。

随着光场采集仪器的研制和相关算法的提出，与计算机紧密联系的光场成像技术已经成为计算摄影技术研究的热点。具体地，目前获取光场的成像系统朝着两个方向发展：

照相机阵列和微透镜式的光场照相机。传统成像是把三维的场景投射到二维的图片的方法来记录,若能实现把三维场景整个光场记录下来,可以记录和还原一个更为真实的世界。

然而,光场数据比普通图像多出的维度信息是以牺牲空间分辨率为代价的,二者之间存在一个折中和平衡。特别是对于现有的微透镜式光场,照相机普遍存在图像空间分辨率不能满足需求的问题,如果加大图像空间分辨率的同时兼顾轴向分辨率,则会对光学器件的加工工艺提出更高要求,这也是当前制约光场成像技术发展的一个主要瓶颈,如何在二者之间获得更好的平衡,是今后研究的一个重点。此外,由于一次曝光获取的数据量巨大,对存储设备的容量和图像处理器速度也提出了更高的要求。因此,光场成像在技术实现、硬件处理能力,算法效率和应用成本等方面还有亟待解决的难题值得研究人员继续努力。

参 考 文 献

- [1] Wilburn B, Adviser-Horowitz M A. High-performance imaging using arrays of inexpensive cameras[M]. Stanford: Stanford University, 2005.
- [2] 聂云峰, 相里斌, 周志良. 光场成像技术进展[J]. 中国科学院研究生院学报, 2011, 28(5): 563-572.
- [3] Vaish V, Wilburn B, Joshi N, et al. Using plane+ parallax for calibrating dense camera arrays[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 1: 1-2-1-9 Vol. 1.
- [4] Image download from <https://www.lytro.com/>.
- [5] Georgiev T, Yu Z, Lumsdaine A, et al. Lytro camera technology: theory, algorithms, performance analysis[C]//IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2013: 86671J-86671J-10.

第 12 章 图像去噪技术

12.1 概 述

从生理学观点来看，凡是妨碍人们工作和生活的因素，都可以被视为噪声。从信号的角度来看，噪声的产生是信号在采集、传输以及记录过程中，受到成像设备自身因素和外界环境的影响而产生的。通过定位和界定噪声类型和噪声水平，并且采用相应算法从噪声信号中恢复出真实信号的过程称为去噪。

现实中噪声是随机分布的，只把它定义成一种模型是不准确的。事实上，噪声无法完全去除，只能使重现信号尽可能接近原始信号。因此，去噪严格意义上应该称为降噪。数字图像信号在数字化和传输过程中由于常常受到成像设备以及外部环境的影响，会包含许多噪声，因此一般将初始得到的图像信号称为含噪图像或噪声图像。噪声会影响图像的视觉效果，也会给后续图像分析，如压缩、编码、识别等，造成不利影响，因此需要对噪声图像进行去噪处理。减少数字图像中噪声的过程称为图像去噪（Image Denoising），其目的是解决由于受噪声污染而导致图像质量下降的问题。通过去噪可以有效地增大图像信号的信噪比，提高图像质量，更好地体现原始图像所携带的信息。

由于图像去噪的重要性，图像去噪方法一直是图像处理领域的研究热点之一，研究人员也陆续提出了多种去噪方法。诸如传统的空间域和变换域滤波、基于小波变换、偏微分方程、非局部特性的图像去噪方法等。图像去噪算法遵循的规则是在去除噪声的同时能尽可能多地保护图像的边缘细节信息，保持原有信息的完整性。然而，在现有的去噪算法中，有的去噪算法在低维信号图像处理中取得较好的效果，却不适用于高维信号图像处理；有的去噪效果较好，却丢失部分图像边缘信息；有的致力于研究检测图像边缘信息，保留图像细节。在抵制噪声和保留细节上找到一个较好的平衡点逐渐受到研究人员的重视，各种“保边”的去噪算法成为近年来研究的重点。当然，随着照相机分辨率的提高，对去噪算法的效率要求也更高，效果和效率的平衡也一直是贯穿整个计算摄影技术研究重点。

图像去噪作为一个经典并且热门的课题已经被研究了几十年，一个很自然的疑问是：这个研究课题是不是已经发展到尽头了？针对这个问题，论文 *Is Denoising Dead?*^[1] 从数学的角度分析并给出了一个否定的答案。也就是说，去噪算法在提高信噪比、改善图像质量上还有很大的研究空间。随着应用需求的不断深入及应用领域的不断拓展，寻求有效的图像去噪算法依旧任重而道远。过完备稀疏表示理论的提出及其在众多领域的成功应用，为图像去噪领域研究提供了一个新的思路 and 方向。比如，基于稀疏表示理论的去噪算法可以将被噪声严重污染的图片恢复到不可思议的效果，如图 12-1 所示。

本章要从图像去噪的基本概念和原理开始，通过分析噪声模型，介绍经典的图像去

噪算法以及近期的研究成果,使读者对于计算摄影技术中图像去噪的理论和方法有基本的认识。



图 12-1 稀疏表示去噪方法效果^[2]

12.2 图像去噪的基本概念

在进行图像去噪处理之前,首先需要知道图像中到底存在哪些噪声,它们到底具有什么样的特性,然后,再针对不同的噪声进行具有针对性的去噪处理。

12.2.1 图像噪声的分类

噪声的来源有许多,因而噪声也多种多样,其性质也是千差万别,这就导致了图像噪声没有一种统一的分类方法,而是根据不同的因素进行分类。通常,有如下几种分类方式。

1) 按噪声产生的原因分类

根据噪声产生的原因,可以将噪声分为外部噪声和内部噪声。所谓外部噪声,即是指由成像系统的外部因素导致的噪声,如外部电气设备、天体放电现象等引起的噪声,它们是以电磁波或者电流的方式对成像系统的内部产生影响,进而产生噪声。所谓内部噪声,即是指由成像系统的内部因素导致的噪声,如由光 and 电的基本性质所引起的噪声、电器的机械运动产生的噪声、器材材料本身引起的噪声以及系统内部设备电路所引起的噪声等。

2) 按噪声与图像信号的关系分类

根据噪声与图像信号之间的关系,又可以将噪声分为加性噪声和乘性噪声。加性噪声与图像信号的强度无关,如图像在传输过程中引进的信道噪声。一张含有加性噪声的图像可以表示为

$$f = g + n \quad (12.1)$$

式中, g 为假设的理想无噪声图像; n 表示加性噪声; f 即为最终含有噪声的图像。

与加性噪声不同,乘性噪声与图像信号的强度相关,它往往随着图像信号强度的变化而变化,如飞点扫描器扫描图像时的噪声,电视扫描光栅时产生的噪声,以及胶片中的颗粒噪声就属于此类噪声。对于含有乘性噪声的图像,可以表示为

$$f = g + n \otimes g \quad (12.2)$$

3) 按概率密度函数 (PDF) 分类

从噪声的概率密度分布情况来看, 可以将图像噪声分为高斯噪声、瑞利噪声、伽马噪声、指数噪声、均匀噪声和脉冲噪声等。

高斯噪声的概率密度函数服从高斯分布 (即正态分布), 其计算公式如下所示:

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right] \quad (12.3)$$

式中, z 表示灰度值; μ 表示 z 的平均值或期望值; σ 表示 z 的标准差。当 z 服从上述分布时, 其值有 95% 落在 $[(\mu - 2\sigma), (\mu + 2\sigma)]$ 范围内。如果一个噪声, 它的幅度分布服从高斯分布, 而它的功率谱密度又是均匀分布的, 则称它为高斯白噪声。在空间域和频域中, 由于高斯噪声在数学上的易处理性, 其噪声模型经常被用在实践中。

瑞利噪声的概率密度函数符合瑞利分布, 其计算公式为

$$p(z) = \begin{cases} \frac{2}{b}(z-a) \exp\left[-\frac{[z-a]^2}{b}\right] & z \geq a \\ 0 & z < a \end{cases} \quad (12.4)$$

其均值和方差分别为

$$\begin{cases} \mu = a + \sqrt{\pi b / 4} \\ \sigma^2 = b(4 - \pi) / 4 \end{cases} \quad (12.5)$$

伽马噪声的概率密度函数为

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} e^{-az} & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (12.6)$$

其密度的均值和方差为

$$\begin{cases} \mu = \frac{b}{a} \\ \sigma^2 = \frac{b}{a^2} \end{cases} \quad (12.7)$$

指数噪声的概率密度函数为

$$p(z) = \begin{cases} az^{-a} & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (12.8)$$

其中, $a > 0$, 其期望值和方差为

$$\begin{cases} \mu = \frac{1}{a} \\ \sigma^2 = \frac{1}{a^2} \end{cases} \quad (12.9)$$

均匀噪声的概率密度函数为

$$p(z) = \begin{cases} \frac{1}{b-a} & a \leq z \leq b \\ 0 & \text{其他} \end{cases} \quad (12.10)$$

其均值和方差为

$$\begin{cases} \mu = \frac{a+b}{2} \\ \sigma^2 = \frac{(b-a)^2}{12} \end{cases} \quad (12.11)$$

脉冲噪声的概率密度函数为

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & \text{其他} \end{cases} \quad (12.12)$$

如果 $b > a$, 灰度值 b 在图像中将显示为一个亮点, 相反, a 的值将显示为一个暗点。若 P_a 或 P_b 为零, 则脉冲噪声称为单极脉冲。如果 P_a 和 P_b 均不为零, 尤其是它们近似相等时, 脉冲噪声值将类似于随机分布在图像上的胡椒和盐粉微粒。由于这个原因, 双极脉冲噪声也称为椒盐噪声。

此外, 还可以按统计理论观点, 按噪声幅度随时间分布形状、按噪声的频谱形状等方式对噪声进行分类, 这里不再赘述。

12.2.2 去噪效果评价指标

目前, 对去噪的效果进行定量的质量评价, 主要有两种客观标准, 峰值信噪比 (PSNR) 和结构相似度测量 (SSIM)。PSNR 定义为

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\text{MSE}} \right) \quad (12.13)$$

式中, L 表示的是图像的动态范围 (8 位/像素的图像); MSE 表示的是原始图像和失真图像之间的均方误差。

PSNR 是最普遍、最广泛使用的评鉴画质的客观量测法, 不过许多实验结果都显示, PSNR 的分数无法和人眼看到的视觉品质完全一致, 有可能 PSNR 较高者看起来反而比 PSNR 较低者差。这是因为人眼的视觉对于误差的敏感度并不是绝对的, 其感知结果会受到许多因素的影响而产生变化。例如, 人眼对空间频率较低的对比差异敏感度较高, 对亮度对比差异的敏感度较色度高, 对一个区域的感知结果会受到其周围邻近区域的影响。

SSIM 定义为

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12.14)$$

式中, x 和 y 表示从原始图像和含噪图像中的局部窗口中提取的图像块; μ_x 、 σ_x^2 和 σ_{xy} 分别表示局部窗口的平均值、方差以及互相关。一个视频帧的整体 SSIM 值由局部 SSIM 值平均获得。

结构相似性理论认为, 自然图像信号是高度结构化的, 即像素间有很强的相关性, 特别是空域中最接近的像素, 这种相关性蕴含着视觉场景中物体结构的重要信息。作为结构相似性理论的实现, 结构相似度指数从图像组成的角度将结构信息定义为独立于亮度、对比度的, 反映场景中物体结构的属性, 并将失真建模为亮度、对比度和结构三个

不同因素的组合。用均值作为亮度的估计，标准差作为对比度的估计，协方差作为结构相似程度的度量。普遍认为，SSIM 是比 PSNR 更好的感知图像质量的标准。

12.3 传统去噪算法

现实中噪声是随机分布的，如果人为地把它定义成一种模型是不准确的。事实上，噪声无法完全去除，只能使重现信号尽可能接近原始信号。图像去噪算法遵循的规则是在去除噪声的同时能尽可能多地保护图像的边缘细节信息，保持原有信息的完整性。

由于图像去噪的重要性，图像去噪方法一直是图像处理领域的研究热点之一，研究人员也陆续提出了多种去噪方法。诸如传统的空间域和变换域滤波、基于小波变换、偏微分方程、非局部特性的图像去噪方法等。

12.3.1 基于空间域的中值滤波

中值滤波器是一种常用的非线性平滑滤波器，图像去噪处理中使用的中值滤波方法的基本原理为：将图像中一点的值用该点邻域内各点的中值代换。设 $f(x, y)$ 表示图像像素点 (x, y) 的灰度值，该点的邻域为 A ，则中值滤波器可以定义为

$$\hat{f}(x, y) = \text{MED}\{f(x, y)\}_{(x, y) \in A} \quad (12.15)$$

式中，当 n 为奇数时， n 个数 x_1, x_2, \dots, x_n 的中值就是按数值大小顺序处于中间的数；当 n 为偶数时，中值即为两个中间数的平均值。图 12-2 所示为对添加了椒盐噪声的含噪图像进行中值滤波后的效果图。



图 12-2 中值滤波效果^[3]

12.3.2 基于小波域的小波阈值去噪

小波萎缩法是目前研究最为广泛的方法。小波萎缩法可分成两类：一类为阈值萎缩，一类为比例微缩。阈值萎缩方法首先设定一个合适的阈值，对于小于阈值的系数，将其置零，否则保留系数的值；然后，进行阈值函数的映射从而得到估计系数；最后，对得到的估计系数进行反变换，就能实现降噪处理。比例萎缩法首先判断系数被噪声污染的程度，然后根据不同的程度引入各种度量方法，进而最终确定萎缩的比例。通常，在进行去噪处理时使用的都是阈值萎缩法，因此下面重点介绍这种方法。

阈值萎缩方法包含两个基本要素：阈值和阈值函数。

如何选择合适的阈值是此种方法的关键所在。目前使用的阈值可以分为全局阈值和局部适应阈值两类。其中，全局阈值对各层所有的小波系数或同一层内的小波系数都是统一的；而局部适应阈值是根据当前系数周围的局部情况来确定阈值。目前，全局阈值主要有以下几种。

(1) Donoho 和 Johastone 统一阈值 (DJ 阈值)^[4]，其公式为

$$\delta = \sigma \sqrt{2.1_n N} \quad (12.16)$$

式中， σ 为噪声标准方差； N 为信号的尺寸或长度。

(2) 基于零均值正态分布的置信区间阈值，其公式为

$$\delta = 3\sigma \sim 4\sigma \quad (12.17)$$

(3) Bayes Shrink 阈值^[5]和 Map Shrink 阈值^[6]。

(4) 最小最大化阈值：这是 Donoho 和 Johnstone 在最小最大化意义下得出的阈值，与上边的阈值不同，它是依赖于信号的，而且没有显式表达式，在求取时需要预先知道原信号。

(5) 理想阈值：理想阈值是在均方差准则下的最优阈值，同最大最小化阈值一样，也没有显式的表达式，并且这个阈值的计算通常也需先知道信号本身。

基于选择的阈值，Bruce 和 Gao 提出了一种半软阈值函数^[7]：

$$\delta(w) = \text{sgn}(w) \frac{T_2(|w| - T_1)}{T_2 - T_1} I(T_1 < |w| < T_2) + w I(|w| > T_2) \quad (12.18)$$

该方法通过选择合适的阈值 T_1 和 T_2 ，可以在软阈值方法和硬阈值方法之间达到很好的折中。图 12-3 所示为采用小波阈值去噪后得到的效果图。



图 12-3 小波阈值去噪的效果^[8]

12.3.3 基于 PDE 的图像去噪

目前，基于 PDE 的图像去噪方法也是图像去噪领域的研究热点，并且，该方法已经取得了一定的理论成果，并且在实际中已经得到了应用。这种方法通过建立噪声图像为某非线性 PDE 的初始条件，然后求解这个 PDE，得到在不同时刻的解，即为滤波结果。

Perona 和 Malik^[9]提出了基于 PDE 的非线性扩散滤波方法（以下简称 P-M），这种方法是一种非线性的各向异性方法，目的是为了克服线性滤波方法存在的模糊边缘和边缘位置移动的缺点。基本思想是：图像特征强的地方减少扩散系数，图像特征弱的地方

增强扩散系数。方程如下：

$$\begin{cases} \frac{\partial u}{\partial t} = \text{div}[g(|\nabla u|)\nabla u] & t \in (0, T) \\ u(x, y, 0) = u(x, y) \end{cases} \quad (12.19)$$

式中， $u(x, y, t)$ 表示的是随时间变化的图像； $|\nabla u|$ 表示的是梯度的模； $g(|\nabla u|)$ 表示的是扩散系数函数，用于控制扩散的速度。理想的扩散系数应当使各向异性扩散在灰度变化平缓的区域快速进行，而在灰度变化急剧的位置（即图像特征处）低速扩散乃至不扩散，所以， $g(|\nabla u|)$ 应具有如下性质：

$$\begin{cases} g(0) = 1 & \text{表示在非边缘处加强扩散} \\ \lim_{|\nabla u| \rightarrow \infty} g(|\nabla u|) = 0 & \text{表示在边缘处停止扩散} \end{cases} \quad (12.20)$$

基于上述的两个性质，P-M 给出了如下的扩散系数函数：

$$g(|\nabla u|) = \frac{1}{1 + (|\nabla u|/k)^2} \quad (12.21)$$

式中， k 为边缘阈值，其作用是判断边缘区域和平坦区域。为了阐明阈值 k 在扩散操作中所起到的作用，需要引入通量函数，其函数定义为

$$g'(|\nabla u|) = \frac{1}{1 + (|\nabla u|/k)^2} \quad (12.22)$$

在抑制噪声与保留图像重要特征方面，P-M 方程取得了一定的效果，如图 12-4 所示，但却存在病态和不稳定的缺陷。

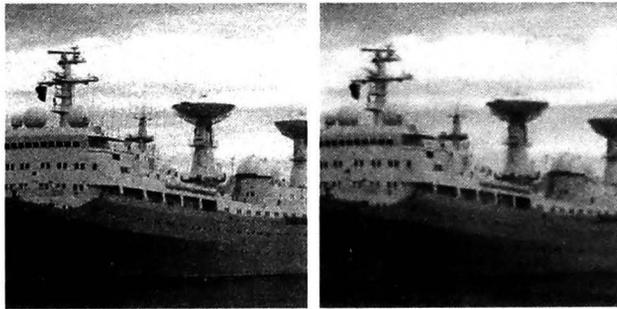


图 12-4 P-M 去噪的效果^[10]

12.3.4 全变分图像去噪

全变分 (TV) 图像去噪方法首先由 Rudin、Osher 和 Fatemi 提出^[11]，这种方法基于变分法的思想来确定图像的能量函数，通过对图像能量函数进行最小化来达到平滑去噪的目的。TV 法也是现在比较流行的图像复原方法，能量函数定义如下：

$$TV[u(x, y)] = \iint_{\Omega} |\nabla u(x, y)| dx dy \quad (12.23)$$

全变分去噪能量泛函为

$$E = \int_{\Omega} \frac{1}{2} (u - u_0)^2 + \lambda \times TV(u) \quad (12.24)$$

为了使能量函数最小，其欧拉-拉格朗日方程为

$$\frac{\partial u}{\partial t} = -\nabla \left(\frac{\nabla u}{|\nabla u|_{\beta}} \right) + \lambda(u - u_0) = 0 \quad (12.25)$$

其中，梯度算子表示为

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \quad (12.26)$$

正则项为

$$|\nabla u|_{\beta} = \sqrt{|\nabla u| + \beta^2} \quad (12.27)$$

其功能是用来减少平坦区域的退化。在进行局部坐标系转换后，方程可以分解为边缘方向和边缘正交的两个方向，分解后各方向的系数控制着该方向的扩散强度。扩散方向实际上是一个分线性的各向异性的扩散方程，其扩散算子仅沿着图像梯度的正交方向扩散，扩散系数为 $1/|\nabla \mu|$ ，而朝着梯度方向无扩散。这样可以通过图像的梯度来判断边缘位置，使得边缘扩散系数最小，从而降低对边缘的模糊程度，但是由于边缘的扩散系数小，噪声得不到很好的抑制，而且当 $|\nabla \mu| > \lambda$ 时，势能函数是非凸的，使得边缘处处理表现得不稳定。所以，如何确定扩散参数的值是一个问题。图 12-5 所示为采用 TV 方法的去噪结果。

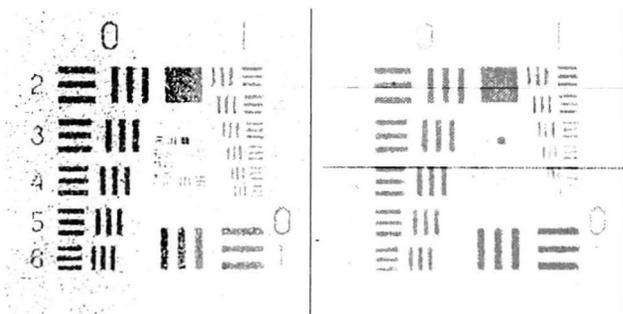


图 12-5 TV 去噪的结果^[11]

12.4 非局域均值去噪算法

12.4.1 NLM

中值滤波、小波变换和偏微分方程 (PDE) 等传统去噪方法的基本思想都是局部平均滤波，即在每个像素一定的邻域内，求所有像素或部分像素的平均或加权平均。换言之，传统去噪算法选择参与平均运算的像素集的标准是与待处理像素点之间的空间距离。近年来，研究人员发现图像恢复可以利用图像本身的自相似性。如图 12-6 所示，大部分图像细节重复出现，图中颜色相同的图像块表明这些区域所包含的信息是十分相似的。很多学者仔细研究了图像自相似性所蕴含的冗余信息，取得了很多成果。例如，在 1999 年，Efros 和 Leung 首先利用图像的自相似性进行了图像的纹理合成以及小洞的填补^[12]。该算法在图像的一个比较大的局部区域中搜索与待恢复像素相似的所有像素

点。这种思想同样也可以应用到图像去噪中，在较大范围内寻找与待处理像素的相似像素点，最后进行加权平均来得到处理后的图像。相对于邻域滤波器，这种方法更为可靠。

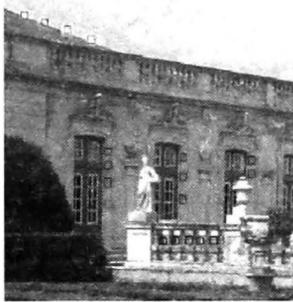


图 12-6 自然图像的相似性^[13]

2005 年，Baudes 等人在前人对纹理合成以及双边滤波的基础上首次提出非局部平均图像去噪算法 (Non-Local Means) ^[14]，这种算法所采用的图像去噪方法与之前所有去噪方法都不同，该算法充分利用图像自身的信息进行各种处理，在图像平滑的基础上能很好地保持细节信息，健壮性较好。NLM 算法是当前图像去噪算法中降噪能力较强的一种。

1) NLM 算法基本步骤

非局部均值滤波的基本思想是：当前像素点的灰度值由图像中所有与其结构相似的像素点的灰度值加权平均得到，结构越相似，权重越大。

假设有含噪图像 $u = \{u(i) | i \in I\}$ ，其中 I 表示图像的域。对于其中任何一个像素 i ，非局部均值的方法利用整幅图像中所有像素点的加权平均来得到该点的估计值，即

$$NL(u(i)) = \sum_{j \in I} w(i, j) u(j) \tag{12.28}$$

式中，加权值 $w(i, j)$ 依赖于像素点 i 与 j 之间的相似性，并且满足条件： $0 \leq w(i, j) \leq 1$ 且 $\sum_I w(i, j) = 1$ 。

两个像素 i 与 j 之间的相似性由与之相对应的灰度值向量 $v(N_i)$ 与 $v(N_j)$ 之间的相似性决定。其中， N_k 为以像素 k 为中心的固定大小的方形邻域， $v(N_k)$ 是该邻域内的灰度值组成的向量。如图 12-7 所示，图中三个像素 i 、 j_1 和 j_2 用小白点表示，以这三个像素为中心的方形邻域分别记为 N_i 、 N_{j_1} 和 N_{j_2} ，用白色的框标示出，由这三个方形邻域内的灰度值组成的向量分别记为 $v(N_i)$ 、 $v(N_{j_1})$ 和 $v(N_{j_2})$ 。设 i 为当前待处理的像素，则像素 i 与像素 j_1 、 j_2 之间的相似性由灰度值向量 $v(N_i)$ 与 $v(N_{j_1})$ 以及 $v(N_i)$ 与 $v(N_{j_2})$ 之间的相似性决定。显然，像素 i 与像素 j_1 具有相似的邻域，但像素 i 与像素 j_2 的邻域差异较大。因此，对于像素 i 的去噪， j_1 的贡献要比 j_2 大些。

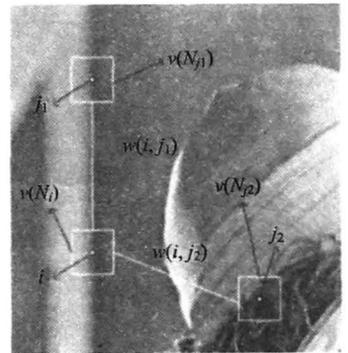


图 12-7 图像自相似性示例^[15]

各邻域之间灰度值向量的相似性通过高斯加权的欧式距离来衡量，即 $\|v(N_i) - v(N_j)\|_{2,\alpha}^2$ ，其中， $\alpha > 0$ 为高斯核函数的标准差，一般取值范围为 2~5。

$w(i, j)$ 是由以为 i 中心的邻域 N_i 和以 j 为中心的邻域 N_j 之间的相似性来决定的，由欧式距离表示的权重可以定义为

$$w(i, j) = \frac{1}{Z(i)} \exp\left(-\frac{\|v(N_i) - v(N_j)\|_{2,\alpha}^2}{h^2}\right) \tag{12.29}$$

式中， $Z(i)$ 为归一化常数，定义为

$$Z(i) = \sum_j \exp\left(-\frac{\|v(N_i) - v(N_j)\|_{2,\alpha}^2}{h^2}\right) \quad (12.30)$$

参数 h 控制指数函数的衰减速度, 因而决定着滤波的程度。该参数的取值与图像中噪声的强度成正比, 取为 $h = \lambda\sigma$, σ 为噪声图像的标准差, 一般取值范围为 0.8~1.5。

在式 (12.29) 中, 当 $i = j$ 时, 会发生过加权的现象, 为了解决这个问题, 按下式计算 $w(i, j)$ 的值:

$$w(i, j) = \max(w(i, j), \forall i \neq j) \quad (12.31)$$

由以上分析可见, 非局部平均的方法建立在对反应邻域相似性的权值进行计算的基础之上, 图像中每一个像素点的邻域要与图像中所有像素点的邻域进行一一比较, 计算量非常大。对于一幅大小为 $N \times N$ 的图像, 需要 $N^2(N^2 - 1)/2$ 对邻域比较。假定邻域大小为 $d \times d$, 则计算复杂度为 $O(N^4 d^2)$ [15]。

NLM 与其他去噪算法的比较如图 12-8 所示, 图中从左到右, 从上至下依次为加噪图像、高斯滤波去噪、各项异性滤波去噪、全变分去噪、邻域滤波去噪和 NLM 去噪效果。



图 12-8 不同去噪算法效果比较

2) NLM 的改进

NLM 使搜索窗内所有的图像块都参与计算, 而那些相似度低或者不相似的图像块会影响算法的去噪效果。另外相似度测量时, 灰度值向量取自含有噪声的图像, 不能很好地反映像素之间的相似性。针对 NLM 中相似度的衡量问题, 文献[16]提出了一种基于字典学习的度量算法。首先利用局部像素群块匹配方法消除不相似的图像块带来的干扰, 然后对含有噪声的相似块采用字典学习的方法降噪。实验结果表明, 相对于传统 NLM 算法, 该算法针对含有较多细节且结构相似性较强的图像, 细节和纹理部分的保持效果更好, 算法的健壮性也更优。

NLM 算法中使用指数型加权核函数, 容易导致图像细节因过度平滑而变得模糊。文献[17]在指数型加权核函数的基础上, 采用余弦系数加权的高斯核函数, 设计一种改进的非局部均值图像去噪算法, 并将其应用于加权系数中。实验结果表明该算法能更好地保留原图像的细节信息, 峰值信噪比最大可以提升 1.6dB。

12.4.2 BM3D

NLM 算法在图像去噪领域可以说是一个里程碑式的工作，它将图像去噪的研究推向了一个新的高峰，随后，一系列 NLM 的改进算法相继被提出。对 NLM 方法最有效的改进是 2007 年 K.Dabov 等^[18]提出的一种局部方法与非局部方法相结合的块匹配三维 (Block-Matching3D, BM3D) 协同滤波方法，这种方法同时利用了局部变换域去噪与非局部平均的优点，既引入较少的假信号，又较好地保留了图像细节，获得了比较理想的去噪效果。因此这种方法得到的图像去噪结果无论是从客观评价标准（如 SNR），还是主观视觉质量来说都是当前的最高水平的图像去噪方法。

BM3D 方法强有力的去噪性能主要来源于该方法对图像增强的稀疏表示，这种增强的稀疏性是因为块匹配后的各个图像块是彼此高度相似的，而且对块匹配群组结果执行了可分的三维变换。

1. BM3D 算法介绍

BM3D 算法的基本思想：首先将图像分成特定大小（比如 8×8 ）的分块，用距离（如欧氏距离等）来衡量图像块之间的相似性。根据距离的差别对所有的图像分块进行分组，组合处于同一组内的二维图像分块，生成一个三维数组；再运用联合滤波对生成的三维数组进行处理，最后通过逆变换，把处理后的结果返回到原图像中，从而得到去噪图像。这种方法在像素点估算过程中引入的人工噪声很小，能保留图像的大部分细节信息。去噪图像不仅有较高的信噪比，而且视觉效果也很好。

将图像按照滑窗的操作方式，分成若干个块，然后对每个参考块进行搜索，在整幅图像中搜索其相似块。根据块的相似程度，将匹配块进行组合形成一个三维矩阵。该矩阵的数据具有很高的相关性，通过三维酉变换可以有效地降低其相关性。同时，通过变换域系数的滤波，可以大幅度降低噪声。随后的三维反变换可以得到每个相似块的预估计值。对整幅图像的每个参考块都进行搜索、变换、滤波、反变换操作。最终将所有相似块的预估计值进行重构加权，得到最终的去噪结果。

1) 块匹配

整幅图像用符号 I 表示。假设图像中正在操作的当前块为 b_r ，其大小为 $N_1 \times N_1$ 。该块的矩阵数值表示为 P_{b_r} ，在搜索过程中的滑窗块表示为 b ，滑窗块的矩阵数值表示为 P_b 。在图像块大小已知的情况下，使用左上角顶点可以代表该块，那么 b_r 、 b 可以看做被代表块的左上角像素点，从而有 $b \in I$ ， $b_r \in I$ 。

对于一幅图像，首先将当前帧从左上角的块开始，按照从左到右，自上而下的顺序，平移分解成多个参考块，平移步长为 N_s 。分解完毕后，对每一个块进行块匹配操作。相似块的寻找过程如图 12-9 所示。

以当前参考块为中心，选取搜索直径为 N_d 的方形区域作为搜索区域。在搜索过程中的距离定义为两个矩阵标量取差后的模除以图像块的大小，其公式为

$$dis(P_{b_r}, P_b) = N_1^{-1} \|P_{b_r} - P_b\| \quad (12.32)$$

利用式 (12.32) 搜索到的相似块结果，组成块矩阵集合 S_{b_r} 。那么有

$$S_{b_r} = \{b \in I / \text{dis}(P_{b_r}, P_b) < \tau_{\text{dis}}\} \quad (12.33)$$

式中, τ_{dis} 是搜索过程中的距离阈值, 只有小于该阈值的搜索块才能属于集合 S_{b_r} 。由于 $\text{dis}(P_{b_r}, P_b) = 0 < \tau_{\text{dis}}$, 所以集合 S_{b_r} 至少包含一个相似块, 即 $|S_{b_r}| \geq 1$ 。

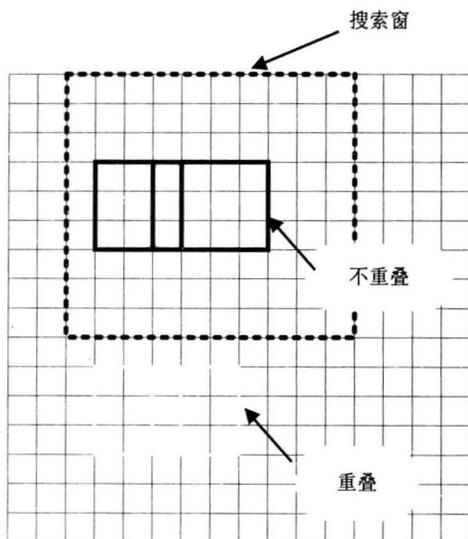


图 12-9 相似块的寻找过程

2) 三维变换域滤波

将 S_{b_r} 中的矩阵按照 $\text{dis}(P_{b_r}, P_{b_r})$ 的大小顺序堆叠, 形成大小为 $N_1 \times N_1 \times |S_{b_r}|$ 的三维矩阵 $T_{b \in S_{b_r}}$ 。对该矩阵进行 3D 酉变换, 设运算符为 κ_{3D} , 那么, 变换域去噪过程中可由式 (12.34) 表示为

$$F_{S_{b_r}} = \kappa_{3D}^{-1} \left(\gamma \left(\kappa_{3D}(T_{S_{b_r}}), \lambda_{thr} \sigma \sqrt{2 \log(N_1^2)} \right) \right) \quad (12.34)$$

式中, σ 为假设的加性高斯白噪声方差; λ_{thr} 是硬阈值滤波的阈值参数; 而 γ 函数描述如下:

$$\gamma(\lambda, \lambda_{thr}) = \begin{cases} \lambda & \text{如果 } |\lambda| > \lambda_{thr} \\ 0 & \text{其他} \end{cases} \quad (12.35)$$

将块匹配结果 $T_{S_{b_r}}$ 进行三维酉变换, 在保证不损耗能量的前提下, 将直流和噪声分量分离。处于三维矩阵能量集中顶点处的变换域系数, 包含了大部分真实图像信息, 而处在较远端的系数则是大部分噪声的携带者。通过变换域硬阈值, 可以滤除大部分噪声分量, 同时保留真实图像的信息。

3) 权值计算

对于每一次三维变换域滤波, 经过变换域硬阈值后, 假设矩阵系数的非零个数为 N_{nz} , 那么该参考块的预估计值由式 (12.36) 计算为

$$w_{b_r} = \begin{cases} \frac{1}{N_{nz}} & \text{如果 } N_{nz} \geq 1 \\ 1 & \text{否则} \end{cases} \quad (12.36)$$

在硬阈值后, 留下的非零系数越少, 说明真实图像中夹杂的噪声分量的值越小, 从

而得到的权值越大。

4) 重构

在对图像中所有的参考块按照以上描述的块匹配和三维变换域滤波后，会得到每个参考块对应的 $F_{b \in S_{b_R}}^{b_R}$ ， $\forall b_R \in I$ 。其中 b_R 代表该参考块，而 $b_R \in S_{b_R}$ 中相似块之一为 b ，而 $F_{b \in S_{b_R}}^{b_R}$ 便是该相似块经过三维滤波后的结果。由于参考块之间存在交叠，至少包含本身的相似块，并且每次都会搜索到多个相似块 $|S_{b_R}| \geq 1$ 。所以，每个像素点都会被预估多次，然后加权得到最后结果。比如 b_{R1} 、 b_{R2} 两个参考块的搜索结果经过三维变换域滤波后，结果分别为 $F_{b_1}^{b_{R1}}$ 、 $F_{b_2}^{b_{R2}}$ 。其中的相似块 b_1 、 b_2 若满足 $b_1 \cap b_2 \neq \phi$ ，那么便有像素点 $i \in (b_1 \cap b_2)$ 且 $i \in I$ ，此时像素点 i 至少被预估了两次。

所以，对于图上的任意像素点 $i \in I$ ，利用下式可计算出其最终重构结果：

$$est(i) = \frac{\sum_{b_R \in I} \sum_{b_m \in S_{b_R}} W_{b_R} F_{b_m}^{b_R(i)}}{\sum_{b_R \in I} \sum_{b_m \in S_{b_R}} W_{b_R} \chi_{b_m(i)}} \quad \forall i \in I \quad (12.37)$$

式中， $est(i)$ 表示像素点 i 的最终去噪估算结果。上式中的分子、分母，在计算最终结果时，对图像中的所有参考块所对应的相似块都进行了遍历。相应的对 $F_{b_m}^{b_R(i)}$ 、 $\chi_{b_m(i)}$ 进行如下零拓展定义：

$$F_{b_m}^{b_R(i)} = \begin{cases} b_m(i) \text{ in } F_{S_{b_R}} & \text{如果 } i \in b_m \\ 0 & \text{如果 } i \notin b_m \end{cases} \quad (12.38)$$

$$\chi_{b_m(i)} = \begin{cases} 1 & \text{如果 } i \in b_m \\ 0 & \text{如果 } i \notin b_m \end{cases} \quad (12.39)$$

这种零拓展仅仅是为了公式表示的简洁性。

2. BM3D 算法流程

BM3D 算法的流程主要分为两步，如图 12-10 所示。

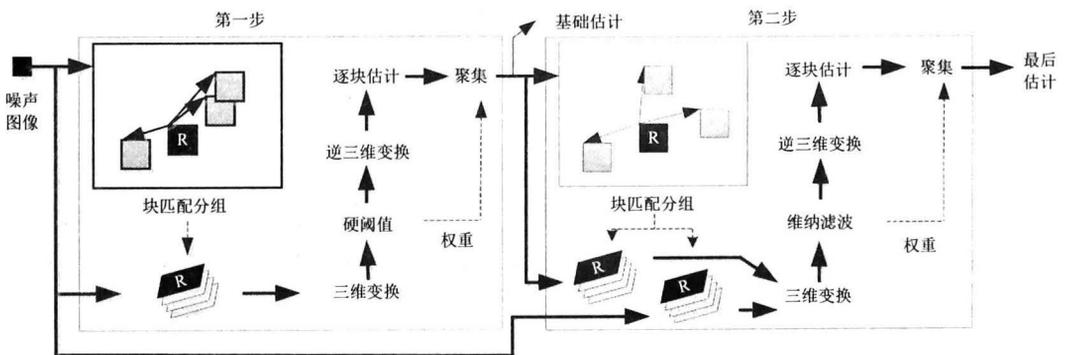


图 12-10 BM3D 算法流程图^[18]

第一步基础估计，对含噪图像中的每一块，找到它的相似块然后把它们聚集到一个三维数组中，对得到的三维数组进行三维变换，通过对变换域的系数进行硬阈值处理而减弱噪声，然后逆变换得到组中所有图像块的估计，然后把这些估计值返回到它们的原始位置。对得到的有重叠的块估计，通过对它们进行加权平均得到真实图像的基础估计。

第二步最终估计，对基础估计中的每个图像块，通过块匹配找到与它相似的相似块

在基础估计图像中的位置,通过这些位置得到两个三维数组,一个是从含噪图像中得到的,一个是从基础估计图像中得到的。对形成的两个三维数组均进行三维变换,以基础估计图像中的能量谱作为能量谱对含噪三维数组进行维纳滤波,然后逆变换得到组中所有图像块的估计,再把这些估计值返回到它们的原始位置。对得到的有重叠的局部块估计,通过对它们进行加权平均得到真实图像的最终估计。

3. BM3D 评价

BM3D 算法的第一步是块匹配,但是块匹配的计算量相当大。BM3D 算法用三维数组来表示二维图像,需要对所有的图像分块都进行分块相似性匹配并进行去噪处理,因此运算量巨大,计算复杂度达到了 $O(MN)$,其中, M 、 N 表示图像的大小。

BM3D 滤波可以抑制错误结构,同时增加图像对比度。在迭代的开始阶段, BM3D 的一个重要特点是能保持高对比度区域的边缘,并可以平滑低对比度区域容易出现的错误边缘和结构,使得重建图像在迭代初期能够在一个更大的可行区域内逼近真实结构,在迭代后期随着参数噪声的减小, BM3D 滤波器对图像的作用越来越小。

12.5 基于稀疏模型的去噪算法

近年来,过完备稀疏模型表示作为一种有效的表示模型广泛应用于各种信号和图像处理任务中。过完备稀疏模型表示采用过完备基来代替传统的正交基函数,在过完备基里寻求少量元素来表示图像的特征。由于所采用基函数的冗余性,过完备稀疏模型表示具有更好的稀疏表示能力。不同于传统去噪方法在变换域中区分信息和噪声,基于过完备稀疏模型表示的去噪方法按照是否是图像中的稀疏成分将图像中的有用信息和噪声分离,能同时满足稀疏性和信号噪声的可分性。本章主要介绍过完备稀疏模型表示的基本理论以及基于稀疏模型的典型去噪算法。

12.5.1 稀疏表示简介

1993 年, Mallat 和 Zhang^[19]第一次提出了信号在过完备字典上进行分解的思想。该思想主要内容是:通过在过完备字典上对信号进行分解,得到信号的稀疏表示。这种在变换域用尽量少的基函数来准确地表示原始信号,就是信号的稀疏表示。它开创了信号稀疏表示这一信号分析的新方向。由于信号稀疏表示的优良特性,随着研究的深入,信号稀疏表示研究很快从一维信号推广到二维图像表示的研究上,并表现出极大的优越性。尤其是近年来在数学和工程领域同时兴起的压缩传感与稀疏表示理论,使得稀疏表示理论的研究和应用越来越引起众多人士的重视。

这类方法通常先将图像分成小块,然后在一定的过完备字典下,对每个小块进行稀疏表示以达到去噪的目的,取得了很好的去噪效果,如图 12-11



(a)

(b)

图 12-11 稀疏表示去噪效果^[2]

所示, 图 (a) 为含噪图像, 图 (b) 为利用稀疏表示方法得到的去噪结果。这种方法给图像去噪提供了一种新的思路。

12.5.2 稀疏去噪原理及模型

1. 稀疏去噪原理

图像统计表明图像碎片在过完备字典上能稀疏表示。稀疏表示理论表明信号稀疏表示能对信号进行准确的重构。图像中有用的信息一般具有一定结构, 是图像的稀疏成分, 可以用过完备字典中的某些原子来表示。噪声是随机分布且没有结构的, 因此无法用字典中的原子表示。基于过完备稀疏表示理论的图像去噪方法, 根据所分析对象是否为图像的稀疏成分把图像中的有用信息和噪声进行区分, 将图像有用信息作为图像稀疏成分, 而将噪声作为图像残差 (去除其中稀疏成分后得到)。

任何噪声都可视为在真实信号上加入的干扰, 关键是如何确定有用和无用成分。稀疏信号定义为: 若信号仅有有限个非零采样点, 而其他采样点均为零 (或接近于零), 则称信号是稀疏的。但现实中时域内自然图像信号通常是非稀疏的, 严格稀疏的信号很少, 尽管有些位置的值很小, 但不一定等于零。于是引入可压缩信号概念。其定义为: 如果某一信号在不丢失任何信息 (大部分信息) 的条件下通过某种变换可以得到稀疏信号, 也就是说信号在某些变换域是稀疏的, 则称为可压缩信号。可以利用可压缩信号去逼近原始信号。

稀疏性可用于图像去噪, 主要是基于以下几个特征^[20]。

(1) 自适应性。稀疏分解不需要先验图像信号与噪声的统计特性, 通过信号在过完备字典上分解, 用以表示信号基, 可自适应地根据图像信号本身特点灵活选取。

(2) 不相关性。原始真实图像信号与观测图像信号之间的不相关性。图像有用信号间存在某种内在联系, 而噪声分布则是孤立的、随机的、离散的。

(3) 原子化。借助于原子能量特性, 对图像信号进行稀疏分解。通过形成大小两类原子, 分别提取时域上分布比较长的信息成分与比较短的信息成分 (信息细节), 利用过完备原子库的稀疏性表示图像。

2. 稀疏表示模型

过完备稀疏表示是针对不确定方程组的求解问题提出的。对于一个不确定性方程组 $Ax = b$, $A \in R^{n \times m}$, 且 $n < m$ 。在 A 满秩的条件下, 由于方程组数目小于未知数的个数, 方程有无数解。给定不同的正则约束函数, 不确定性方程组的解也会不同。过完备稀疏表示模型是从解的稀疏性来给不确定性方程组一种约束, 使得方程组有唯一解。将其引入到信号表示理论, 信号的过完备稀疏表示即是从过完备字典里寻求最少原子的线性组合来表示信号。过完备稀疏表示模型如图 12-12 所示。

过完备字典 $D \in R^{n \times K}$ 包含 K 列基本原子, $\{d_j\}_{j=1}^K$, 则一个信号 $y \in R^n$ 能用字典中的少量原子线性表示, 即 $y \approx Dx$, 有

$$\hat{y} = \arg \min_x \|y - Dx\|_2^2 \quad \text{满足 } \|x\|_0 < t \quad (12.40)$$

式中, $\|\cdot\|_0$ 为 l_0 范数, 表示一个向量的非零元的个数。式 (12.40) 的解的稀疏表示可以有以下两种方式:

$$(P_0) \min_x \|x\|_0 \quad \text{满足 } y = Dx \quad (12.41)$$

或者为

$$(P_0) \min_x \|x\|_0 \quad \text{满足 } \|y - Dx\|_2 \leq \varepsilon \quad (12.42)$$

式 (12.41) 是针对无噪声图像求解, 式 (12.42) 是针对有噪声图像求解, 一般采用式 (12.42) 进行求解。

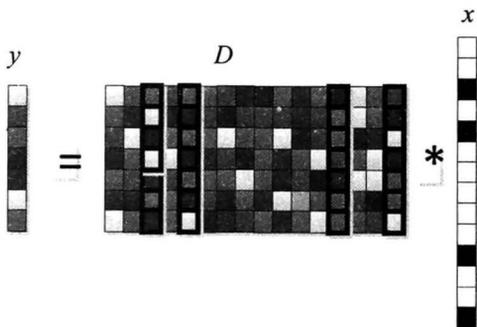


图 12-12 过完备稀疏表示模型示意

一般情况下自然图像的数据信息也具有冗余性, 因此可以在过完备字典上进行稀疏表示。如果将图像信号看作一维向量 y , 其稀疏表示是指信号 y 在过完备字典 D 下的低维投影, 投影系数即为稀疏表示系数 x 。则可以用上述的过完备稀疏表示模型对图像进行稀疏表示为

$$\hat{x} = \arg \min_x \|x\|_0 \quad \text{满足 } Dx = y \quad (12.43)$$

$$\hat{x} = \arg \min_x \|x\|_0 \quad \text{满足 } \|y - Dx\|_2^2 < \varepsilon \quad (12.44)$$

上述两式中, $\|x\|_0$ 为 l_0 范数, 即向量 x 中非零元素的个数; \hat{x} 为信号在过完备字典 D 上的稀疏表示。

12.5.3 字典构建算法

基于稀疏模型的图像去噪方法, 通常先将图像分成小块, 然后在一定的过完备字典下, 对每个小块进行稀疏表示来达到去噪的目的, 故字典构建和稀疏分解是过完备稀疏表示的两个主要方面。

字典的过完备性决定稀疏分解过程的计算复杂度, 同时也决定稀疏分解结果的简洁性和稀疏性。构建合适的过完备字典有助于得到最优稀疏表示。稀疏表示中字典的发展从最开始的正交基到冗余的正交基, 再到现在的超完备字典, 体现了向冗余发展的趋势。超完备字典中的原子能更好地解析图像信号结构, 符合人类的视觉特性, 同时对噪声与误差也具有更好的健壮性。目前, 过完备字典的构建方法可以分为以下两类。

(1) 选择目前已有的某种变换域中的正交基作为字典, 即固定字典 (如 DCT 字典、Contourlet 字典、Wavelet 字典)。Ting Fu 在文献[21]中采用了小波变换域中的基函数构

成超完备字典实现对信号的稀疏表示去噪；文献[22]提出一种基于多尺度脊波字典的稀疏性约束图像去噪算法。通过构造一个多尺度脊波字典，使得字典具有多分辨率以及多方向等特性。这些方法都具有快速变换和逆变换的特点。然而由于字典构成的单一性，不能完全有效地稀疏表示图像信号，无法最大程度分离信号和噪声。

(2) 根据训练样本学习得到过完备字典。通过训练样本可以得到一组适合某一类信号特征的字典，这类方法构造的过完备字典能更好地符合信号的结构特性，能够很好地保证信号分解的稀疏性，从而实现信号与噪声的分离。但由于需要对样本进行训练，这种方法一般运算时间较长，而且仅适用于与训练样本结构特性类似的信号，因此这类方法构造的过完备字典使用范围较小。这种方法是目前较为常用的字典构建方法。

12.5.2 节中提到的过完备稀疏表示模型可表示为如下优化问题：

$$\min_{D, X} \|Y - DX\|_2^2 \quad \text{满足 } \forall i, \|\alpha_i\|_0 \leq T \quad (12.45)$$

式中， Y 为输入信号集合； X 为 Y 对应的稀疏系数矩阵； D 为所求的过完备字典。

上述优化问题对于 (D, X) 不是凸的，但是当 D 和 X 中一个固定时，相对另一个却是凸的，所以求解该问题一般使用交替优化方法，即先在 D 固定的情况下对 X 进行优化，然后在 X 固定的情况下对 D 进行优化，不断交替地对 D 和 X 进行优化直至收敛^[23]。

字典训练有多种方法，其中较为经典的是 K-SVD 算法^[24]。该算法利用噪声图像本身提取训练样本得到过完备字典，主要分为以下两个步骤。

(1) 稀疏编码。使用跟踪算法对信号进行稀疏分解。

(2) 字典更新。假设给定稀疏变换系数矩阵，采用奇异值分解方法对残差矩阵进行分解，更新字典中对应的原子；与此同时更新对应的稀疏表示系数。再返回到稀疏编码步骤，迭代循环直至最优。

基于 K-SVD 的去噪算法虽然能够有效区分图像细节信息和噪声，但随着噪声的增大，该方法会出现过拟合和自由度过大的问题，导致图像纹理细节丢失，出现模糊。

其他字典构建算法如最优方向法 (Method of Optical Directions, MOD) 采用式 $D = YX(XX^T)^{-1}$ 更新字典。Online Learning 算法^[25]采用梯度下降的方法求解，该算法的优点是效率大大提高，且支持增量式字典学习。

目前基于训练样本学习的字典构建方法还有最大似然法、最大后验概率法、组合正交基学习算法、广义 PCA 算法、CNDL-FOCUSS 算法^[26]和 KLLD 算法^[27]等。

12.5.4 稀疏分解算法

基于过完备字典的稀疏表示的最优解问题是一个 NP 问题。发展至今，针对该问题较为经典的稀疏分解方法主要包括贪婪算法^[28]和全局最优化方法^[29]。

贪婪算法的原理是在过完备字典中提取与信号最匹配的原子，从而得到信号的稀疏表示。1994 年，S. Mallat 等人^[30]提出了基于稀疏表示理论的匹配追踪 (Matching Pursuit, MP) 算法。MP 算法的原理是根据一定的规则在每次迭代的过程中依次从字典里选取用于稀疏表示的最佳原子，最终得到图像信号的稀疏表示。但 MP 算法中采用的字典不是正交的，因此在选定的原子空间中，得到的信号稀疏表示不一定是最优的。之后人们对 MP 算法进行了各种改进，提出了正交匹配追踪 (Orthogonal Matching Pursuit, OMP)

算法^[31]、弱贪婪算法^[32]及其变种^[33]等。

OMP 算法在分解过程的每一步均对所选原子进行 Gram-Schmidt 正交化处理, 再将信号投影到正交化后的原子所构成的空间上, 获取信号在各个已选原子上的投影分量和残余分量; 再循环对残余分量进行分解。经过 M 次迭代分解之后, 原信号可以表示为 M 个原子的线性组合。在每一步分解中, 原子都要满足相应的条件才能被选为最佳原子, 因此残余分量会随着分解快速变小, 这样用少量原子组合就可以表示原始信号。与 MP 算法相比, 要达到相同的精度要求, OMP 算法能够选用更少的原子对信号进行稀疏表示, 同时可保持一个较高的收敛速度对信号进行逼近。

全局优化方法的原理是预先选取目标函数 (典型的目标函数一般为凸函数), 使其在预先设定的阈值条件下获得最小值, 该目标函数的局部最优解即为全局最优解。目前较为经典的全局优化方法包括基追踪 (Basis Pursuit, BP) 算法^[34]和在此基础上改进的基追踪去噪 (Basis Pursuit Denoising, BPD) 算法^[35]。

除了上面给出的稀疏分解经典算法, 现有的稀疏分解的算法主要包括以下几种。

(1) FOCUSS (Focal Underdetermined System Solver) 算法^[36]。FOCUSS 算法是一种非参数算法。它的优点是能从有限的的数据中找出局部解^[37]。

(2) 框架方法^[38]。

(3) 非线性规划方法。例如, 内点法 (Interior Point Method)^[39]、块调整松弛 (Block Coordinate Relaxation)^[40], 它们都能收敛到基追踪的解。

(4) Bayesian 算法^[41]。该算法基于概率学理论, 通过 Bayesian 方法获得信号的稀疏表示。

(5) 最优正交基法^[42]。

12.5.5 稀疏表示去噪效果

图 12-13 展示了过完备 DCT 字典和通过训练样本得到的全局训练字典。图中将字典原子变为 8×8 的像素块, 使字典的表达形式更为直观。其中, DCT 字典也被用于初始字典学习中的字典。全局字典由 K-SVD 算法训练得到, 训练所需的样本集来自干净图像, 并且这些图像都和后续的去噪实验使用的图像无关。图中自适应字典是从需要处理的噪声图像本身获取训练样本集, 再进行字典学习, 根据要处理的图像的不同特点来调整字典, 最终得到图像的自适应字典 (所选的训练图像为 Zebra, 大小为 256×256)^[42]。

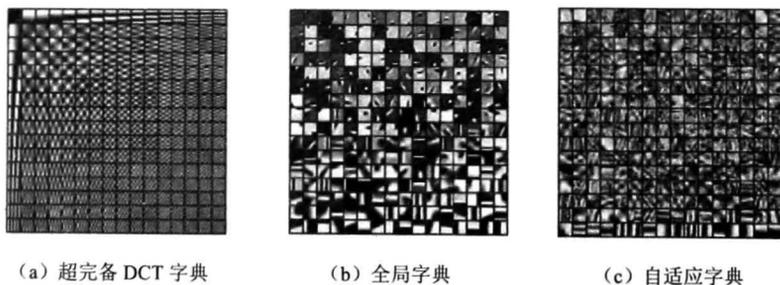


图 12-13 不同的过完备字典^[42]

将加入噪声 ($\sigma = 20$) 的 Barbara 图像作为实验图像, 分别用上述三种过完备字典

进行去噪，去噪效果如图 12-14 所示。

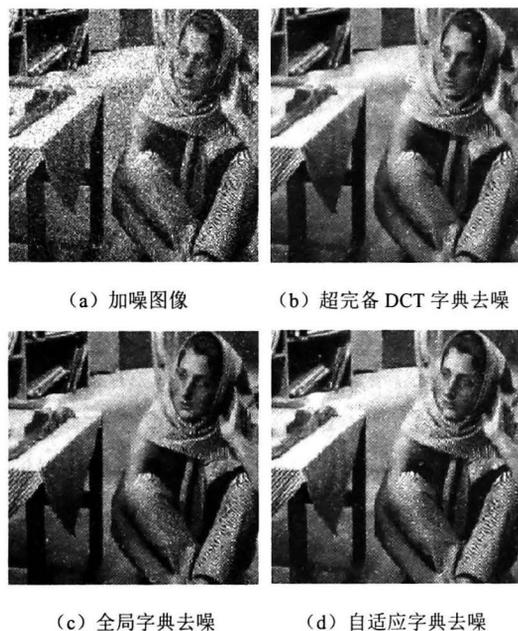


图 12-14 Barbara 图像去噪效果^[42]

12.6 总 结

随着各种数字仪器和数码产品的普及，数字图像已成为人类活动中最常用的信息载体，因为其包含着人们关注的大量信息，成为人们获取外界原始信息的主要途径。然而在数字图像的获取、传输、存储和处理过程中常常会受到各种因素的干扰形成噪声并使得图像质量下降。不仅如此，图像噪声还会影响其他图像处理算法的效果和效率，如图像去模糊、目标识别、背景建模等，所以为了获取高质量数字图像，能够保持原始信息完整性的同时，又能够去除信号中无用的信息的去噪算法一直受到研究人员的关注。过去数十年里，作为一种重要的图像处理技术，研究人员对图像去噪滤波算法进行了广泛而深入的研究，提出了很多经典的滤波和变换域处理方法。

本章对图像去噪技术进行了概述，包括噪声的概念和去噪原理，并对一些基本的图像去噪方法做了介绍。由于篇幅有限，本章并没有涵盖所有方法和理论。况且，仍有很多去噪方面的新思想、新方法出现，不断地充实图像去噪方法。而且，噪声的研究范围也在不断扩大，由高斯噪声到非高斯噪声。去噪技术有很广泛的应用和研究的前景，而且，研究领域也在不断地拓展。

绝大多数现有的图像去噪算法是基于单幅图像的，随着近年来，在安防领域从模拟照相机到高清数码摄像机的更新换代，视频去噪技术开始有了广泛的应用，去噪也成为摄像机 ISP 处理中一个很重要的模块，称为评价摄像机性能的主要指标。本书前面讲到的 ISO 会使得噪声水平加大，基于视频的去噪方案，成为应用工业界急需的技术。视频去噪不能简单看成单张图像的传统去噪方法在连续图像的重复使用。因为传统的去噪研

究主要针对图像进行滤波,未考虑视频信号的时域与空域相关性。没有充分利用视频信号的时空联系进行去噪。因此,在原有的帧内 2D 空间的去噪技术降噪的基础上,增加了帧与帧之间降噪的功能的视频去噪方法也称为 3D 降噪。本章由于篇幅的关系并没有涉及这方面的知识和算法。随着摄像机向超高清发展,对于去噪算法的要求也越来越苛刻,虽然在计算处理的硬件上有所提升,众多研究领域的视频去噪算法既要做到实时,又要有好的效果,还需要在摄像机本身的平台上运行,也是未来一个具有挑战性的课题。

参 考 文 献

- [1] CHATTERJEE P, MILANFAR P. Is Denoising Dead? [J]. IEEE Transactions on Image Processing, 2010, 19 (4): 895-911.
- [2] MARIAL J, ELAD M, SAPIRO G. Sparse Representation for Color Image Restoration [J]. IEEE Transactions on Image Processing, 2008, 17(1).
- [3] Gonzalez R C, Woods R E, Eddins S L. Digital image processing using MATLAB [M]. Pearson Education India, 2004.
- [4] DONOHO D, JOHNSTONE I. Minimax Estimation via Wavelet Shrinkage[J]. The Annals of Statistics, 1998, 26 (3): 879-921.
- [5] ABRAMOVICH F, SAPANTINAS T, SILVERMAN B. Wavelet Thresholding via a Bayesian Approach[J]. 2002, 60 (4): 725-749.
- [6] HANSEN M, YU B. Wavelet Thresholding via MDL for Natural Images[J]. IEEE Transactions on Information Theory, 2000, 46 (5): 1778-1788.
- [7] GAO H, BRUCE A. Waveshrink and Semisoft Shrinkage. StatSci Division of MathSoft, Inc. 1995.
- [8] KAUR L, GUPTA S, CHAUHAN R. Image Denoising Using Wavelet Thresholding[J]. ICVGIP, 2002, 2:16-18.
- [9] PERONA P, MALIK J. Scale-Space and Edge Detection Using Anisotropic Diffusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12 (7): 629-639.
- [10] 韩希珍, 陈朝东, 赵建. 基于 PDE 的非线性图像去噪与增强[J]. 液晶与显示, 2011, 26 (001): 111-115.
- [11] RUDIN L, OSHER S, FATEMI E. Nonlinear Total Variation based Noise Removal Algorithms[J]. Physica D: Nonlinear Phenomena, 1992, 60 (1): 259-268.
- [12] EFROS A, LEUNG T. Texture Synthesis by Non-parametric Sampling [C]. The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, 2: 1033-1038.
- [13] BUADES A, COLL B, MOREL J. A Review of Image Denoising Algorithms with a New One [J]. Multiscale Modeling and Simulation, 2005, 4 (2): 490-530.
- [14] BUADES A, COLL B, MOREL J. A Non-local Algorithm for Image Denoising [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, vol. 2, pp. 60-65, 20-25 June.
- [15] 尹勇, 孙伟峰, 卢洁, 等. 一种基于非局部平均的 PET 图像去噪方法 [J]. 生物医学工程学杂志, 2010, 27(2): 274-277.
- [16] 崔学英, 张权, 桂志国. 基于字典学习的非局部均值去噪算法[J]. 计算机应用, 2013, 33 (5): 1420-1422.
- [17] 刘晓明, 田雨, 何微, 等. 一种改进的非局部均值图像去噪算法[J]. 计算机工程, 2012, 38 (4): 199-201.
- [18] DABOV K, FOI A, EGIAZARIAN K. Video Denoising by Sparse 3D Transform-domain Collaborative Filtering [C]. 15th European signal processing conference (eusipco) 2007, September 3-7.
- [19] MALLAT S, ZHANG Z. Matching Pursuit with Time-frequency Dictionaries [J]. IEEE Transactions on Signal Processing, 1993, 41 (12): 3397-3415.
- [20] 王建英, 尹忠科, 张春梅. 信号与图像的稀疏分解及初步应用[M]. 成都: 西南交通大学出版社, 2006.
- [21] TING F, HUAFU C, DEZHONG Y. Noise Reduction by a New Iterative Weighted Sparse Decomposition Algorithm [C]. IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, 2002, 2: 909-913.
- [22] 邓承志. 基于多尺度脊波字典的图像去噪算法[J]. 计算机工程, 2010, 36 (23): 207-208.
- [23] MAIRAL J, FRANCIS B, JEAN P. Online Dictionary Learning for Sparse Coding [J]. ICML, 2009.

- [24] AHARON M, ELAD M, BRUCKSTEIN. The K-SVD: an Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation [J]. IEEE Transactions on Signal processing, 2006, 54 (11): 4311-4322.
- [25] MAIRAL J, BACH F, PONCE J. Online Learning for Matrix Factorization and Sparse Coding [J]. Journal of Machine Learning Research, 2010, 11 (1): 19-60.
- [26] DELGADO K, MURRAY J, RAO B. Dictionary Learning Algorithms for Sparse Representation [J]. Neural Computation, 2003, 15 (2): 349-396.
- [27] CHATTERJEE P, MILANFAR P. Clustering-based Denoising with Locally Learned Dictionaries [J]. IEEE Transactions on Image Processing, 2009, 18 (7) , 1438-1451.
- [28] FRIEDMAN J, STUETZLE W. Projection Pursuit Regression [J]. Journal of the American Statistical Association, 1981, 76 (376): 817-823.
- [29] CHEN S, DONOHO D, SAUNDERS M. Atomic Decomposition by Basis Pursuit [J]. SIAM Journal Scientific Computing, 1999, 20 (1): 33-61.
- [30] MALLAT S, ZHANG Z. Matching Pursuit with Time-frequency Dictionaries [J]. IEEE Transactions on Signal Processing, 1993, 41 (12): 3397-3415.
- [31] PATI Y, REZAIFAR R, KFISHNAPRASED P. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition [C]. The Proceedings of 27th Asilomar Conference on signals. Systems and Computers. 1993, 40-44.
- [32] TEMLYAKOV V. Weak Greedy Algorithms [J]. Advances in Computational Mathematics, 2000, 12 (2): 213-227.
- [33] TEMLYAKOV V. Greedy Algorithms in Banach Spaces [J]. Advances in Computational Mathematics, 2001, 14 (3): 277-292.
- [34] CHELA S, DONOHO D, SAUNDCRS M. Atomic Decomposition by Basis Pursuit [J]. SIAM Journal on Scientific Computing, 1998, 20 (1): 33-61.
- [35] TROPP A. Topics in sparse approximation. Stanford University, CA, 1995.
- [36] MURRAY J, DELGADO K. An Improved FOCUSS-based Learning Algorithm for Solving Sparse Linear-Inverse Problems [C]. The Proceedings of the IEEE 35th Asilomar Conference, 2001, 1: 347-351.
- [37] K. Delgado, J. Murray, B. Rao. Dictionary Learning Algorithms for Sparse Representation [J]. Neural Computation, 2003, 15 (2): 349-396.
- [38] DAUBECHIES L. Time-Frequency Localization Operators: A Geometric Phase Space Approach [J]. IEEE Transactions on Information Theory, 1998, 34 (4): 605-612.
- [39] SHANNO D, VANDERBEI R. Interior-point Methods for Non-convex Nonlinear Programming: Orderings and Higher-Order Methods [J]. Mathematical Programming, 2000, 87 (2): 303-316.
- [40] SARDY S, BRUCE A, TSENG. Block Coordinate Relaxation Methods for Nonparametric Wavelet Denoising [J]. Journal of Computational and Graphical Statistics, 2000, 9 (2): 361-379.
- [41] WIPFAND D, RAO. Bayesian Learning for Sparse Signal Reconstruction [C]. The Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2003, 6: 601-604.
- [42] COIFMAN R, WICKERHAUSER M. Entropy-based Algorithms for Best Basis Selection [J]. IEEE Transactions on Information Theory, 1992, 38 (2): 713-718.

第 13 章 压缩成像技术

13.1 概 述

在现代数字信号处理的基础上,信号采样是从模拟的物理世界通向数字的信息世界的必备手段。传统的信号采样建立在 Nyquist 采样定理基础之上,要求采样频率必须大于信号最高频率的两倍,才能精确重构信号。随着社会的进步,科技的发展,人们对信息量的需求日益增加,随之出现了大量高分辨率的数码设备,通过这些设备采样产生了庞大的数据,海量数据的存储与传输成为一大难题,对信号处理能力提出了更高要求。在实际应用中,为了降低存储、处理和传输的成本,人们常采用数据压缩技术,用较少的数据量表示信号,大量的冗余数据被抛弃。这种“先采样,再压缩”的信号处理流程浪费了大量的采样资源,于是很自然地引出一个问题:既然采样得到的大部分数据是冗余的,可以被丢弃,能否将采集和压缩结合在一起,直接采集那部分重要的、在最后没有被丢弃的数据,并能够精确重构原始信号呢?

近几年来 Donoho 和 Candès^[1~5]等人提出的一种新颖的理论——压缩感知(Compressed Sensing 或 Compressive Sampling, CS)表明这是可能的。压缩感知建立在信号的稀疏性或可压缩性基础上,其主要目标是从少量的非自适应线性测量集中准确且有效地重建一个信号。压缩成像(Compressive Imaging, CI)^[6~8]技术是在 CS 理论的基础上发展起来的崭新科学研究方向,是压缩感知理论在成像领域的具体应用。CS 理论建立在信号的稀疏性或可压缩性基础上,提供了将模拟信号直接采样压缩为数字信号的有效途径,突破了现有的“先采集,再压缩存储”的信号处理模式。而数字图像自然场景的数字化描述,在特定的稀疏基下一般都具有可压缩性。因此,只要能够找到合适的稀疏基,就可以有效地同步进行压缩采样,利用较少数目的探测器获取高分辨率的目标图像。

13.2 压缩感知基本理论

CS 理论在采样的同时对信号进行压缩编码,利用信号的稀疏性或可压缩性,以远低于 Nyquist 采样率对信号进行自适应的测量编码,然后利用测量值通过稀疏重构算法高概率地恢复出原始信号。压缩感知理论主要包括信号的稀疏表示、测量矩阵和信号重构算法三个方面。

13.2.1 稀疏表示

信号的稀疏表示是指将信号投影到某个变换基时,如果变换系数中只有 K 个非 0 元素,而其他 $N-K$ 个系数为 0 或者绝对值非常小,则称信号是 K 稀疏的。信号在某种表示方式下的稀疏性,是包括压缩感知在内的很多现代信号处理理论的基础。这里的表示

方式，可以是特定完备基、超完备字典或框架下的变换，如 DCT 变换、傅里叶变换、小波变换或其他多尺度表示方法；也可以是通过数据驱动方法学习、训练得到的字典（图像片段集合）。

考虑 R^N 空间中长度为 N 的离散时间信号 x ，给定正交基矩阵 $\Psi \in R^{N \times N}$ ， x 在基 Ψ 下可表示为

$$x = \sum_{i=1}^N \alpha_i \psi_i \text{ 或 } x = \Psi \alpha \tag{13.1}$$

式中， α 为 x 在基 Ψ 下的变换系数。如果 x 在基 Ψ 下是 K 稀疏，式 (13.1) 可以表示为

$$x = \sum_{i=1}^K \alpha_{ni} \psi_{ni} \tag{13.2}$$

式中， $\{ni\}$ 是 K 个非 0 元素对应的系数列向量和基向量的索引，此时， α 是只有 K 个非 0 元素的 N 维列向量，即 $\|\alpha\|_0=K$ ，其中 $\|\cdot\|_p$ 为向量的 l_p 范数，定义为

$$\|\alpha\|_p = \left(\sum_i |\alpha_i|^p \right)^{\frac{1}{p}} \tag{13.3}$$

当 $p=0$ 时，式 (13.3) 等价于计算向量 α 中非 0 元素的个数。

合理选择变换基 Ψ ，使信号的 l_0 范数尽可能得小，不仅有利于提高采集信号的速度，而且有利于减少存储、传输信号所占用的资源。信号的稀疏性是压缩感知理论的先验条件，即信号必须在某种变换下可以稀疏表示。

13.2.2 测量矩阵

在 CS 中，式 (13.1) 中向量 α 的 K 个非 0 元素不是直接测量得到，而是通过 $M \times N$ ($M \ll N$) 维测量矩阵 Φ 对向量 α 进行线性投影，将高维信号投影到低维空间，从而获得观测信号为

$$y = \Phi x = \Phi \Psi \alpha = \Theta \alpha \tag{13.4}$$

式中， y 为 M 维测量数据； Φ 为 $M \times N$ 维测量矩阵； Ψ 为所采用的稀疏基； Θ 为 $M \times N$ 维压缩感知矩阵。由于观测值 y 的维数 M 远小于原始信号 x 的维数 N ，在没有更多约束条件的情况下满足条件的原始信号 x 是无穷多的。为了保证信号能够精确重构，要求压缩感知矩阵 Θ 满足有限等距性质 (Restricted Isometry Property, RIP) [9]。

定义： $M \times N$ 维测量矩阵 Φ 的约束等距常数 δ_k 指的是对于所有的 K 稀疏信号 x 满足不等式：

$$(1 - \delta_k) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k) \|x\|_2^2 \tag{13.5}$$

的最小常数。

如果式 (13.5) 成立，则称矩阵 Φ 对恰当的 K 和 δ_k 满足 RIP。

Baraniuk^[10]证明了当矩阵 Φ 和 Ψ 不相关时， Θ 满足有限等距性质。直接验证压缩感知矩阵是否满足有限等距性质很难做到，但由于变换基 Ψ 是固定的，可以通过设计测量矩阵 Φ 使得 Θ 满足有限等距性质。Baraniuk^[11]等同时还证明当 Φ 为高斯随机矩阵时， Θ 能够很大概率地满足有限等距性。常见的能使压缩感知矩阵满足约束等距性质的测量矩阵还包括二值随机贝努里矩阵、局部傅里叶矩阵、部分哈达玛矩阵、托普利兹矩阵以及

循环-托普利兹矩阵等。

13.2.3 重构算法

CS 重构指的是由压缩采样得到的长度为 M 的观测向量 y 恢复出长度为 N ($M \ll N$) 的高维原始信号 x , 其实质是求解一个最小 l_0 范数问题:

$$\hat{\alpha} = \min \|\alpha\|_0 \quad \text{满足 } \Theta\alpha = y \quad (13.6)$$

式 (13.6) 的求解是一个 NP-hard 的非凸优化问题, 需要穷举向量 α 中所有可能的 C_N^K 个非零项的组合, 直接求解非常困难。因此需要寻找可用多项式时间求解的近似模型和算法, 常用的主要有最小 l_1 范数凸优化模型、贪婪算法、最小 l_p ($0 < p < 1$) 范数非凸优化模型、平滑 l_0 算法以及一些其他组合算法。图 13-1 总结了目前常用的稀疏重构模型、算法及其相互之间的关系。这里主要介绍在图像处理领域应用最广泛的最小 l_1 范数凸优化模型。

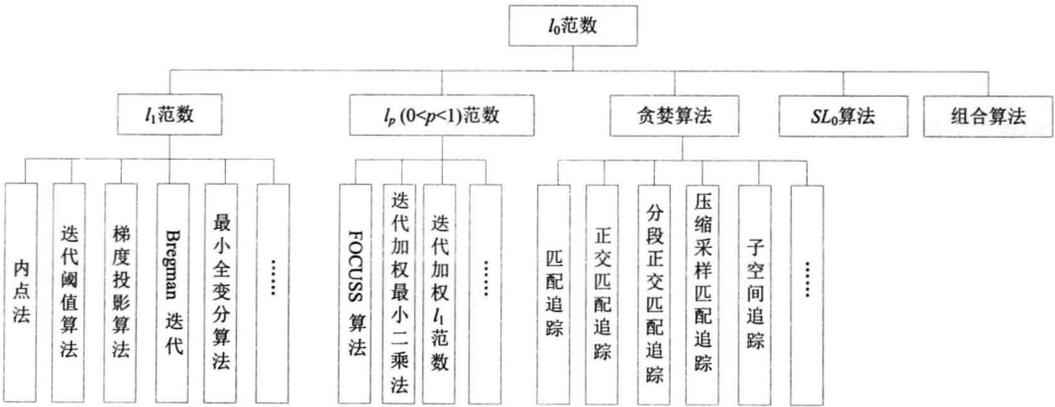


图 13-1 重构算法关系图

基于 l_1 范数凸优化的稀疏重构模型描述为

$$\hat{x} = \min \|x\|_1 \quad \text{满足 } \|y - \Phi x\|_2 \leq \sigma \quad (13.7)$$

式 (13.7) 的求解可以通过拉格朗日准则, 将约束条件转化为惩罚项, 从而构建非约束极值问题:

$$\hat{x} = \min \|y - \Phi x\|_2^2 + \lambda \|x\|_1 \quad (13.8)$$

式 (13.7) 的问题模型是一个二阶锥规划问题, 利用内点法可实现锥规划问题的求解。然而成像系统中涉及的稀疏重构一般为大数据量问题, 基于内点法的优化算法, 由于计算复杂度较高 (需要求解大规模线性方程组) 等原因, 而不适合用于成像系统的稀疏重构算法。在保证求解精度的条件下, 为了大幅提高计算效率, 学者们提出了许多其他快速算法, 其中包括不动点延拓 (Fixed Point Continuation, FPC)^[12]、Bregman 迭代^[13]以及最小全变分^[4, 14]算法等。

13.3 压缩成像技术

压缩成像技术是压缩感知理论在光学成像中的崭新应用, 其目的是从源头上降低所

需采集的图像数据量，从根本上解决海量图像数据的生成和传输问题，使得采用低成本的传感器实现高分辨率的数字成像成为可能。压缩成像技术突破了传统的成像系统设计理念，利用硬件实现场景的非自适应线性投影，从而达到利用低分辨率的探测器获取高分辨率图像的目的。

13.3.1 单像素照相机

2007年，美国的Rice大学基于压缩感知理论首次成功研制出单像素照相机^[15, 16]。其设计原理是通过光路系统将成像目标投影到数字微镜设备(DMD)，以DMD作为测量矩阵，将接收的入射光线反射，通过透镜聚焦到单个光敏二极管，通过A/D转换以数字信号的形式记录下来，即为关于成像目标的一个测量值，如图13-2所示。由于DMD是由数字电压信号控制微镜片的机械运动完成对入射光线的调整，因而可实现0~1的伪随机测量矩阵的构建。如果用 I 表示成像场景对应的数字图像，表示 f 按列序组织成一列形成的向量，则感光器件上的电压就等于测量值 $y_i = \langle \phi_i, f \rangle$ ，将这一过程重复 M 次即可以得到全部的测量值向量 $y = \Phi f$ 。在获得了一系列测量值之后，可以根据CS重建原理选择适当的优化算法从这些观测值中重建出原始图像，如图13-3所示，单像素相机针对 64×64 的字母图像“R”，在测量次数 M 分别为820和1638时的重构结果。

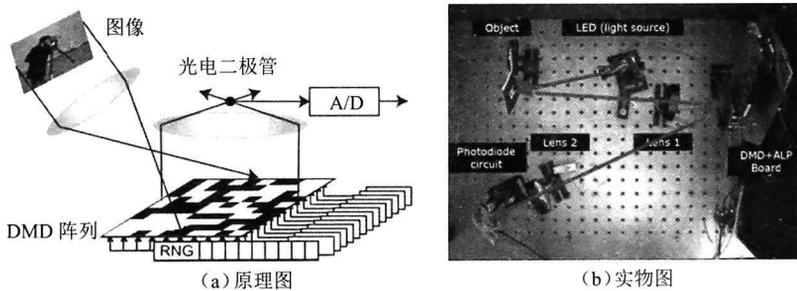


图 13-2 单像素照相机^[16]



图 13-3 实验结果^[16]

单像素照相机直接获取的是 M 次随机线性测量而不是原始图像的 N 个像素值，这就为采用低像素照相机拍摄高分辨率图像提供了可能。然而单像素照相机采用DMD对成像目标进行编码，需要连续采样投影 M 次才能获得重建原始图像所需的全部测量值，因此成像系统较为耗时，无法进行实时图像获取。

13.3.2 基于编码孔径的压缩成像

美国杜克大学的研究小组将压缩感知理论和编码孔径成像技术相结合,提出了一种压缩编码孔径(Compressive Coded Aperture, CCA)成像机制^[17, 18],当目标图像在变换基下可稀疏表示时,可通过低分辨率观测值重构出高分辨率图像。该技术利用光线的直线传播特性,入射光线经编码孔径后直接投影在探测器像面上,每个物点在探测器像面上形成一个编码孔径的投影图像。不同物点产生的像因相互错开、叠加而在接收平面上形成退化的、重叠的二维分布信号,即编码像,整个成像光路原理如图 13-4 所示。

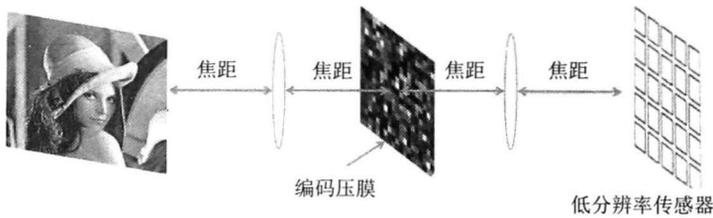


图 13-4 编码孔径光路示意

压缩编码孔径的成像模型可以用式 (13.9) 表示为

$$y = \Phi^{\text{CCA}} f = D(f \cdot H^{\text{CCA}}) \quad (13.9)$$

式中, y 为测量值; f 为原始图像; D 表示下采样操作; Φ^{CCA} 为编码掩膜对应的测量矩阵; H^{CCA} 表示编码掩膜。这里, 编码模板的大小和需要重建的图像分辨率相同, 该分辨率大于图像传感器阵列的分辨率。因此, 式 (13.9) 中模拟此成像过程为先和编码模板卷积再进行降采样。

设 F 为 $n \times n$ 的傅里叶矩阵, 其元素为 $F_{k,l} = e^{2\pi i(k-1)(l-1)/n}$, 令 $\mathcal{F} = F \otimes F$, 其中, \otimes 表示克罗内克积 (Kronecker Product), 对于 $n \times n$ 的矩阵 X 进行二维傅里叶变换有

$$\text{Vec}(FXF) = (F^T \otimes F)\text{Vec}(X) = (F \otimes F)\text{Vec}(X) = \mathcal{F}\text{Vec}(X) \quad (13.10)$$

式中, $\text{Vec}(\cdot)$ 为向量化算子; $\text{Vec}(X)$ 表示依 X 的列顺序将 X 转换为一个列向量。

结合式 (13.10) 有

$$\text{Vec}(f \cdot H^{\text{CCA}}) = \mathcal{F}^{-1} C_H \mathcal{F} f \quad (13.11)$$

式中, C_H 为对角矩阵, 其对角元素为 $\mathcal{F} H^{\text{CCA}}$ 的元素。根据式 (13.9) 有

$$y = \Phi^{\text{CCA}} f = D(\mathcal{F}^{-1} C_H \mathcal{F}) f \quad (13.12)$$

压缩编码孔径成像的目标是合理设计测量矩阵 Φ^{CCA} 以求解编码掩膜 H 。基于编码孔径的压缩成像系统光学结构紧凑、易于实现。针对目标图像的测量可一次性完成, 更加适用于动态场景和视频的压缩成像。

图 13-5 给出了基于压缩编码孔径进行图像分辨率增强的实验结果, 图 13-5 (a) 所示为原始图像, 图 13-5 (b) 所示为传统成像模型下, 低分辨率传感器采集图像, 图 13-5 (c) 所示为基于压缩编码孔径模型, 低分辨率传感器采集图像, 图 13-5 (d) 所示为传统的分辨率增强方法得到的高分辨率图像, 图 13-5 (e) 所示为在 CS 框架通过稀疏重构算法得到的高分辨率图像。对比图 13-5 (d) 和图 13-5 (e), 后者的边缘明显比前者锐利, 细节更加丰富。



图 13-5 仿真实验结果

13.3.3 CMOS 压缩成像

传统的 CMOS 图像传感器信号采集处理流程如图 13-6 所示，对采集原始数据的压缩是在 A/D 转换之后进行的，对于高分辨率、高帧率图像的获取，需要高速率宽频带的 A/D 转换模块，不仅导致了功耗的提高，同时大大增加了制造成本；在压缩编码时仅对采样数据变换（DCT 或者小波）后少数绝对值较大的系数进行，而抛弃大量为零或者接近于零的系数，从而对采样资源造成大量浪费。

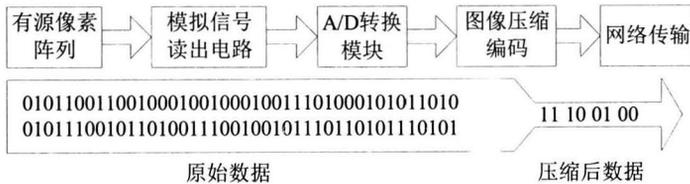


图 13-6 传统 CMOS 信号处理流程

如果在 CMOS 图像传感器中引入压缩感知技术，在模拟信号光电转换之后，直接进行压缩感知采样，将信号采集和压缩同时进行，然后对压缩采样的数据进行 A/D 转换，可以有效减少输入 A/D 转换模块的数据量，提高 A/D 转换的速度，降低系统功耗。基于压缩感知的 CMOS 图像传感器信号采集处理流程如图 13-7 所示。

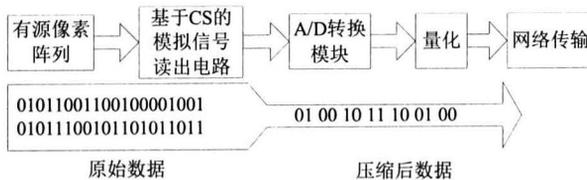


图 13-7 基于压缩感知的 CMOS 信号采集处理流程

瑞士联邦理工学院的 Jacques^[19]等人基于上述思路,在 CMOS 图像传感器中引入压缩感知技术,通过随机卷积实现模拟信号的压缩采样,设计开发出一款 CMOS 压缩感知成像器件。成像器件由标准的 CMOS 传感器和位移寄存器构成,如图 13-8 所示。紧邻 CMOS 像素单元的位置额外设置有 1 位的闪存,用于存储卷积滤波的系数 a_i 。CMOS 像素单元中的测量值将与存储在该 1 位闪存中的卷积滤波系数相乘作为输出。存储在 1 位闪存中的卷积系数按照从左至右,从上至下的顺序遍历整个传感器。系统采用多路复用的形式在生成的卷积图像中随机抽取 M 个像素作为压缩感知图像输出。本质上,该方法与基于编码孔径压缩成像系统的数学理论是一致的,都是在生成的卷积图像上下采样生成测量值,即满足

$$y_i = (\Phi f)_i = \sum_j a_{r(i)-j} f_j = (f * a)_{r(i)} \quad (13.13)$$

式中, a 是随机卷积滤波器。与编码孔径压缩成像系统不同的是,卷积编码 CMOS 压缩成像系统是通过改造 CMOS 硬件器件完成的。CMOS 成像的优点在于:压缩采样是通过 CMOS 电路在模拟域完成的,可以有效降低数据率,缓解 A/D 转化压力的同时降低数据存储和传输压力。

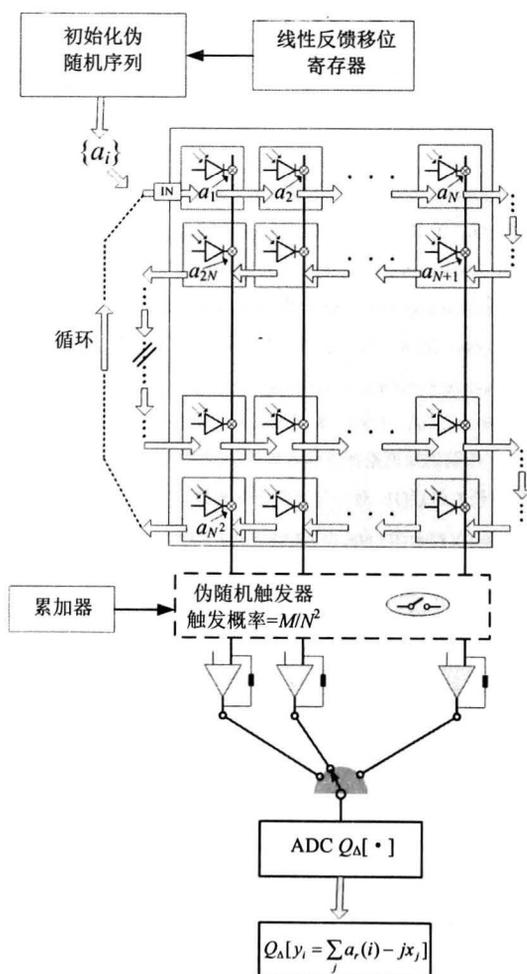


图 13-8 CMOS 压缩成像示意^[19]

13.4 总 结

压缩感知理论利用信号固有的稀疏性或可压缩性,将信号的采集和压缩结合在一起,以远低于 Nyquist 采样率对原始信号进行采样,然后通过稀疏重构算法得到原始信号,从而有效缓解了高维信号的采集、传输以及存储的压力。目前关于压缩感知在成像中的应用大多停留在数据仿真和实验室原型研究层面,研究主要集中在如何利用硬件实现目标物体的非自适应线性投影以及如何利用信号恢复算法重构原始图像。如何根据压缩感知理论制造实用的成像设备,以下几个问题尚待解决。

(1) 测量矩阵的物理实现。实际物理器件无法实现测量矩阵中的元素取负值功能,测量矩阵的非负值约束必将对重构算法造成影响;即使确定性的测量矩阵,光学器件的精度也难以保证,通过光学设备实现也比较困难。

(2) 实时压缩的实现。要使压缩成像系统实用化,必须要解决图像实时获取问题,然而重构算法的计算量一般比较大,难以达到实时要求。

参 考 文 献

- [1] DONOHO D. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52 (4): 1289-1306.
- [2] CANDÈS E, WAKIN M. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine, 2008, 25 (2): 21-30.
- [3] DONOHO D, TSAIG Y. Extensions of compressed sensing[J]. Signal Processing, 2006, 86 (3): 533-548.
- [4] CANDÈS E, ROMBERG J, TAO T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory, 2006, 52 (2): 489-509.
- [5] CANDÈS E, TAO T. Near Optimal signal recovery from random projections: universal encoding strategies?[J]. IEEE Transactions on Information Theory, 2006, 52 (12): 5406-5425.
- [6] HAUPT J, NOWAK R. Compressive sampling vs. conventional imaging[C]. In Proceedings of International Conference on Image Processing (ICIP), Atlanta, GA, USA, 8-11 October 2006: 1269-1272.
- [7] 肖龙龙, 刘昆, 韩大鹏, 等. 压缩感知理论在光学成像中的应用[J]. 应用光学, 2012, 33 (1): 71-77.
- [8] 陈靖, 王涌天. 压缩成像技术研究进展[J]. 激光与光电子学进展, 2012(49): 030002.
- [9] CANDÈS E. The restricted isometry property and its implications for compressed sensing[J]. Academic des Sciences, 2006, 346 (1): 589-592.
- [10] BARANIUK R. A lecture on compressive sensing[J]. IEEE Signal Processing Magazine, 2007, 24 (4): 118-121
- [11] BARANIUK R, DAVENPORT M, DEVORE R, et al. A simple proof of the restricted isometry property for random matrices[J]. Constructive Approximation, 2008, 28 (3): 253-263.
- [12] HALE E, YIN W, ZHANG Y. A fixed-point continuation method for l_1 -regularized minimization with application to compressed sensing[R]. CAAM Technical Report TR07, 2007.
- [13] YIN W, OSHER S, GOLDFARB D. Bregman iterative algorithms for l_1 minimization with applications to compressed sensing[J]. SIAM Journal on Imaging Sciences, 2008: 143-168.
- [14] XU J, MA J, ZHANG D. Improved total variation minimization method for compressive sensing by intra-prediction[J]. Signal Processing, 2012(92): 2614-2623.
- [15] TAKHAR D, LASKA J, WAKIN M, et al. A new compressive imaging camera architecture using optical-domain compression[J]. Computational Imaging IV, 2006(6065), 43-52.

-
- [16] DUARTE M, DAVENPORT M, TAKHAR D. Single-pixel imaging via compressive sampling[J]. *IEEE Signal Processing Magazine*, 2008, 25 (2): 83-91.
- [17] MARCIA R, WILLETT R. Compressive coded aperture superresolution image reconstruction[C]. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008: 833-836, Las Vegas.
- [18] WILLETT R, MARCIA R, NICHOLS M. Compressed sensing for practical optical imaging systems: a tutorial[J]. *Optical Engineering*, 2011, 50 (7): 072601.
- [19] JACQUES L, VANDERGHEYNST P, BIBET A. CMOS Compressed imaging by random convolution[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, 19-24.

第 14 章 模糊图像复原技术

14.1 概 述

模糊图像复原得到全清晰/全聚焦图像在计算摄影和数字图像处理领域是一个很具有挑战性的课题。造成图像模糊的原因很多，照相机抖动、低照度、聚焦不准、光学系统的像差、拍摄物体的运动、大气湍流效应、环境随机噪声等都会导致图像模糊。另外，图像的编解码、传输过程都可能导致图像的进一步模糊。

上述因素引起的图像质量下降称为图像退化。图像退化的典型表现是图像出现模糊、失真和统计噪声等。由于图像的退化，显示输出的图像已不再是传输的原始图像，图像质量明显变差。对退化的图像进行处理，恢复出真实的原始图像的过程就称为图像复原。和计算摄影学相关的图像复原技术的核心部分是由两部分组成，一是用于图像复原的创新光学装置，二是相应的图像复原算法。由于引起图像退化的因素众多且性质各异，目前没有统一的复原方法，研究人员根据不同的应用物理环境，采用了不同的退化模型、处理技巧和估计准则，从而得到了不同的复原技术，本章将详细介绍两种典型的应用了图像复原技术的计算摄影方法：编码曝光和编码孔径。对于退化模型中的噪声而言，现在有两种看法，一种看法是把图像去噪技术视为图像复原技术的一部分，支持这种看法的研究人员把图像噪声视作图像退化的一个因素并进行处理。另一种看法是把去噪技术视为一个独立研究课题，毕竟去噪算法本身也有很多特点和相应的处理方案，本书有专门的章节介绍图像去噪技术。

需要注意的是，对于退化模型中图像模糊而言，通常说的图像去模糊技术（Image Deblurring）包含更广泛的概念范畴，包括图像增强和图像复原。虽然图像复原和图像增强的目的都是为了改善图像的质量，但是所不同的是图像复原过程实际上是一个估计过程，需要根据某些特定的图像退化模型，对退化图像进行复原。图像增强不考虑图像是如何退化的，只有通过试探各种技术来增强图像的视觉效果。而图像复原就完全不同，需知道图像退化过程的先验知识，据此找出一种相应的逆过程方法，从而得到复原的图像。图像复原主要取决于对图像退化过程的先验知识所掌握的精确程度。本章将介绍模糊图像复原技术，这并不说明图像增强技术不重要，而是图像复原技术和计算摄影贴合得更紧密。

早期的模糊图像复原方法有：非邻域滤波法、最近邻域滤波法以及效果较好的维纳滤波和最小二乘滤波等。随着数字信号处理和图像处理的发展，新的复原算法不断出现，在应用中可以根据具体情况加以选择。目前国内外图像复原技术的研究和应用主要集中于空间探索、天文观测、物质研究、遥感遥测、军事科学、生物科学、医学影像、交通监控、刑事侦查等领域。如生物方面，主要是用于生物活体细胞内部组织的三维再现和重构，通过复原荧光显微镜所采集的细胞内部逐层切片图，来重现细胞内部构成；医学

方面,如对肿瘤周围组织进行显微观察,以获取肿瘤安全边缘与癌肿原发部位之间关系的定量数据;天文方面,如采用迭代盲反卷积进行气动光学效应图像复原研究等。本章重点介绍离焦、运动、大气湍流等原因引起的图像模糊,图像复原的方法效果较好的常用算法包括维纳滤波算法、小波算法、基于训练的方法等。在知道退化模型的情况下,相对图像增强来说,模糊图像复原可以取得更好的效果。

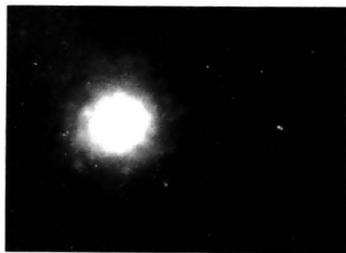
在本章 14.2 节将会介绍图像模糊的三种主要类型,在 14.3 节中将会分析图像复原的数学模型和求解方法。接下来,在 14.4 节中将重点介绍两种模糊图像复原的计算摄影技术:用于去除运动模糊的编码曝光技术和用于去除散焦模糊的编码孔径技术。在 14.5 节中将对模糊图像复原技术进行总结和展望。

14.2 图像模糊的类型

根据不同的物理性质,图像模糊大致可以分为大气模糊、散焦模糊和运动模糊三大类。这一节将分别介绍这三类图像模糊各自产生的原因和呈现的特点。

14.2.1 大气模糊

大气模糊 (Atmospheric Blur) 是模糊图像复原研究领域提的相对较少的一种模糊形式。它源于光学湍流和气溶胶造成的小角散射。光学湍流是由于大气中的某些属性因随机的时空变化引起的,如温度和密度的光折射的变化。同时,不同大小的化学物质也对不同类型的、不同波长的光波具有不同的散射性。在大气科学中,这种现象通常被称为气溶胶散射。举个简单的例子,可以想象火焰热浪后的物体,因为不同温度气体的折射,产生出了图像模糊的情况,如图 14-1 所示。在一般情况下,湍流模糊更接近地球表面和气溶胶模糊的问题,在海拔较高的地方较为普遍。光学湍流和气溶胶散射这两个主要原因导致了图像信号在空间传递中的混合,造成了图像模糊。大气模糊目前在大气科学、卫星成像和遥感成像等领域是一个重要的研究主题。



(a) 拍摄星体的大气模糊示例



(b) 气体热浪造成的图像模糊

图 14-1 大气模糊实例

14.2.2 散焦模糊

散焦模糊 (Out-of-Focus Blur) 也通常被称为光学模糊,在本书前面章节介绍光圈时,就详细介绍了由于光圈产生的散焦模糊的原因。小孔成像是不存在散焦模糊问题的。由于大光圈的使用,不在焦平面上的物体,会产生散焦模糊,并且越远离焦平面,被拍

摄的物体在图像中就会成像越模糊，如图 14-2 所示。对于图像的散焦模糊，前面章节也介绍了利用光场成像技术实现“先拍摄，后对焦”的原理，当然光场照相机也有自己本身的缺点，比如说造价高、分辨率较低等。本章将详细介绍使用编码孔径技术处理图像散焦模糊的问题。



图 14-2 散焦模糊示意

通常在处理散焦模糊图像复原时，由于不能预知拍摄场景的深度信息，一般不会将整幅图像一并处理，而是将图像划分成很多小块，如图 14-3 所示。对于同一小块内的图像，认为其深度是一致的，模糊程度是一致的。对每个图像块分别处理后，再重新拼成一幅完整的图像。这种图像的分块处理方式，在第 7 章局部色调映射算法中也曾看到。还有很多传统的图像处理算法，也会用到这种处理方式。



图 14-3 散焦模糊图像分块示意

14.2.3 运动模糊

运动模糊 (Motion Blur) 的产生是因为在一次曝光时间内，由于运动使得整个场景或者拍摄目标产生了拖影。运动的产生主要是因为两个因素，一是拍摄目标的快速移动，二是拍摄过程中照相机本身的晃动。尤其是在低光照条件下，这个问题会变得更加突出。因为光线较暗的时候，需要通过较长的曝光时间来获得高信噪比、亮度适中的图像，这就需要增加曝光的时间。然而，在长曝光情况下很难保持手持照相机的稳定。一旦照相机因为某种原因发生抖动，被拍摄的目标本身的像素会在运动的反方向出现多个像素的拖影，去除运动模糊的目标就是要将这些拖影去除。

这里首先分析照相机晃动的情况，照相机晃动造成图像模糊是整个场景都变得模糊了。所以图像复原技术的关键在于分析和估计照相机晃动的运动方式，包括运动方向、轨迹和速度等。这些运动有些是线性的，有些是非线性的，现实中大部分晃动的运动是非线性的。图 14-4 所示说明了四种常见的摄像机运动，包括平行焦面内和非平行焦面内的照相机运动。

平行焦面内的位移是最常见的照相机运动方式，它是指照相机在平行于对焦平面的平面内上下左右地晃动/抖动，如图 14-4 所示。平行焦面内的运动模糊不仅取决于摄像机的运动，而且还和拍摄场景的景深相关。尤其是用短焦镜头拍摄的照片，模糊内核的大小变化是与距照相机的距离成反比，因此，前景中的物体被模糊的程度要比背景要大。如果假设摄像机运动是垂直于光轴，特别是用长焦距对场景和目标进行拍摄的时候，所

具有视差效应可以忽略不计。也就是说，照相机的位移将不取决于场景深度的变化的影响，因而可以认为这种模糊是具有空间不变性的，这个假设被广泛采用在大多数的研究工作。除了位移的变化，平行焦面内的旋转也会导致图像模糊，如图 14-4 (a) 所示，并且离旋转中心越近，运动模糊的程度就越小。对于平行焦面内的旋转造成的运动模糊，无论是否考虑视差效应，旋转模糊是空间不变的，也就是说和场景深度无关的。

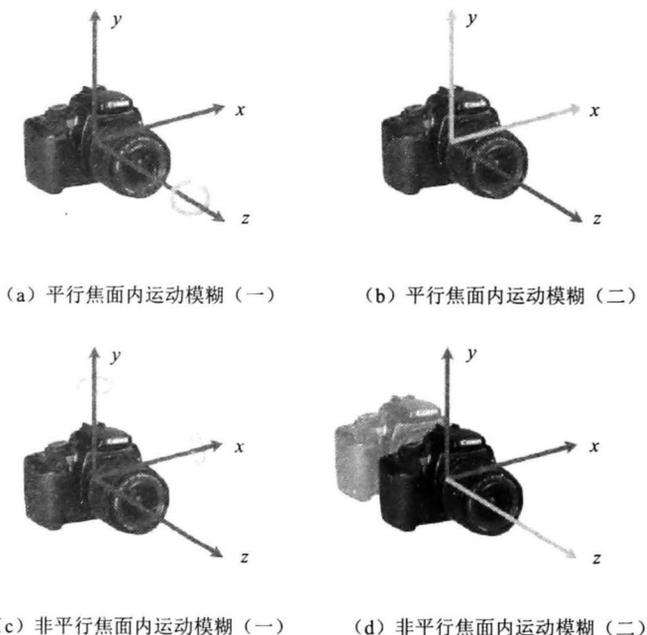


图 14-4 导致图像模糊的照相机运动方式示意

非平行焦面内的位移导致的运动模糊[见图 14-4 (d)]是指照相机在垂直焦平面的方向上的相对位移造成的模糊。所有像素的图像将被模糊径向向内或向外，这种照相机运动可以被称为径向模糊。非平行焦面内的旋转也会造成运动模糊，如图 14-4 (c) 所示，其主要原因为镜头的畸变导致倾斜角度的模糊图像。因此，非平行焦面旋转的这种情况是如图 14-4 所示的四种类型的摄像机运动中非常复杂的一种，现有的技术还不能很好地定义的这种复杂的运动，这也是未来研究的方向之一。

当照相机在一次曝光时间内保持静态，拍摄物体的移动造成的运动模糊是不同于照相机抖动造成模糊的另一类研究课题。场景中的拍摄对象可能有不同的运动方式和速度，并且会有多个目标以多种运动姿态造成成像的模糊。图 14-5 显示了被拍摄的汽车所引起的运动模糊。在实际问题中，可能同时引起运动模糊和照相机晃动模糊这两种模糊叠加的情况。这种情况下，运动模糊变得更加复杂，目前科学家的研究还是把这两种模糊情况分开进行分析和处理。



图 14-5 运动物体造成的图像模糊

14.3 图像复原的数学模型

在数学上, 图像退化模型为式 (14.1) 所示的线性关系:

$$B(x, y) = \int_{\Omega} I(x-r, y-s) K(x-r, y-s; r, s) dr ds + N(x, y) \quad (14.1)$$

式中, I 为原图像, K 为卷积核或点扩展函数, N 为附加的独立噪声信号, B 为所观察到的模糊图像, $\Omega \subset \mathbb{R}^2$ 。

式 (14.1) 的积分部分可以看做是每个像素 (x, y) 的能量在其相邻点上的传播, 而卷积核 $K(x, y, r, s)$ 则决定了在不同点上的权重。这就解释了为什么卷积核又叫做点扩展函数 (Point Spread Function)。如果点扩展函数不依赖于图像中的位置 (x, y) , 即 $K(x, y, r, s) = K(r, s)$, 那么式 (14.1) 的积分部分就变为标准卷积, 这样的点扩展函数称为空间不变点扩展函数。相反, 如果点扩展函数随 (x, y) 的变化而变化, 则称为空间变化点扩展函数。解决空间上变化的模糊的困难性促使人们采用空间不变模型, 这也将是本节所讨论的主题。对于空间不变模型, 式 (14.1) 变为如下形式:

$$B(x, y) = \int_{\Omega} I(x-r, y-s) K(r, s) dr ds + N(x, y) \quad (14.2)$$

在数字图像处理过程中, 等式 (14.3) 可以更好地表示离散模型:

$$B(x, y) = (I \otimes K)(x, y) + N(x, y) = \sum_{\Omega} I(x-r, y-s) K(r, s) + N(x, y) \quad (14.3)$$

式中, 操作符 \otimes 表示 2D 卷积。

通过 2D 离散傅里叶变换, 可以将等式 (14.3) 表示在离散频域内:

$$B_F(f_x, f_y) = I_F(f_x, f_y) \Theta K_F(f_x, f_y) + N_F(f_x, f_y) \quad (14.4)$$

式中, 操作符 Θ 表示逐个元素相乘; B_F 、 I_F 、 K_F 分别表示图像 B 、 I 和点扩展函数 K 的傅里叶变换。

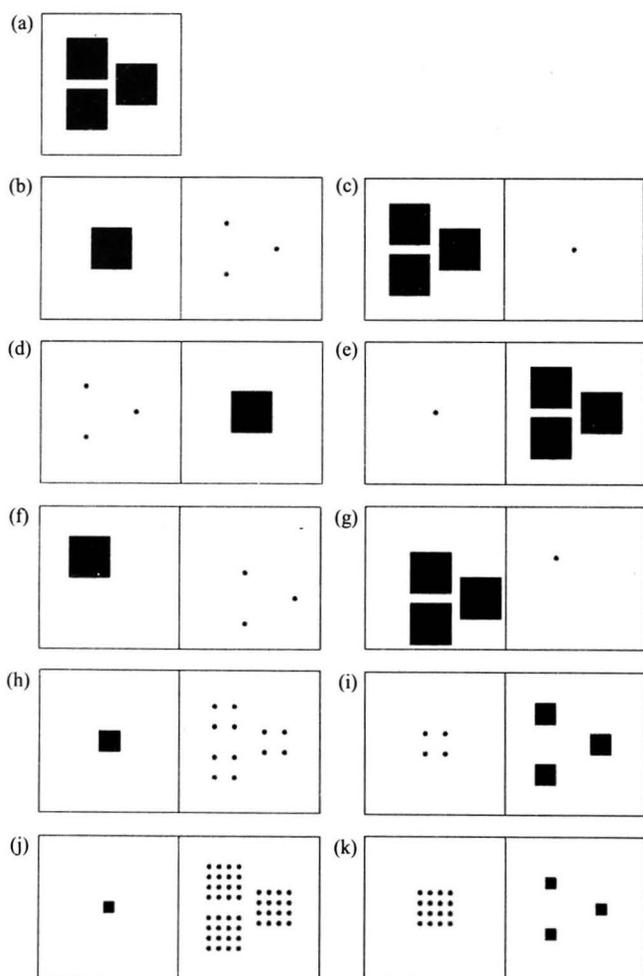
如果原图像 I 和点扩展函数 K 均不知道, 那么从模糊图像 B 中恢复 I 的过程称为盲卷积问题。盲卷积是一个众所周知的受约束问题。对于 B , 将会有无限种可能存在的估计对 (\hat{I}, \hat{K}) 。当点扩展函数 K 已知时, 那么从模糊图像 B 中复原 I 的过程可以看做是非盲卷积问题。通常, 非盲卷积算法作为闭环优化的一部分, 是盲卷积过程中十分重要的组成。在接下来的部分, 将介绍当前关于盲卷积和非盲卷积的工作。与此同时, 将介绍如何对模糊核进行估计。

14.3.1 盲卷积

首先介绍在盲卷积 (Blind Image Deconvolution) 中所面对的主要挑战。其次, 将所有的盲卷积方法分为分离方法和联合方法两类。在分离方法中, 首先要对模糊核进行估计, 然后结合模糊图像对原图像进行恢复。通常, 这种方法可以分为两步, 模糊核估计和非盲卷积, 将在 14.3.2 和 14.3.3 节中分别介绍。联合方法是对原图像和模糊核同时进行估计。然而, 大部分联合方法只是交替对模糊核和原图像进行估计, 而不是真正意义上的同时进行。

1. 重解

盲卷积是一个定问题，因为都需要对原图像和模糊核进行估计，这就导致式 (14.3) 的解有无限种可能的组合。图 14-6 表明了一些有效的解决方案。而这些解决方案可能只是从三个正方形对象的一个观察方向出发的。显然，误差度量有多个全局最优解，因此也可能有多个局部次优解。当噪声达到一定程度时，将会进一步增加求解的难度。这意味着需要额外的先验信息来确保解的形成。在图 14-6 中，图 14-6 (a) 是观察到的图像，图 14-6 (b) 是正确的解，图 14-6 (c) 是另外一种解，图 14-6 (d) 和图 14-6 (e) 分别是交换图像和模糊核，图 14-6 (f) 和图 14-6 (g) 分别是对图像和模糊核进行变换，图 14-6 (h) ~ 图 14-6 (k) 是将解分解为更小的正方形。

图 14-6 解的几种可能^[1]

2. 联合盲卷积

现有的大部分盲卷积方法基本属于联合盲卷积 (Joint Blind Deconvolution) 类。贝叶斯理论的一个基本原则就是将所有参数和显著的变量视为未知的随机量，然后依据主

观想法为其设定概率分布。因此，在盲卷积中，式(14.3)中的原图像 I 、模糊核 K 和噪声 N 都被视为随机场中的样本。最后，根据相应的先验概率密度函数来建立关于成像过程和图像性质的模型。通过贝叶斯框架，可以写出联合全局分布为

$$p(\Omega, I, K, B) = p(\Omega) p(I, K|\Omega) p(B|\Omega, I, K) \quad (14.5)$$

通常假设 I 和 K 是有条件的先验独立，即

$$p(I, K|\Omega) = p(I|\Omega) p(K|\Omega)$$

然后，采用后验概率进行推导，可得

$$p(I, K, \Omega|B) = \frac{p(B|\Omega, I, K) p(I|\Omega) p(K|\Omega) p(\Omega)}{p(B)} \quad (14.6)$$

对于联合估计方法，既可以忽略 I 作为一种先验模糊核估计方法，如式(14.7)，也可以忽略 K 来估计原图像，如式(14.8)所示。

$$p(K|B) = \iint p(I, K, \Omega|B) dI \cdot d\Omega \quad (14.7)$$

$$p(I|B) = \iint p(I, K, \Omega|B) dK \cdot d\Omega \quad (14.8)$$

在这里，将介绍两种常用的基于贝叶斯框架的联合盲卷积方法。第一种方法是找到基于泊松噪声假设的能量函数 $\Psi(I)$ 的最小值，能量函数

$$\Psi(I) = K \otimes I - B \odot \log(K \otimes I) + \log(B!) \quad (14.9)$$

假设已经得到 (I', K') ，那么更新准则为

$$\begin{cases} K^{t+1} = \frac{K'}{\sum I'} \odot \left(\frac{B}{I' \otimes K'} \cdot I' \right) \\ I^{t+1} = I' \odot \left(\frac{B}{I' \otimes K'} \cdot K' \right) \end{cases} \quad (14.10)$$

第二种方法是基于高斯噪声，采用类似 L_1 范数的边缘保持正则函数。能量函数

$$\Psi(I) = \frac{1}{2} \|K \otimes I - B\|^2 + \lambda_1 \iint \|\nabla I\| + \lambda_2 \iint \|\nabla K\| \quad (14.11)$$

如上所示，更新准则为

$$\begin{cases} I' \cdot (I' \otimes K^{t+1} - B) - \lambda_2 \operatorname{div} \left(\frac{\nabla K^{t+1}}{\|\nabla K^{t+1}\|} \right) = 0 \\ K^{t+1} \cdot (I^{t+1} \otimes K^{t+1} - B) - \lambda_1 \operatorname{div} \left(\frac{\nabla I^{t+1}}{\|\nabla I^{t+1}\|} \right) = 0 \end{cases} \quad (14.12)$$

在上述方法中，最优化都是通过交替进行的最小化来实现的。

14.3.2 模糊核的估计

根据上文所描述，只采用模糊图像的模糊核估计是一个病态问题，因为模糊核和原图像间存在多种组合。基于模糊核或原图像的先验知识能够很好地消除一些潜在的解。通常，许多关于模糊核估计的研究工作涉及三部分知识：参数模型、核约束、先验图像。

1. 参数模型

模糊核的先验分布参数直接由参数模型来定义。在以往的工作中,许多方法采用直接估计模糊核的参数,该方法简单且有效率。对应 14.2 节中所提到的三种图像模糊类型,将依次介绍其参数模型。

大气模糊: Hufnagel 在 1964 年提出一个基于大气湍流的退化模型:

$$K(x, y) = \int_{\Omega} \exp\{-\beta(f_x^2 + f_y^2)^{5/6}\} \exp\{j2\pi(f_x x + f_y y)\} df_x df_y \quad (14.13)$$

式中, x 、 y 为空间坐标; f_x 、 f_y 为频域坐标; β 为常数。

那么模糊核就可以采用下式进行估计:

$$K(x, y) \propto \left(1 + \frac{x^2 + y^2}{\sigma^2}\right)^{-\alpha} \quad (14.14)$$

散焦模糊: 在很多图像处理的应用中,一个基于几何光学的简单模型被广泛应用。该模型只考虑通过无像差凸薄透镜的孔径成像。基于几何光学和理想透镜假设,散焦模糊的模糊核可以表示为一个圆盘:

$$K(x, y) = \begin{cases} \frac{1}{\pi r^2} & \sqrt{x^2 + y^2} > r \\ 0 & \sqrt{x^2 + y^2} \leq r \end{cases} \quad (14.15)$$

式中, r 为圆盘的半径。

通常,真正的散焦模糊是图像平面上一个平滑衰减的圆形模糊点。因此,一个更广泛散焦模糊参数模型是一个圆对称的 2D 高斯函数:

$$K(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (14.16)$$

运动模糊: 对于使用参数形式的运动模糊,通常假设其模糊核是线性的且沿着一个维度,可以充分近似为平滑的快速运动:

$$\vec{K} = \frac{2}{\lambda L^2 + (2-\lambda)L} ((\lambda-L)[1, 1, \dots, 1] + \lambda[1, 2, \dots, L]) \quad (14.17)$$

在这个水平模型中,有两个十分重要的参数:运动模糊的形状 λ 和核的长度 L 。对于式 (14.17) 所示的模型,有三个代表性的模糊核。

(1) 斜坡 (Ramp): 加速运动,此时 $\lambda=1$, 有

$$\vec{K} = \frac{2}{L^2 + L} [1, 2, \dots, L] \quad (14.18)$$

(2) 梯形 (Trapezoid): 混合匀速运动和加速运动,如 $\lambda = \frac{1}{L+1}$, 有

$$\vec{K} = \frac{2}{3L^2 + L} [L+1, L+2, \dots, 2L] \quad (14.19)$$

(3) 平方加 (Square plus): 匀速运动,此时 $\lambda=0$, 有

$$\vec{K} = \frac{1}{L} [1, 1, \dots, 1] \quad (14.20)$$

2. 核约束

有几种常见的假设来约束模糊核的形式，下面将依次介绍。

非负性：模糊核的值一定要是正值或者为 0，即

$$K(x, y) \geq 0 \quad \forall x, y \in \Omega \quad (14.21)$$

没有真正的模糊核有负值，因为世界上没有负的光照。非负性是一个很强的约束条件，几乎适用于所有类型的模糊核。

能量守恒性：在去模糊过程中，模糊核保证能量守恒，并不移除光照也不损失任何能量。能量守恒约束表示为

$$\sum_{x,y} K(x, y) = 1 \quad (14.22)$$

与正约束类似，能量守恒约束也是一个适用于任何有效核函数的通用性质。它几乎可以用于所有的模糊核估计方法。

稀疏性：稀疏性先验广泛用于运动模糊核。运动核函数表明照相机或者物体的运动轨迹，通常是连续的稀薄曲线。因此，模糊核的值大部分为零，趋近于稀疏。一般情况下，稀疏性先验的核函数值服从混合指数分布为

$$p(K(x, y)) = \sum_d \alpha_d \exp\left(-\frac{K(x, y)}{\lambda_d}\right) \quad (14.23)$$

平滑性：在散焦模糊中，模糊核近似为一个对称的圆盘或者高斯函数。显而易见，在散焦模糊中，模糊核的值平滑地扩散。在运动模糊中，由于速度和加速度的原因，运动轨迹是连续的而且是二次可微的，模糊核的值也是均匀地扩散到其邻域上的。因此，假设模糊核的先验值满足高斯分布为

$$p(K(x, y)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|K(x, y)\|^2}{2\sigma^2}\right) \quad (14.24)$$

具体性：核函数的表现可以通过参数的形式来描述。例如，用圆盘来近似表示散焦模糊核的形状。显然，运动核模糊具有重要的几何性质：核函数为一条稀疏的平滑曲线。因此，样条曲线或者曲波系统被用来表示运动模糊核。因为它们能够稀疏、连续、平滑地表示出这类模糊核。

3. 先验图像

通常，先验图像是为了获得自然场景的统计数据，一般应用在机器视觉方面，如光流、复原、超分辨率、基于图像的渲染等。下面主要考虑几种通过图像统计上的先验概率而得到的模糊核。

图像梯度分布：关于自然影像统计的研究表明，尽管真实世界的场景图像颜色分布差异很大，但是它们的梯度仍然服从重尾分布，即梯度分布大部分值都很小，但是相比高斯分布，它赋予较大的、值更高的概率。也就是说，图像包括两部分：一部分为恒定强度的平滑区域，这部分梯度较小或者为 0；另一部分为尖锐的结构，有较大的梯度值，如边缘或闭合的边界。

图 14-7 展示了自然图像的对数梯度分布直方图。对于一张给定的图像，运动模糊

的方向可以视为最小的导数变化方向。此外，导数分布的形状还用来确定模糊核的大小。这种简单的内核逼近仅限于水平或垂直运动和一维箱式滤波器模糊。最后，对图像进行分层，保证每层都有相同的模糊核。

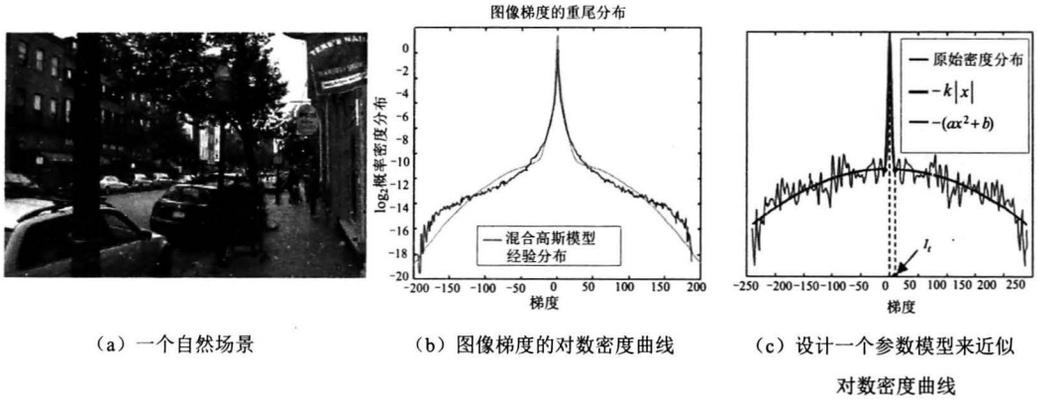


图 14-7 采用图像梯度先验的核估计^[2]

Alpha 通道约束 (Alpha-Matte Constraint): 假如一个固态的、不透明的物体在拍摄过程中由于运动而变得模糊了，那么它的边界就会融合到背景中，导致它变得透明。对于这种模糊的对象，在对象中心的像素有两种 Alpha 值，即 0 或 1。而对于原始为模糊的对象，每个像素不是属于背景就是属于前景，这表明可以通过只有 0 和 1 的二值图来近似表示发生模糊的透明图像。采用透明度约束的运动模糊核估计有一个突出的优点：二值图不仅能够大大减少图像重建的模糊度，而且使得有可能对模糊核估计采用高效离散的优化方法。

图 14-8 所示为采用 Alpha 通道的一个实例。尽管相比梯度先验，Alpha 通道能够很好地降低未知值的自由度，但是它在模糊图像中的准确性是有限的。

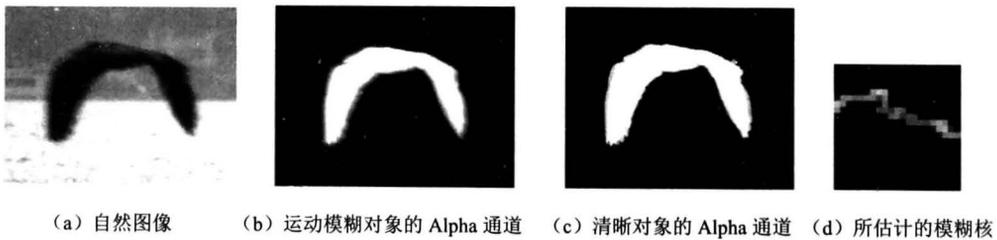


图 14-8 采用 Alpha 通道约束的核估计^[3]

边缘轮廓 (Edge Profile): 对于某些类型的模糊，图像边缘的位置也有可能检测到。造成模糊的尖锐边缘能够被预测到的前提是在模糊前，这个边缘是一个阶跃边缘。每对预测和模糊的边缘能够带来关于模糊核径向剖面的先验知识。如果一张图像包含横跨所有方向的边缘，那么模糊图像和所预测的清晰图像就包含足够多的信息来解决一般的二维模糊核问题。

在图 14-9 中，首先采用高斯边缘检测器来获得模糊图像中边缘的位置和方向，然后沿着边缘轮廓找到局部像素的最大值和最小值。通过从一条边缘的两侧像素传播最大和最小值到子像素边缘位置，来实现对理想尖锐边缘的预测。虽然基于图像先验的方法

能够很好地估计模糊核，甚至复杂的运动模糊核，但是这类方法的主要缺点是对于一般的图像，假设先验概率并不总是成立的。

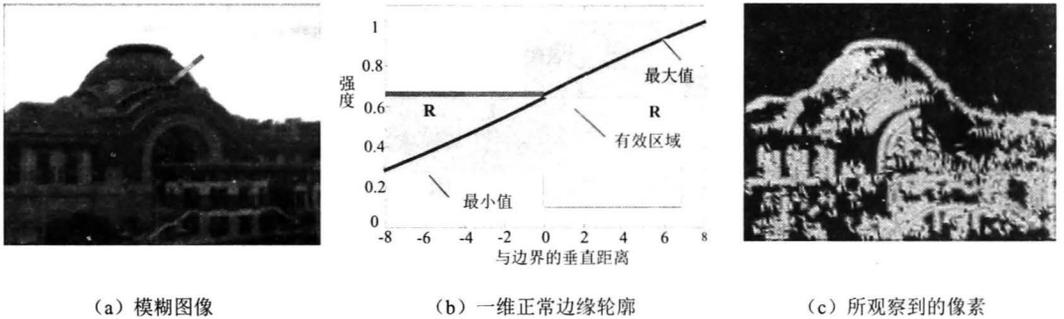


图 14-9 采用边缘轮廓的边缘预测^[4]

14.3.3 非盲卷积

非盲卷积 (Non-blind Image Deconvolution) 方法已经研究了很多年，但是，实际上不管模糊核多么完美，总会有意想不到的情况出现在重建图像里。贝叶斯公式可能只是一种能够描述经典最小化问题的方法。贝叶斯方法是为了建立条件概率密度关系，有

$$p(I|B, K) = \frac{p(B|I, K) p(I) p(K)}{p(B) p(K)} \propto p(B|I, K) p(I) \quad (14.25)$$

贝叶斯解就是在模糊图像 B 和模糊核 K 已知的前提下，求得原图像 I 。如果假设原始图像有相同的最大化，那么 $p(I)$ 就是一个常数，则最大似然解 (ML) 就为概率密度 $p(B|I, K)$ 的最大值：

$$ML(I) = \max_I p(B|I, K) p(I) \quad (14.26)$$

如果考虑到原图像 I 的先验概率，那么最大后验概率解 (MAP) 为

$$MAP(I) = \max_I p(B|I, K) p(I) \quad (14.27)$$

当假设 $p(I)$ 是均匀的概率密度函数时，最大似然解等同于最大后验概率解。下面，将会分别介绍不同形式的似然性和先验。

1. 带有高斯噪声的最大似然

如果式 (14.3) 中的噪声为高斯噪声，那么似然概率函数可以表示为

$$p(B|I, K) = \frac{1}{\sqrt{2\pi}\sigma_I} \exp\left(-\frac{\|B - I \otimes K\|^2}{2\sigma_I^2}\right) \quad (14.28)$$

假设 $p(I)$ 是一个常数，那么 $p(B|I, K)$ 最大等同于下式最小：

$$\Psi(I) = \left(\frac{\|B - I \otimes K\|^2}{2\sigma_I^2}\right) \quad (14.29)$$

上式的解可以分为两类：非迭代方法和迭代方法。

非迭代方法的优势是容易实现和快速计算。但是，对于噪声放大很难控制。傅里叶变换提供了一种直接快速的图像卷积方法。如果噪声可以忽略不计，通过采用离散傅里叶反卷积就可以确定图像，即

$$I_F(f_x, f_y) = \frac{K_F^*(f_x, f_y) B_F(f_x, f_y)}{\|K_F(f_x, f_y)\|^2} \quad (14.30)$$

迭代方法通常采用对数似然函数作为评价函数。在这里，将会介绍几种迭代方法来满足对数似然函数，也就是统计学中通常所说的期望最大化方法。

Van Cittert 的方法是最早的、最简单的图像重建方法，它将数据和图像定义在同一网格上。该方法从零阶图像 $I^{(0)} \equiv 0$ 开始遍历所有的网格点，按照下式依次迭代：

$$I^{t+1} = I^t + \alpha(B - K \otimes I^t) = \alpha B + (1 - \alpha K) \otimes I^t \quad (14.31)$$

采用 Van Cittert 方法的关键在于选择好参数 α 的值和建立相应的停止标准，这样计算时间、噪声放大、复原图像的分辨率是可以接受的。

Landweber 方法是另外一种迭代方法：

$$I^{t+1} = I^t + \alpha K \cdot (B - K \otimes I^t) \quad 0 < \alpha < \frac{2}{|A_{\max}|} \quad (14.32)$$

式中， A_{\max} 为 $\bar{K}^T \bar{K}$ 的最大特征值， α 为一个较小的正参数。

另外一种迭代方法是共轭梯度法，该方法首先通过初始图像 I^0 计算对数似然函数的负梯度 r^0 ，并设置初始共轭梯度方向为 $q^0 = r^0$ ，然后根据下式来建立一系列的负梯度 r^t 和共轭梯度方向 q^t ：

$$\begin{aligned} r^0 &= B - K \otimes I \\ q^0 &= r^0 \\ \alpha^t &= \arg \min (I^t + \alpha^t q^t) = (r^t)^T q^t / (q^t)^T \bar{K} q^t \\ I^{t+1} &= I^t + \alpha^t q^t \\ r^{t+1} &= r^t - \alpha^t \bar{K} q^t \\ \beta^t &= (r^{t+1})^T r^{t+1} / (r^t)^T r^t \\ q^{t+1} &= r^{t+1} + \beta^t q^t \end{aligned} \quad (14.33)$$

2. 带有泊松噪声的最大似然

如果式 (14.3) 中的噪声为泊松噪声，那么似然概率函数可以表示为

$$p(B|I, K) = \prod \frac{(I \otimes K)^B \exp(-I \otimes K)}{B!} \quad (14.34)$$

通过计算对数的导数可以得到 $p(B|I, K)$ 的最大值：

$$\frac{\partial \log p(B|I, K)}{\partial I} = 0 \quad (14.35)$$

由于 $-\log p(I|B, K) = \sum [-I \log(K \otimes I) + K \otimes I]$ ，将会使上式变为

$$\frac{B}{I \otimes K} \cdot K = 1 \tag{14.36}$$

在等式的两边同时乘以图像 I ，并采用基于梯度的迭代算法来估计 I^{t+1} ：

$$I^{t+1} = I^t \left[K \cdot \frac{B}{K \otimes I^t} \right] \tag{14.37}$$

上述方法就是著名的 Richardson-Lucy 算法，也是通常所说的期望最大化算法。该方法的一个重要特性是如果第一个估计是非负的，那么整个结果都是非负的。

3. 采用不同先验的最大后验

对于上述两种主要似然，可以添加先验或者全局约束到不适定问题的解中。一种常用的方法就是为对数似然函数增加一个正则化项 $\Phi(I)$ 来改变能量函数：

$$\tilde{\Psi}(I) = \Psi(I) + \lambda \Phi(I) \tag{14.38}$$

λ 为控制正则化相对强度的归一化参数。实际上，正则化项 $\Phi(I)$ 是先验概率 $p(I)$ 的负对数，二者并没有本质上的区别。

最著名的正则化项方法是 Tikhonov-Miller (TM) 算法，该算法按下式设计正则化项：

$$\Phi_{\text{TM}}(I) = \|\nabla I\|^2 = \int \left(\frac{\partial I(x, y)}{\partial x} \right)^2 + \left(\frac{\partial I(x, y)}{\partial y} \right)^2 dx dy \tag{14.39}$$

该方法能够通过傅里叶变换快速实现，然而，由于先验概率采用 L_2 范数，并不能很好地保护边缘部分。

Total Variation (TV) 方法采用 L_1 范数，在保持边缘的同时能够很好地平滑均匀区域，表示式为

$$\Phi_{\text{TV}}(I) = \|\nabla I\| = \int \sqrt{\frac{\partial I(x, y)^2}{\partial x} + \frac{\partial I(x, y)^2}{\partial y}} dx dy \tag{14.40}$$

TV 方法能够很好地适用于非纹理对象，对于纹理图像的处理效果并不是很好。最大熵方法是对不适定问题增加约束的尝试，最常用的熵函数如下所示：

$$\Phi_{\text{entropy}}(I) = -\sum_x \sum_y I(x, y) \log(I(x, y)) \tag{14.41}$$

当正则化项 $\Phi_{\text{entropy}}(I)$ 的值较大时，会造成该方法的不稳定。

最近，采用稀疏、自然图像的先验成为图像重建的一项重要工作。该方法能够提供更加清晰的重建图像：

$$\Phi_{\text{parse}}(I) = \|\nabla I\|^{0.8} = \int \left(\left| \frac{\partial I}{\partial x} \right|^{0.8} + \left| \frac{\partial I}{\partial y} \right|^{0.8} \right) dx dy \tag{14.42}$$

式中， $|\cdot|^{0.8}$ 是一个重尾函数。该方法能够提供更加清晰的边缘，移除噪声和一些不希望出现的伪像。该方法的不足之处在于优化问题不再是一个简单的最小二乘法，不能在闭环中取得最小值。

图 14-10 展示了稀疏先验与经典 Richardson-Lucy 方法之间的差异。可以发现，稀疏先验能够得到一个更加清晰的结果。

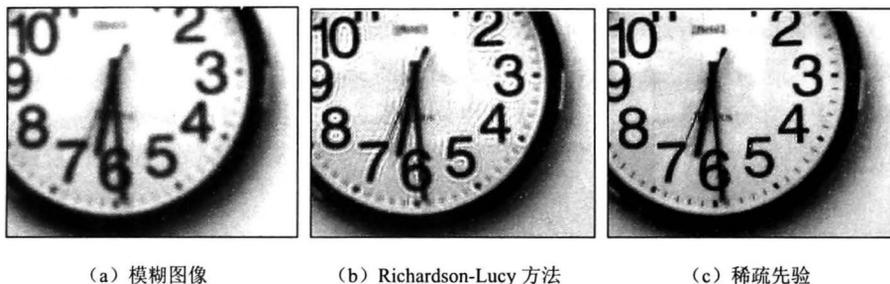


图 14-10 采用不同方法的对比图

14.4 计算摄影中的技术应用

图像运动模糊复原问题是图像处理以及计算机视觉领域中的一个基础性问题，也是一个公认的难题。根据前面介绍的数学模型，知道模糊图像的复原包括精确的点扩展函数（PSF）估计和图像反卷积两大方面的问题。对于运动模糊图像的 PSF 问题，需要有运动模糊模型和散焦模糊模型进行分析和估计。对于图像反卷积问题的病态性也是困扰运动模糊图像复原的关键因素。由于病态问题的解不连续地依赖于观测值，所以微小的偏差就会导致复原图像中存在振铃效应、寄生波纹等不良问题^[15]。本节针对去除运动模糊和散焦模糊，介绍编码曝光和编码孔径两种用于模糊图像复原的计算摄影技术。

14.4.1 编码曝光技术

分析运动模糊图像产生的原因，由于传统照相机快门在拍摄过程中始终处于开启状态，这相当于一个时间轴上的低通滤波器，因而传统照相机所获得的图像损失了很多高频信息。在图像的复原过程中，由于高频信息的损失，卷积滤波器含有多个频域零点，这导致滤波器不可逆，图像复原成为一个病态问题。近几年，一些采用计算摄影技术进行运动模糊图像恢复的方法被提出并取得较理想的复原效果。采用计算摄影技术的复原方法与以往仅仅通过单幅运动模糊图像复原方法不同，新的复原方法通常会引入一些辅助传感器或改变传统图像的获取方式来改进传统复原方法的不足^[16]。

针对传统照相机拍摄的运动模糊图像在复原时的病态性问题，受天文摄影中采用“改进一致冗余阵列（Modified Uniformly Redundant Arrays）”改进拍摄图像的信噪比的方法的启发，Raskar^[5]等提出了一种在照相机曝光时间段内快速开关快门的编码曝光（Coded Exposure, CE）方法，该方法能够改变运动模糊图像的形成过程，将反问题的病态性转化为良态，从而能够快速地使用反卷积方法获得清晰图像。

那么，什么是编码曝光方法呢？先分析一下普通曝光的过程。对于传统照相机的一次成像过程，照相机的快门从曝光开始时刻到曝光结束时刻一直是处于开启状态。也就是说，快门的一“开”一“关”这样一组动作就可以形成一个影像，如图 14-11 所示，拍摄一张图像的曝光时间为 T ，等价于在时域内与一个盒状滤波器（Box Filter）的卷积。若曝光时间段内照相机和场景之间不存在运动，则每个场景点的能量都积分到一个对应的传感器成像单元中形成清晰的图像。但如果在曝光时间段内，照相机与场景之间产生了相对

运动，那么场景点将沿着运动路径分散到若干个传感器成像单元中，形成模糊图像。

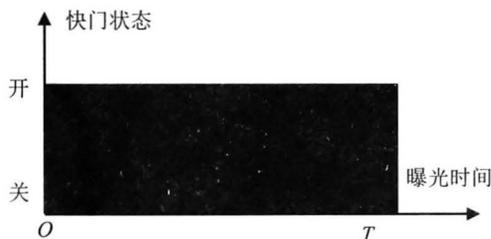


图 14-11 传统照相机曝光过程

以水平匀速直线运动为例，场景点的能量沿着运动方向均匀地分布到运动路径上的每个传感器成像单元中，形成运动模糊图像。这一过程写为点扩展函数的形式为

$$h(x) = \begin{cases} 0 & d < x < \infty \\ \frac{1}{d} & 1 \leq x \leq d \end{cases} \quad (14.43)$$

根据前面提到的图像模糊模型，对水平匀速直线运动模糊图像的一个清晰估计值可以由点扩展函数的离散傅里叶变换除法获得：

$$\hat{F}(x, y) = G(x, y)V(x, y) \quad (14.44)$$

式中， F 表示估计图像的傅里叶变换， V 表示反卷积滤波器 v 的傅里叶变换。

$$V(x, y) = \frac{1}{H(x, y)} = \frac{H^*(x, y)}{|H(x, y)|^2} \quad (14.45)$$

下面以一个简单的匀速直线运动模糊为例，论述图像运动模糊的病态性。以图 14-11 所示的目标为例，设该目标曝光期间在图像中的运动距离为 $d=16$ ，形成直线运动模糊图像，图像尺寸为 252。根据式 (14.43) 可知，该运动模糊图像的点扩展函数可以写为

$$h(x) = \begin{cases} 0 & 16 < x < 252 \\ \frac{1}{16} \times [1111111111111111] & 1 \leq x \leq 16 \end{cases} \quad (14.46)$$

根据运动模糊图像的形成原理，用上式中点扩展函数与图 14-12 (a) 中的图像进行卷积运算形成直线运动模糊仿真图像，得到如图 14-12 (a) 所示图像。如图 14-12 (b) 中，字母“ A ”中的像素点被均匀地分散到运动方向上的 16 个像素中。

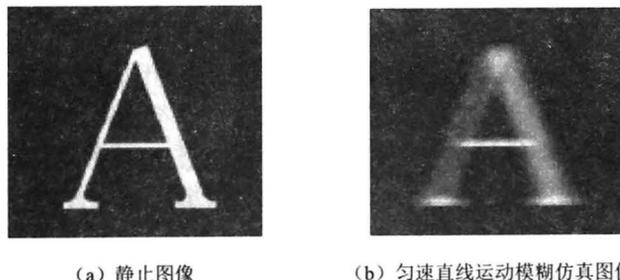


图 14-12 运动模糊仿真图像

图 14-13 所示是式 (14.46) 中点扩展函数经离散傅里叶变换后得到的幅值曲线。从

图中可以看出, 曲线在很多频率处的幅值为“0”, 即对应于式 (14.45) 中处于分母位置的数值为 0。这说明匀速直线运动模糊图像的复原过程是一个病态问题, 由于幅值零点的存在, 获得结果稳定性变差, 若考虑到噪声的存在, 某些频率内的噪声值将被无限放大, 这对估计图像来说是非常严重的影响。

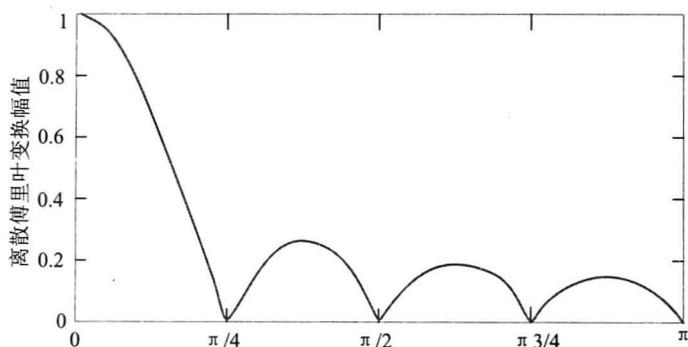


图 14-13 传统照相机匀速直线运动点扩展函数的离散傅里叶变换幅值曲线

对传统的图像复原来讲, 消除频域幅值零点是不可能的, 这也是图像复原成为著名难题的重要原因之一。为了解决频域幅值零点问题, 一些算法采用最大似然估计方法来近似求解, 如著名的 R-L 算法就是用一个图像统计模型以及贝叶斯框架求得最优估计的图像恢复算法。此类算法需要进行非线性迭代求解, 时间消耗非常巨大, 通常用于一些事后处理的场合, 对于一些要求实时性较高的应用, 则需要寻找更直接的算法来快速复原运动模糊图像。计算摄影则从照相机的工作原理角度来想办法解决点扩展函数离散傅里叶变换幅值存在“0”点这一难题。为了消除传统匀速直线运动模糊图像的点扩展函数频域幅值零点, Rasker 改变传统照相机的曝光过程, 通过快门的开/关状态将一次曝光过程拆分为多次间断的曝光过程。以 4 次曝光为例, 如图 14-14 所示, 照相机在时刻 t_0 打开快门开始第一次时间为 5ms 的曝光, 在时刻 t_1 关闭快门完成一次曝光; 在时刻 t_2 再次打开快门开始时间为 10ms 的曝光, 在时刻 t_3 关闭快门完成第二次曝光, 依此类推直到完成所有四次曝光, 每段曝光时间长度均为 5ms 的整数倍。由于单次曝光的单位时长非常短 (ms 量级), 在整个曝光过程中快门处于快速开关动作, 因此这一过程也称“闪动快门”曝光。

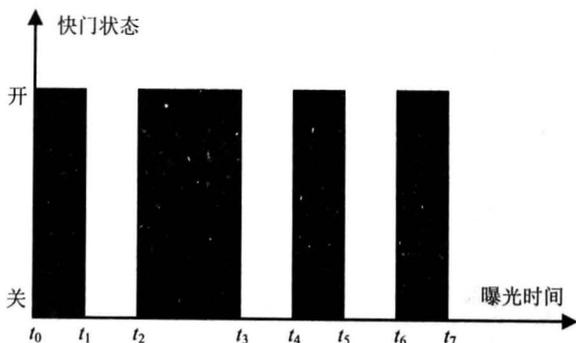


图 14-14 快门完成四次曝光动作组合示意图

将整个曝光过程的总时间分为若干个时间片，快门在每个时间片内的状态可以用“1”或“0”表示，“1”表示打开，“0”表示关闭，这样照相机快门在整个曝光时间段内的状态就可以用一个二进制的序列来表示。其快门状态可以用下面的二进制序列来表示 T ：

$$10110101 \tag{14.47}$$

这种用二进制序列来控制快门开关状态的曝光过程，称之为“编码曝光”；对应的二进制序列，称之为“码字”。照相机在编码曝光过程中，可以认为二者之间的运动是匀速直线运动，因此在子时间段内的模糊所对应的点扩散函数也符合匀速直线运动模型。假设整个曝光时间为 20 个单位时间，运动模糊的总长度为 20 个像素，即一个单位时间内的运动模糊长度为 1 个像素，在这样的条件约束下，其对应的点扩散函数与上述二进制序列（或码字）就有着密切的联系，即点扩展函数可以表示为

$$\frac{1}{1+2+1+1} \times [1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1] \tag{14.48}$$

如图 14-14 所示，照相机快门在一次曝光时间内完成了四次开关动作组合，但同时传感器却一直处于工作状态，这就使四次曝光量累加到一起形成一幅断续运动模糊图像。图 14-15 所示为上述分析得到的点扩展函数与清晰图像卷积操作得到模糊图像的仿真结果，从仿真结果中可以看出，由于在图像形成过程中快门由编码控制开关造成字母“A”的模糊形式是不均匀的，而是类似于重影一样的模糊。

为了说明编码曝光思想的优势，对编码曝光运动模糊图像的点扩展函数进行了频域分析，为了与图 14-13 中频域分析对比，将其扩展为同样尺寸，经离散傅里叶变换后计算其幅值，结果如图 14-16 所示。

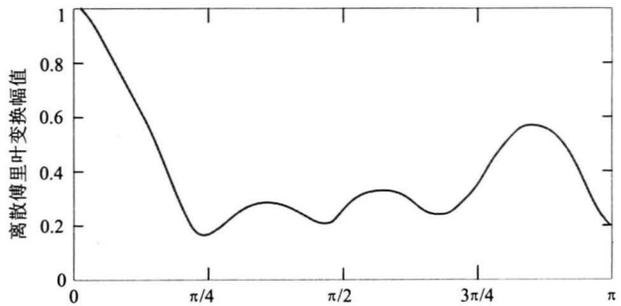
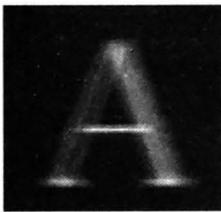


图 14-15 编码曝光仿真结果图 图 14-16 编码曝光运动模糊图像点扩展函数离散傅里叶变换幅值曲线

从图 14-16 中可以看出，编码曝光运动模糊图像点扩展函数的离散傅里叶变换没有幅值“0”点，因而在图像复原时可以进行反卷积操作，直接获得复原图像。编码曝光实际上是将传统曝光的盒状滤波器改造成一种没有频域“0”点的宽带滤波器，从而保护了图像中的频域信息，将原有病态问题转化为良态问题。

之前仅仅给出了编码曝光模型的一种特例，为不失一般性，这里可将编码曝光模型描述如下：对于给定的一段有限的曝光时间，将其均匀分割成子时间段（为正整数）。使用“1”或“0”来表示每个子时间段内照相机快门的的状态，“1”表示打开，“0”表示关闭，则照相机在整个曝光时间内的开关状态所对应的码字（即二进制序列）为

$$C_w = c_1 c_2 \cdots c_i \cdots c_m \quad 1 \leq i \leq m \tag{14.49}$$

由上述介绍的内容可以知道：编码曝光技术作为计算摄影的具体应用，其核心思想

是在照相机曝光期间根据预先设计的二进制编码序列（称为码字）快速的开/关照相机快门以保留高频信息。与在曝光期间快门一直处于打开状态的传统照相机不同，编码曝光照相机的点扩展函数是宽带滤波器，其傅里叶变换频谱不含零点，编码曝光图像的运动模糊复原可以采用快速的直接反卷积方法来获取清晰图像，其点扩展函数是可逆的。由此，编码曝光技术将病态的模糊图像复原问题转化为良态问题。

寻找控制相机快门的最优码字是编码曝光技术的关键问题。研究者们已经做了许多卓有成效的工作。Raskar 等^[6]采用随机线性搜索的方法得到一个近似最优的码字，并且首次提出了选取编码曝光相机最优码字的两条标准。Agrawal 和 Xu 同时考虑点扩展函数的可逆性和易估计性提出了一种寻找最优码字的方法^[7]。Agrawal 等^[8]随后研究了一种针对运动去模糊的图像最优获取策略问题，从实验的角度探讨了编码曝光码字构造对图像恢复结果的影响。McCloskey 证明了最优码字依赖于被拍摄物体的运动速度，并提出了一种速度依赖的最优码字搜索方法^[9]，同时探讨了编码曝光照相机的读出噪声（Readout Noise）和激发脉冲次数的关系，提出在编码曝光相机获取图像过程中应使用尽量少的激发脉冲次数以减少读出噪声。当然，还有许多其他研究人员提出自己的码字排列，只要有合理的证明，读者也可以提出自己的码字设计方案来做运动模糊图像复原。

14.4.2 编码孔径技术

编码孔径技术（Coded Aperture）最初应用在高能天文学领域，近年来受到计算机视觉领域研究人员的重视，在场景的深度信息提取和超分辨率重建等多个分支都可以看到其相关研究。利用编码孔径技术做散焦模糊图像复原是最受关注的研究热点之一。麻省理工学院和哥伦比亚大学作为编码孔径技术领域的佼佼者积极推动了编码孔径技术在普通图像复原领域的发展。编码孔径技术是利用特殊设计的掩膜来改进成像装置（见图 14-17），对还没有到达图像传感器的光束进行了有针对性的“编码”加工，为图像采集后的处理过程提供更多有效的信息。迄今为止，针对不同的应用，国内外科学家已经设计出了近 20 种不同的编码孔径掩膜，其中一些如图 14-17（a）所示，从左至右分别为：随机方案、MURA^[10]、Zhou^[11]、Levin^[12]、Veeraraghavan^[13]。图 14-17（b）显示的是采用 Levin^[12]的编码孔径设计方案对普通孔径镜头进行简单改造，可得到编码孔径镜头。



(a)



(b)

图 14-17 编码孔径的不同方案和实现途径^[14]

比起普通孔径的成像系统，应用了编码孔径技术的成像系统优势在于：经过编码孔径获得的图像能够更准确地辨识出各个区域的模糊尺度（辨识 PSF 准确度），进而更准确地利用点扩散函数的尺度和反卷积算法进行图像复原。但是编码孔径技术有其自身的问题，最主要的是，编码孔径相机采集的单幅编码图像由于掩码遮挡损失了特定频率的信息，降低了图像质量。

为什么在镜头前加一个掩膜，就可以显著提高对散焦图像模糊尺度的辨识呢？先从造成散焦模糊的原因说起。当成像平面没有落在焦平面上时，就出现了散焦模糊。此时，人们看到的图像，通俗地说就是虚了。从数学模型的角度来看，散焦模糊可以表示为一个卷积过程：

$$F(u, v) = G(u, v) \otimes H(u, v) \tag{14.50}$$

式中， $F(u, v)$ 表示获取的散焦模糊图像； $G(u, v)$ 表示清晰图像； $H(u, v)$ 表示模糊核，也就是上面说到的模糊尺度。图像去散焦模糊的数学过程，就是已知获取的模糊图像 $F(u, v)$ ，通过估计模糊核 $H(u, v)$ ，同时设计反卷积算法，最终求得清晰图像 $G(u, v)$ 的过程。估计模糊核 $H(u, v)$ ，也就是模糊尺度是一个有挑战的问题。近几年，采用计算摄影的方法来估计散焦模糊图像的模糊尺度成为研究热点，一些研究人员尝试将编码孔径引入到模糊尺度的估计中，取得不错效果。

在确定模糊尺度之前，需要对成像系统进行校准。校准过程就是获取一套该成像系统下的卷积核尺度标准，如图 14-18 所示。得到了卷积核尺度标准之后，就可以对散焦图像的模糊尺度进行估计了。

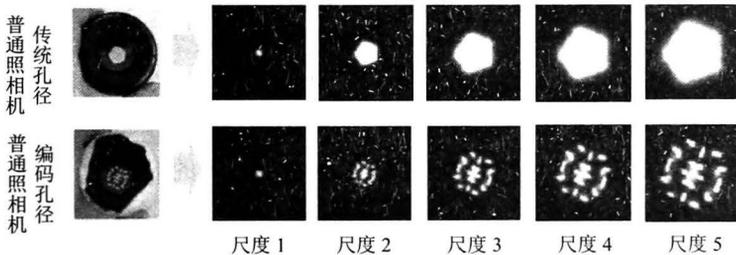


图 14-18 普通孔径与编码孔径卷积核尺度标准^[12]

为什么要引入编码孔径呢？普通的孔径就不能很好地估计模糊尺度吗？首先来看图 14-19。

从图 14-19 中可以明显看出，对于普通孔径，正确的模糊尺度以及比正确模糊尺度更小的尺度都能得到令人满意的结果，这就给模糊尺度的估计带来了困难与挑战。相比普通孔径，经过编码的孔径对于错误尺度和正确尺度的区分能力要强很多，从图 14-19 中可以看出，错误的模糊尺度得到的图像均不理想，只有正确的尺度能得到理想的结果。为什么简单地在镜头上加个掩膜，就可以提高对模糊尺度的区分程度呢？

解释这个问题之前，先要简单介绍如何估计散焦图像的模糊尺度。在建立了卷积核尺度标准之后，通过图像的概率数学模型来选择最有可能的卷积尺度，也就是模糊程度。这里为了推导方便，假设自然图像 X 的梯度值符合高斯分布，则清晰图像的概率模型可以表示为

$$P(X) \propto N(0, \sigma) \quad (14.51)$$

即

$$P(X) \propto e^{-\frac{1}{2}X^T \sigma^{-1} X} \quad (14.52)$$

其中

$$\sigma^{-1} = \lambda \text{diag} \left(|G_x(v, w)|^2 + |G_y(v, w)|^2 \right) \quad (14.53)$$

式中, $G_x(v, w)$ 、 $G_y(v, w)$ 表示在 X 、 Y 方向上的梯度; v 、 w 是频域坐标。经过简单推导可知, 其模糊图像 Y 也符合高斯分布:

$$P_s(Y) \propto N(0, \Sigma_s) \quad (14.54)$$

通过散焦数学模型可知:

$$\Sigma_s = C_{F_s} \sigma C_{F_s}^T + \eta^2 I \quad (14.55)$$

式中, C_{F_s} 是在编码孔径 F 形式下 s 尺度的卷积核。那么, 在尺度 s 下得到模糊图像的概率可以表示为

$$P_s(Y) \propto \exp \left(-\frac{1}{2} \sum_{v, w} |Y(v, w)|^2 / \text{diag} \Sigma(v, w) \right) \quad (14.56)$$

其中

$$\text{diag} \Sigma(v, w) = |F^s(v, w)|^2 \sigma^{-1} + \eta^2 \quad (14.57)$$

这里将之前制定的卷积核尺度标准的每一个候选的 F_s 代入到式 (14.56) 中, 就可以求得何种尺度下, 能使采集到的全向模糊图像概率最大。这样就可以估算出最可能的模糊程度 s 。

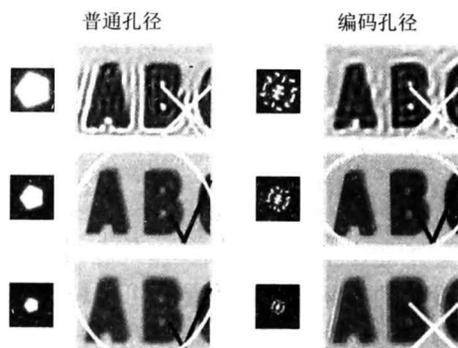


图 14-19 普通孔径和编码孔径不同模糊尺度下反卷积结果^[12]

回到先前提出的问题, 为什么简单地在镜头上加个掩膜, 就可以提高对模糊尺度的区分程度? 首先看看不同模糊尺度信号在空域和频域中特点, 如图 14-20 所示。

可以看出不同模糊尺度信号在频域下有“过零点”, 即频率值为零的点。这些“过零点”在区分模糊尺度上起着重要作用。当频率值为零时, 式 (14.56) 中的值将较大, 根据这个特性, 来确定模糊尺度。那么有人会问不同尺度之间的“过零点”相同怎么办? 这就是编码孔径的特点和优势所在。如图 14-21 所示, 可以看出普通孔径下, 不同模糊尺度信号的“过零点”重合程度高, 导致区分不同尺度较困难。相反, 编码孔径下, 不同模糊尺度信号的“过零点”相互错开, 区分度明显, 可以很好区分开不同模糊尺度信号。

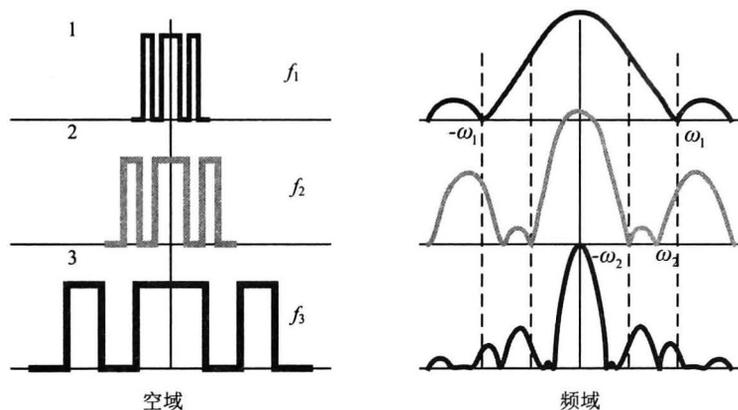


图 14-20 不同模糊尺度信号在空域和频域中示意^[12]

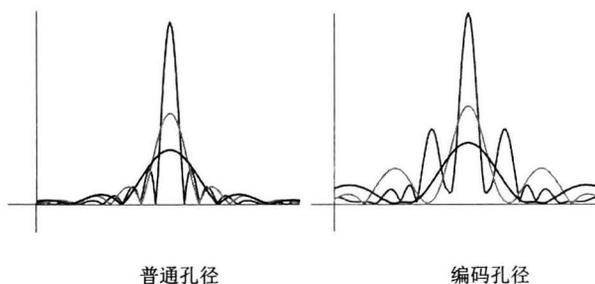


图 14-21 在普通孔径和编码孔径下，不同模糊尺度的频域“过零点”示意^[12]

编码孔径虽然增强了模糊尺度的区分程度，但是通过对光线遮挡“编码”实现的。也就是说以损失成像过程中光谱信息为代价，频域中表现出的形式就是“过零点”的增多。图 14-21 可以很好地说明编码孔径与普通孔径的优缺点。

普通孔径成像可以获得更全面的光谱信息，但是尺度区分能力较弱；编码孔径提高了模糊尺度区分能力，但是损失了部分光谱信息，对图像复原有着一定程度的影响。研究人员通过设计优秀的编码孔径形式以及采用图像先验信息等方法来弥补这个缺点，编码孔径形式设计将在下面介绍。

既然编码孔径可以很好估计出散焦图像的模糊尺度，那么编码孔径的形式设计就至关重要了。编码孔径形式设计的好坏，决定了是否可以精准地找到图像的模糊程度，也在很大程度上决定了图像复原的优劣。当前，编码孔径技术主要分为两类：单个编码孔径技术和多个编码孔径技术。下面分别对这两类进行简单介绍。

1) 单个编码孔径技术

单个编码孔径技术有其自身优势。

(1) 其对硬件的改造较为简单和便捷，只需要在透镜内侧植入一个编码的遮光片，这个遮光片可以由不透光的塑料片或者硬纸片做成。

(2) 单个编码孔径在设计阶段，具有相比于多个编码孔径更低的计算代价。

(3) 单个编码孔径成像系统中，只需要采集一幅图像，对实验场景的变化要求较为宽松。

接下来简单介绍单个编码孔径的设计和遴选过程。

对于单个编码孔径设计方案, Levin^[11]采用 KL (Kullback Leibler) 收敛法来遴选编码孔径设计。KL 收敛法计算出的 D_{KL} 表明两个不同尺度分布的差别, 值越大, 表示两个分布的差别越大。通过计算这个值来选择编码孔径。对于每一个编码设计形式, 假定场景中的模糊尺度(卷积尺度)是 10 个, 结合 KL 收敛的标准, 来对这 10 个模糊尺度信息两两求得到 D_{KL} 。平均值最大的形式成为编码孔径设计方案。针对搜索量过大的问题, 采用梯度递进算法。首先, 利用遗传算法 (Genetic Algorithm) 来搜索低精度的编码孔径设计方案, 找到最佳结果后, 将这个编码形式作为高精度搜索的起始搜索方案。依此类推, 通过这种梯度搜索算法得到需要的编码精度。

但是单个编码孔径技术有其局限性。

(1) 不能分辨图像模糊区域的尖锐边缘和清晰区域的平滑纹理。

(2) 编码孔径会造成采集图像有“过零点信息”, 实际就是模糊图像的信息缺失, 无法很好地得到散焦图像复原效果。

2) 编码孔径对技术

针对单个编码孔径技术的不足, 一些研究人员提出通过两个或多个不同设计的孔径, 对相同场景采集多张图像信息, 使获得的图像信息可以互补, 然后再一起做后期的分析处理, 进行散焦模糊图像复原。当然, 多个编码孔径的设计也有其自身的缺点: 拍摄图片间场景的变化对图像处理有很大影响。

对于两个互补的编码孔径形式的设计, 可以通过建立相关概率模型设计信息互补的卷积核获得最优的编码孔径组合。可以参考 Zhou^[11]的做法, 这里就不具体介绍了。

对于多个编码孔径的组合, 已经有一些研究人员针对普通成像设备提出了可编程的 LCD 或者 DMD 做编码遮光片, 如图 14-22 所示, 这样就可以多个编码孔径间快速切换得到多张不同编码的图像。

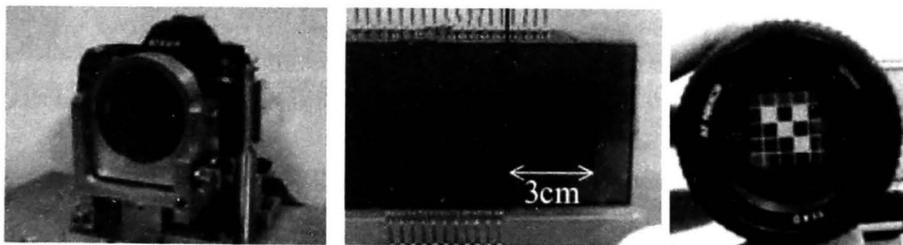


图 14-22 可编程编码孔径设备^[15]

14.5 总结与展望

图像复原是在图像的获取、传输以及保存过程中, 由于各种因素的影像所引起的几何失真, 都难免会造成图像模糊。图像去模糊技术在许多领域作为基础研究课题, 如图像处理、计算机视觉和空间探索等。总体来说, 虽然模糊图像处理算法已经取得了非常广泛的应用, 但是图像算法毕竟有其局限性, 不能将所有问题都寄希望于图像算法, 对于不同种类的模糊问题, 要区别对待。

目前, 模糊图像复原技术在国外已经取得了较好的成果。早期的图像复原是利用光

学的方法对失真的观测图像进行校正,而数字图像复原技术最早则是从对天文观测图像的后期处理中逐步发展起来的。其中一个成功的例子是NASA的喷气推进实验室在1964年用计算机处理有关月球的照片。照片是在空间飞行器上用电视摄像机拍摄的,图像的复原包括消除干扰和噪声,校正几何失真和对比度损失以及反卷积。另一个典型的例子是对肯尼迪遇刺事件现场照片的处理。由于事发突然,照片是在照相机移动过程中拍摄的,图像复原的主要目的就是消除移动造成的失真。美国的Cognitech软件是相当成熟的一套模糊图像复原商业软件,在美国FBI及其他执法机构中已有多年实际应用,其恢复出的图像可以直接当成法庭证据使用,可见模糊图像处理技术已经取得了相当成功的实际应用。以上这些应用都是基于图像复原算法的应用,在计算摄影的概念提出以后,在算法上创新的同时,也会有更多类似于编码曝光和编码孔径这样创新的光学成像装置和方法被提出,可以达到更好的图像复原效果,这是模糊图像复原技术未来发展的大方向。

参 考 文 献

- [1] L. Yuan. Image Deblurring[D]. Hong Kong: Hong Kong University. 2009.
- [2] Fergus R, Singh B, Hertzmann A, et al. Removing camera shake from a single photograph[C]//ACM Transactions on Graphics (TOG). ACM, 2006, 25(3): 787-794.
- [3] J. Jia. Single Image Motion Deblurring Using Transparency[C]// Computer Vision and Pattern Recognition, 2007. CVPR 2007. IEEE Conference on. IEEE, 2007: 1-8.
- [4] Joshi N, Szeliski R, Kriegman D J. PSF estimation using sharp edge prediction[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.
- [5] Raskar R. Computational photography[C]//Computational Optical Sensing and Imaging. Optical Society of America, 2009.
- [6] Raskar R, Agrawal A, Tumblin J. Coded exposure photography: motion deblurring using fluttered shutter[C]//ACM Transactions on Graphics (TOG). ACM, 2006, 25(3): 795-804.
- [7] Agrawal A, Xu Y. Coded exposure deblurring: Optimized codes for psf estimation and invertibility[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 2066-2073.
- [8] Agrawal A, Raskar R. Optimal single image capture for motion deblurring[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 2560-2567.
- [9] McCloskey S. Velocity-dependent shutter sequences for motion deblurring[M]//Computer Vision—ECCV 2010. Springer Berlin Heidelberg, 2010: 309-322.
- [10] Gottesman S R, Fenimore E E. New family of binary arrays for coded aperture imaging[J]. Applied Optics, 1989, 28(20): 4344-4352.
- [11] Zhou C, Lin S, Nayar S K. Coded aperture pairs for depth from defocus and defocus deblurring[J]. International journal of computer vision, 2011, 93(1): 53-72.
- [12] Levin A, Fergus R, Durand F, et al. Image and depth from a conventional camera with a coded aperture[J]. ACM Transactions on Graphics (TOG), 2007, 26(3): 70.
- [13] Caroli E, Stephen J B, Di Cocco G, et al. Coded aperture imaging in X-and gamma-ray astronomy[J]. Space Science Reviews, 1987, 45(3-4): 349-403.
- [14] Liu Y, Li H, Li Y, et al. Coded aperture enhanced catadioptric optical system for omnidirectional image deblurring[J]. Optik-International Journal for Light and Electron Optics, 2014, 125(1): 11-17.
- [15] C. Liang, T. Lin, B. Wong, C. Liu, H. Chen. Programmable aperture photography: multiplexed light field acquisition[J]. ACM Transactions on Graphics (TOG), 2008, 27(3): 55.



图1-4 蒙太奇处理



图1-13 PiCam合成后图像



图1-14 PiCam计算得到的深度图
(不同颜色表示不同的深度)



图2-1 摄影是用光的艺术

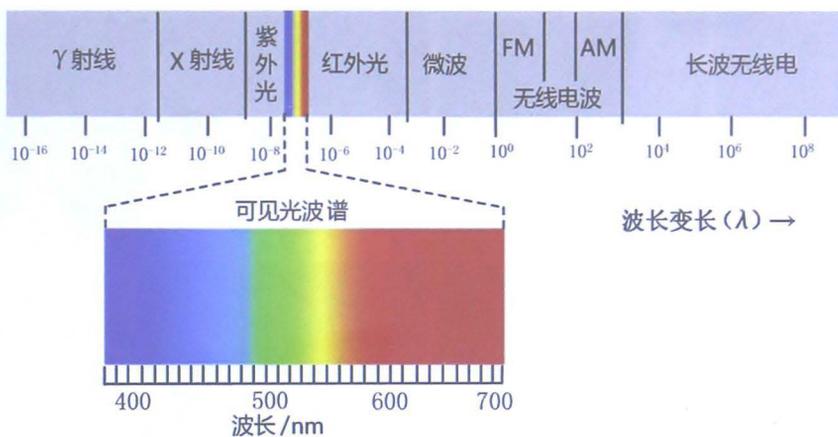


图2-3 光谱图

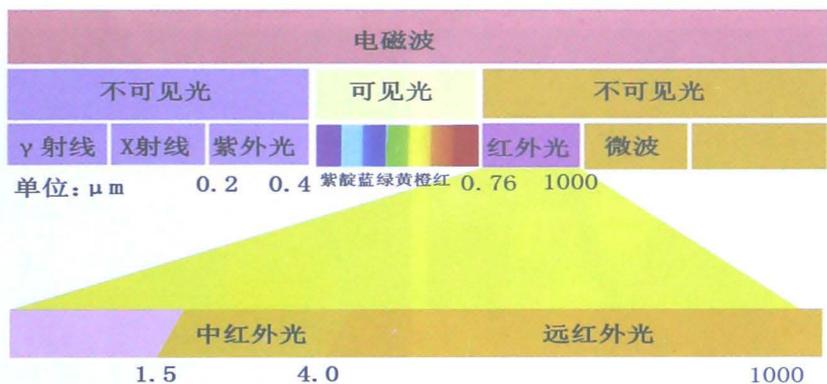


图2-5 红外光谱图

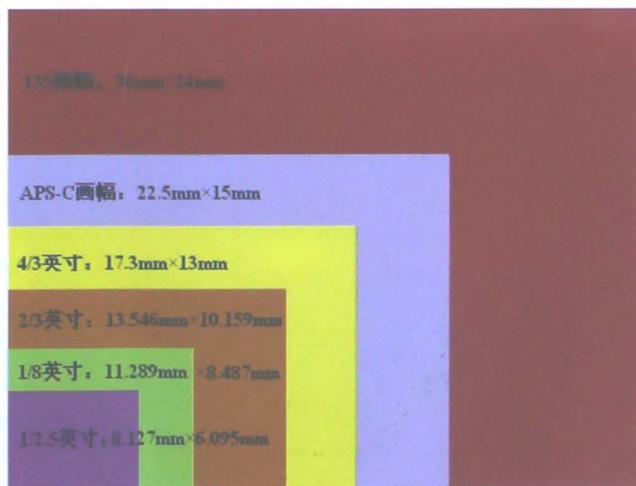


图2-21 画幅定义



图2-23 一幅背景与前景同样清晰的图像



图2-24 一幅背景被虚化的图像

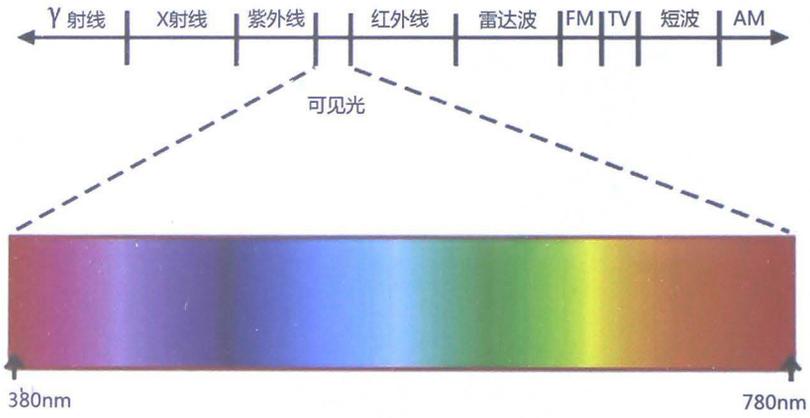


图3-2 电磁波/可见光的波谱划分

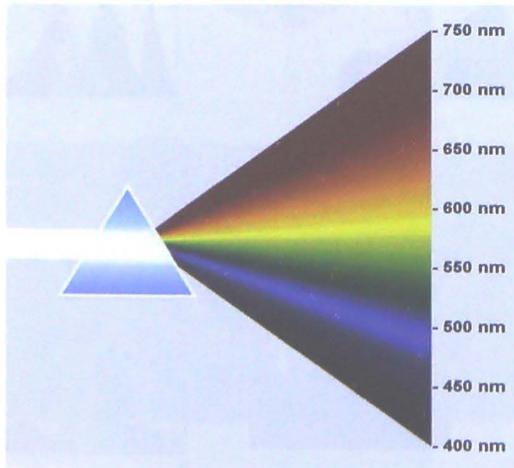


图3-3 基于棱镜的白光色散

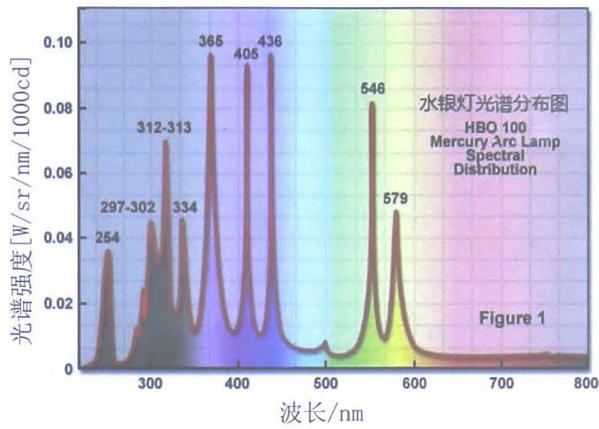


图3-4 水银灯的典型发射光谱分布

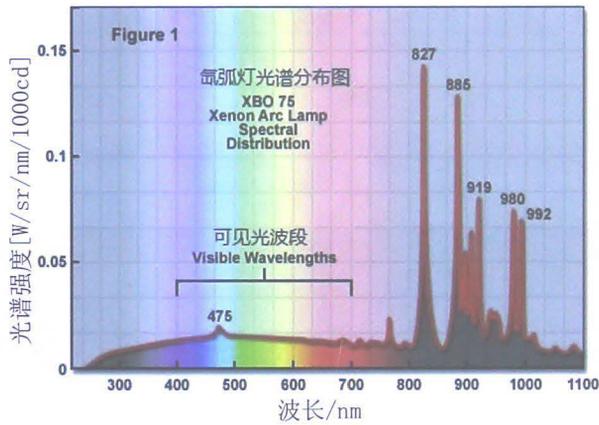


图3-5 氙弧灯的典型发射光谱分布

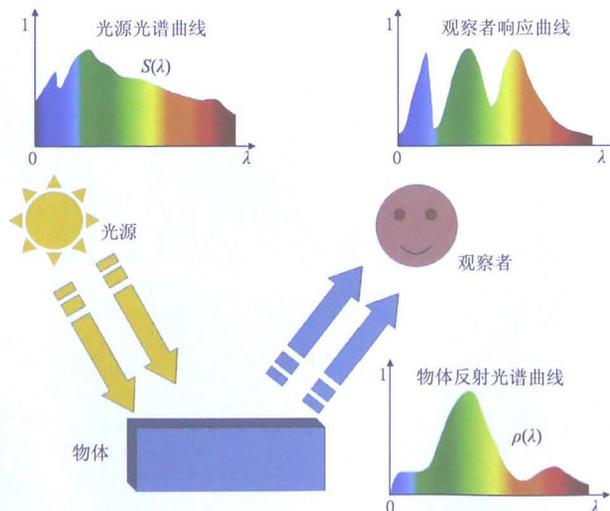


图3-6 不同物体光谱特性示意

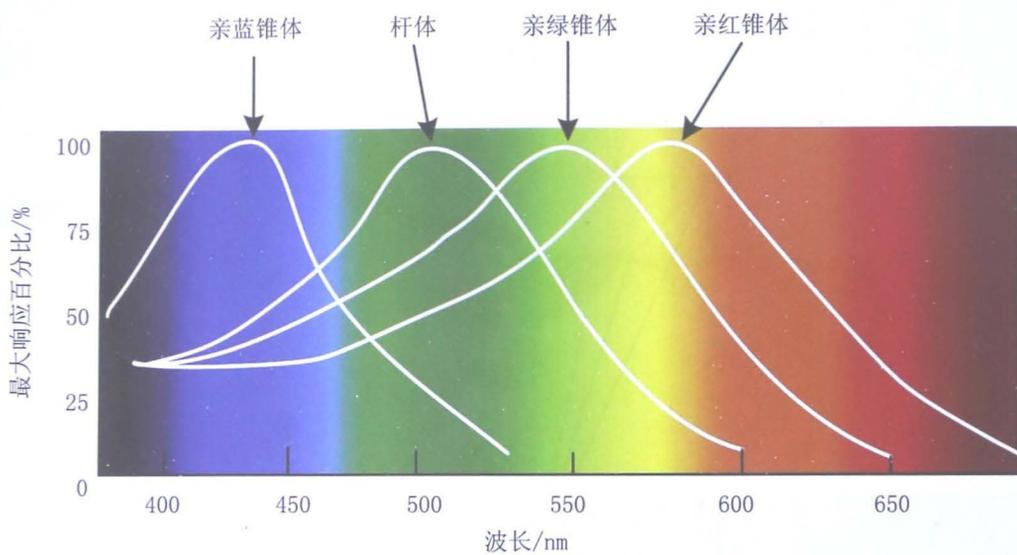
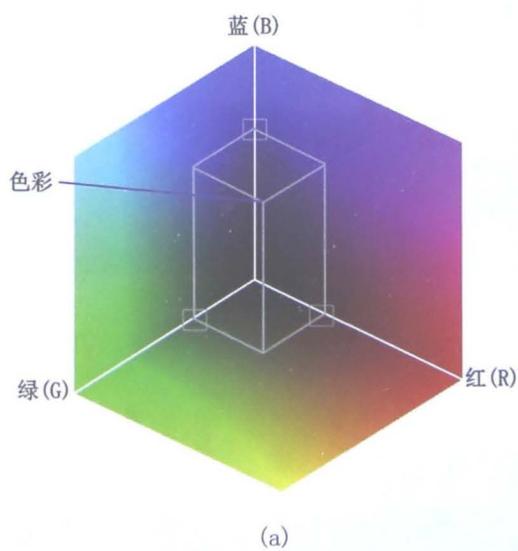
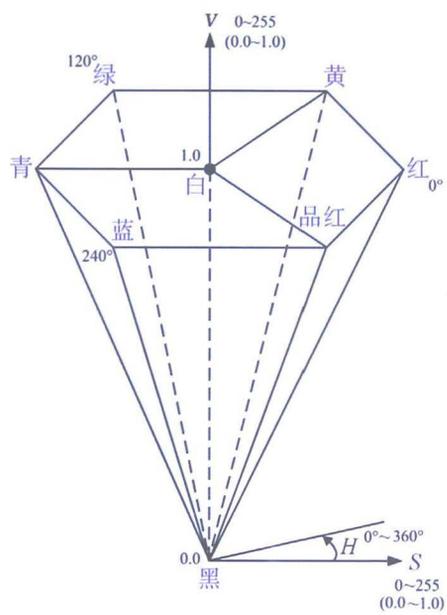


图3-8 视网膜锥体和杆体感光细胞的感光特性



(b)

图3-11 RGB颜色空间，RGB图像及R、G、B颜色通道分量



(a)



(b)

图3-12 HSV颜色空间，HSV图像及H、S、V颜色通道分量

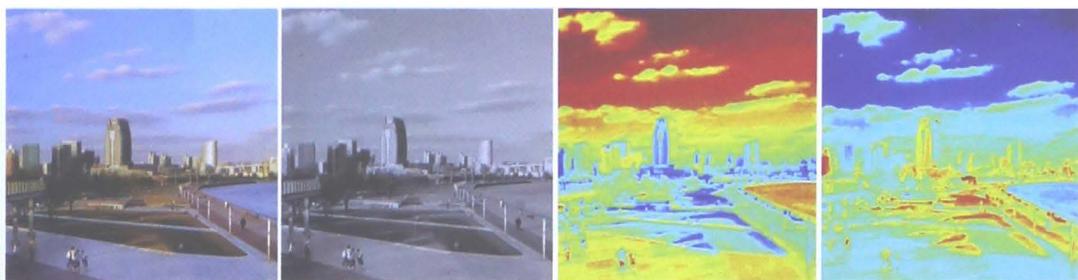
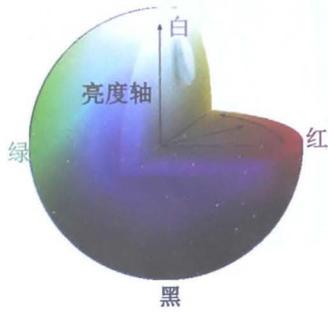


图3-13 YCbCr图像及Y、Cb、Cr颜色通道分量



(a)



(b)

图3-14 Lab颜色空间，Lab图像及Lab颜色通道分量

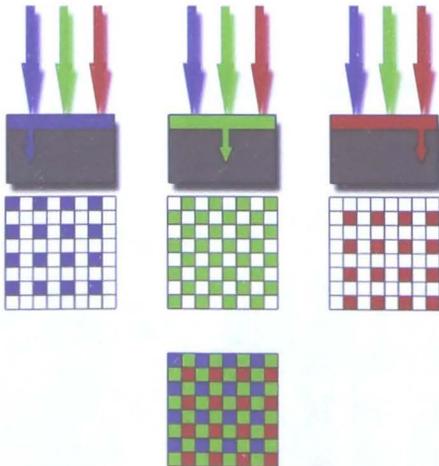


图3-16 Bayer格式图像获取示意

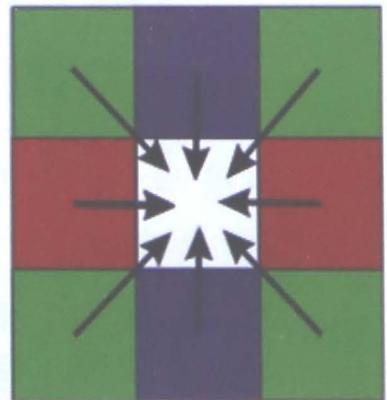
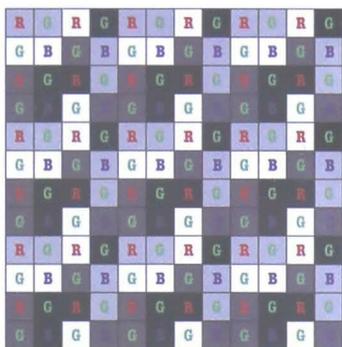


图3-17 邻域插值方法计算每个像素缺失的颜色通道



(a) Assorted Pixels设计

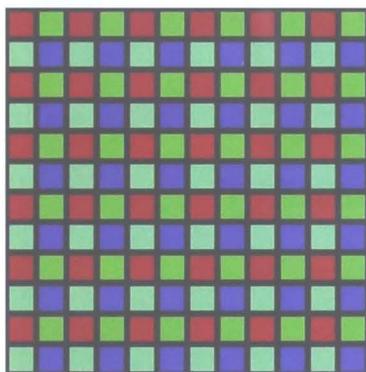


(b) 传统Bayer结果

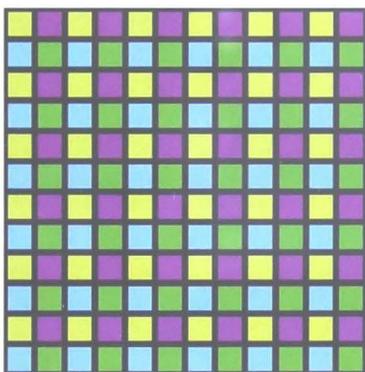


(c) Assorted Pixels结果

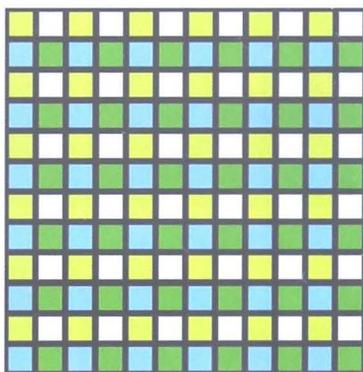
图3-18 Assorted Pixels彩色滤镜阵列及插值后的效果对比



(a) 红/绿/蓝/翠绿



(b) 减色法三原色+绿色



(c) 黄/蓝绿/绿/白

图3-20 其他创新的传感器彩色滤镜阵列设计

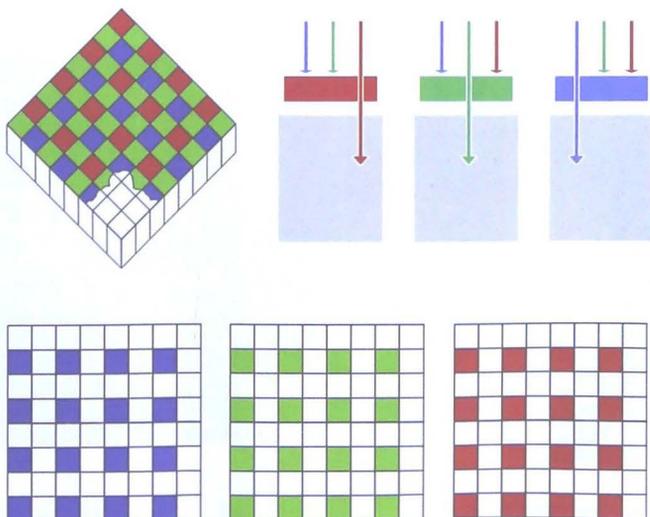
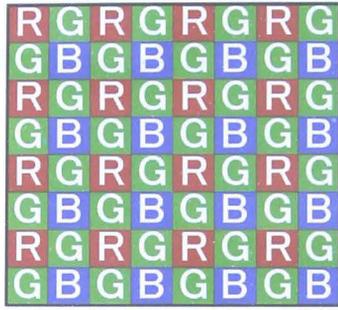
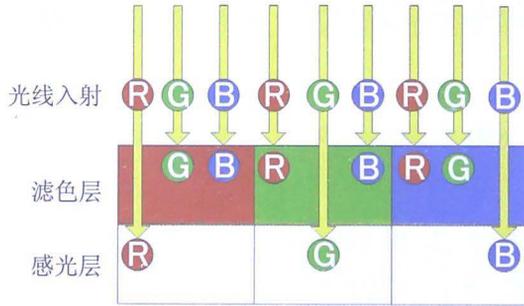


图4-8 CCD滤光片

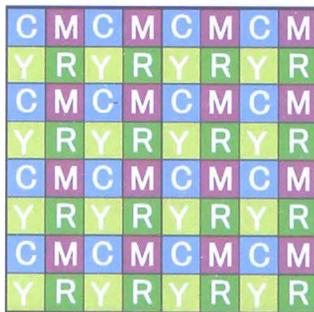


(a)

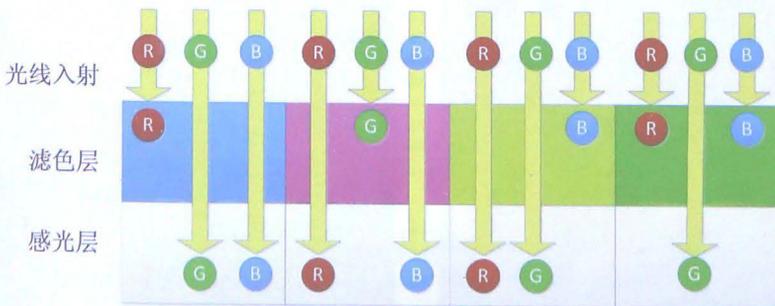


(b)

图4-9 RGB原色分色法



(a)



(b)

图4-10 CMYK互补分色法



图4-24 不同ISO拍摄效果

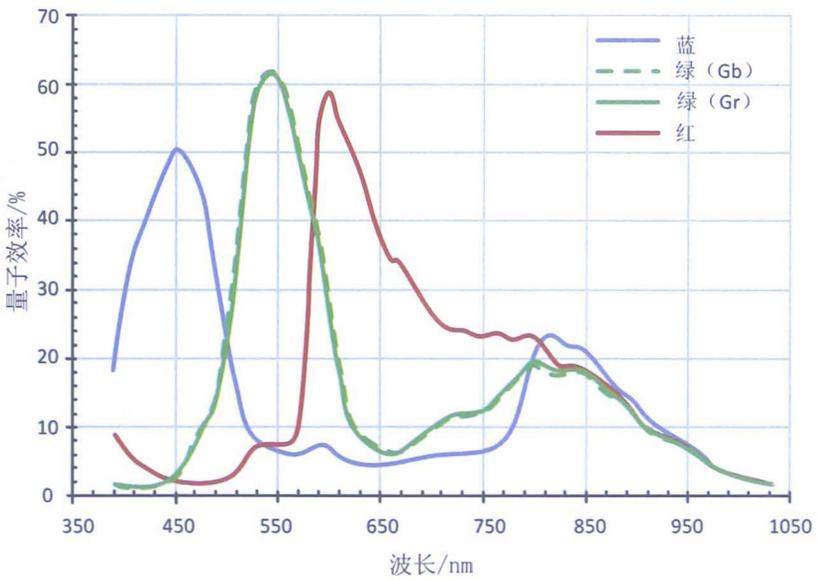
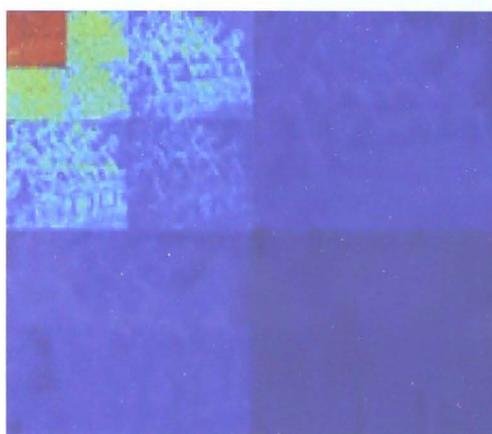
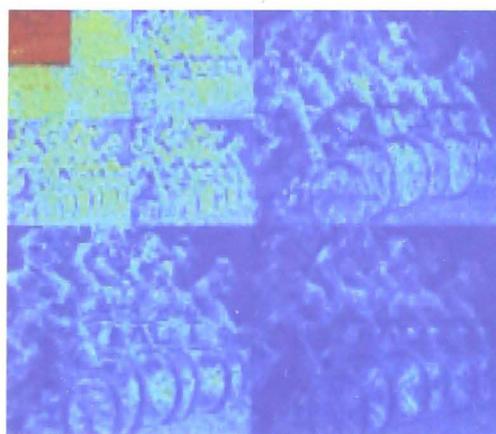


图4-27 CMOS量子效率图



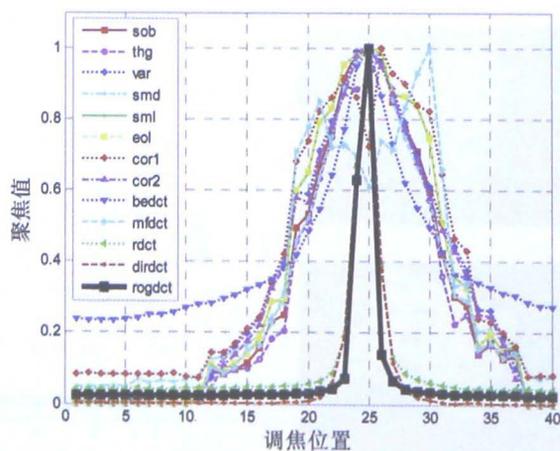
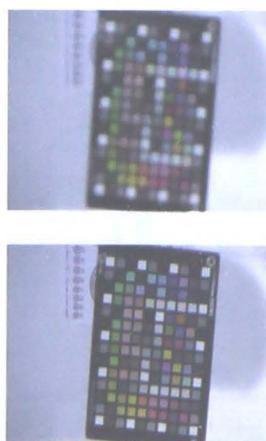
图5-7 8×8块状DCT变换



(a) 无模糊的原图即高斯模糊 $\delta = 0$

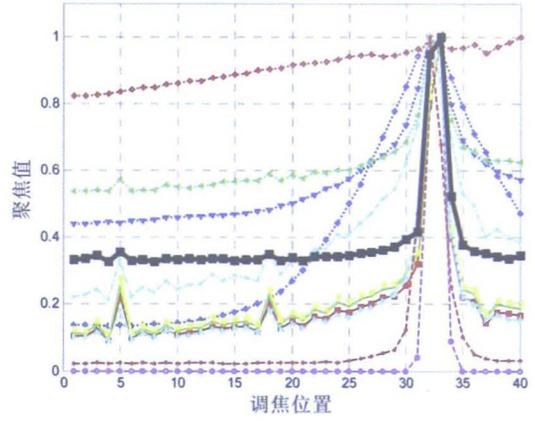
(b) 高斯模糊 $\delta = 1.3$

图5-9 两幅清晰度不同的图像对应的重组后的DCT图

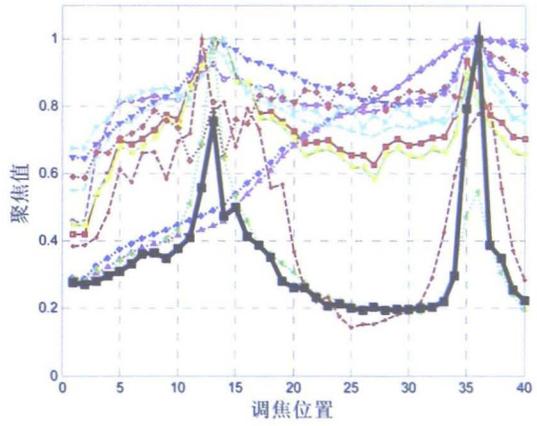
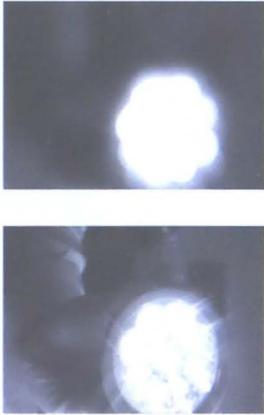


(a) 第一组, 准确聚焦位置为25

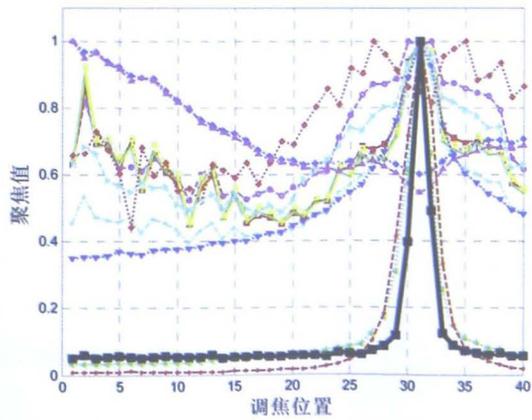
图5-10 聚焦图像序列样图及各函数的聚焦曲线



(b) 第二组, 准确聚焦位置为33



(c) 第三组, 准确聚焦位置为36



(d) 第四组, 准确聚焦位置为31

图5-10 聚焦图像序列样图及各函数的聚焦曲线 (续)



(a) $\delta = 0.677051$



(b) $\delta = 1.307261$



(c) $\delta = 0.791634$



(d) $\delta = 1.536428$

图5-12 具有不同程度的高斯模糊图像， δ 为高斯模糊核标准差

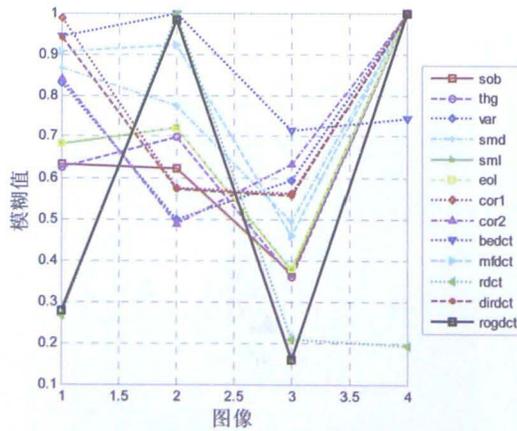


图5-13 对图5-12各图像采用锐度评价函数计算得到的模糊度值

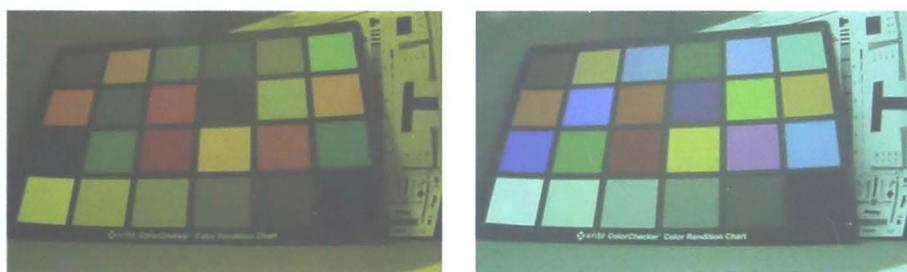


(a) 过曝

(b) 正常曝光

(c) 欠曝

图6-1 不同曝光的图像



(a)

(b)

图6-2 偏色的图像



图6-3 色温变化示意图

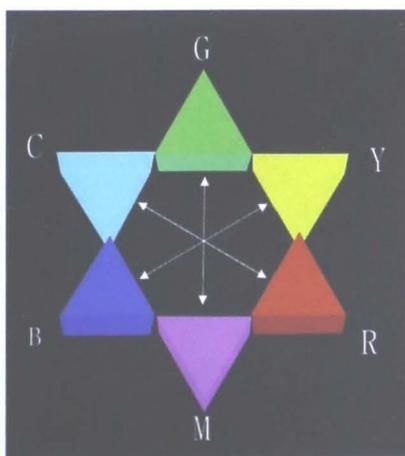


图6-4 白纸的色度和画面色度成补色关系



(a) 中央平均测光



(b) 点测光



(c) 中央部分测光

图6-9 高对比度条件下的测光



图6-10 曝光补偿前后效果图



(a) 未经白平衡校正



(b) 灰世界法校正

图6-16 灰世界法自动白平衡效果



(a) 未经白平衡校正



(b) 灰世界法校正

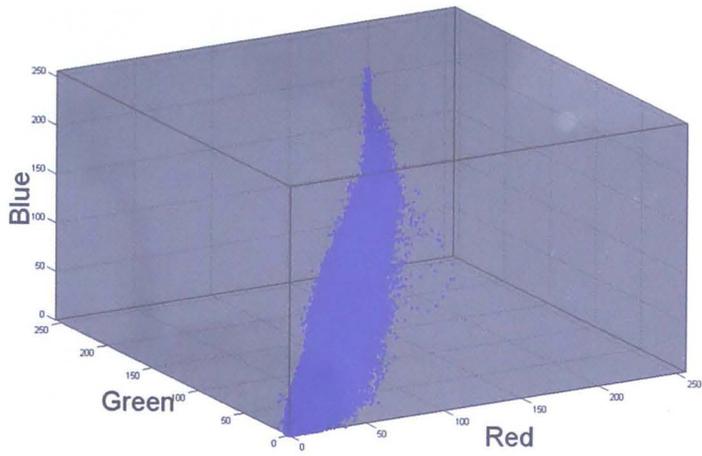


(c) 上述简单灰边缘法校正

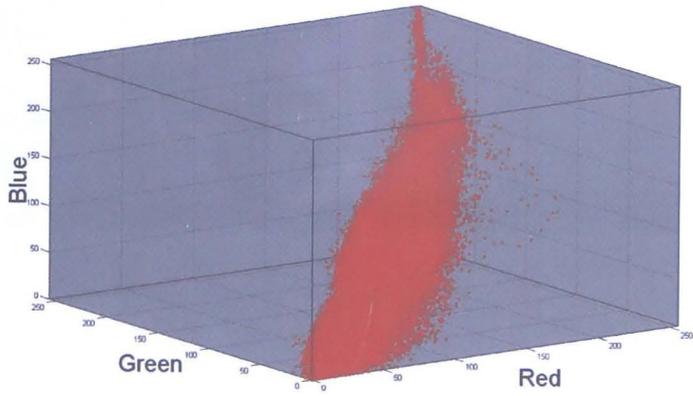
图6-19 简单灰边缘自动白平衡效果对比



图6-20 暖光源下的场景



(a) 冷光源下的色域



(b) 暖光源下的色域

图6-21 不同光源下的色域

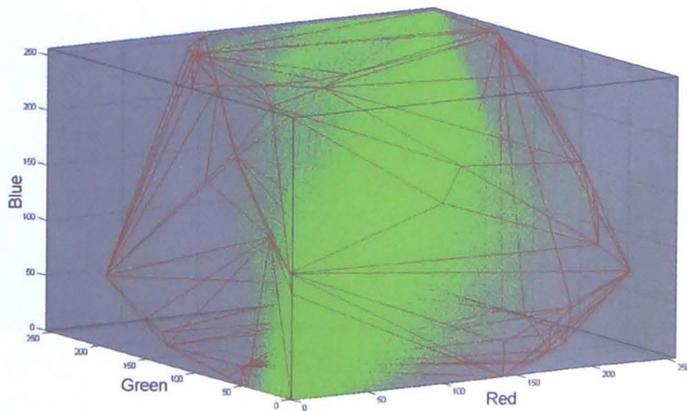


图6-22 标准色域的凸壳



(a) 原始图像

(b) 校正后的图像

图6-23 本章基于数据驱动的自动白平衡效果



图7-3 高动态范围图像再现图7-2中原始场景结果



图7-5 同一场景的不同曝光度图像序列和通过合成之后再再现出的结果图



$\tau = 0.01$ (太亮)



$\tau = 0.1$ (手动选择值)



$\tau = 1.0$ (太暗)



$\tau = 0.09$ (自动选择值)

图7-9 用式 (7.18) 得到的高动态图像的色调映射图

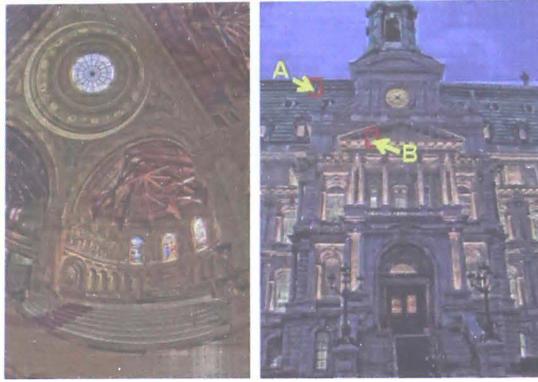
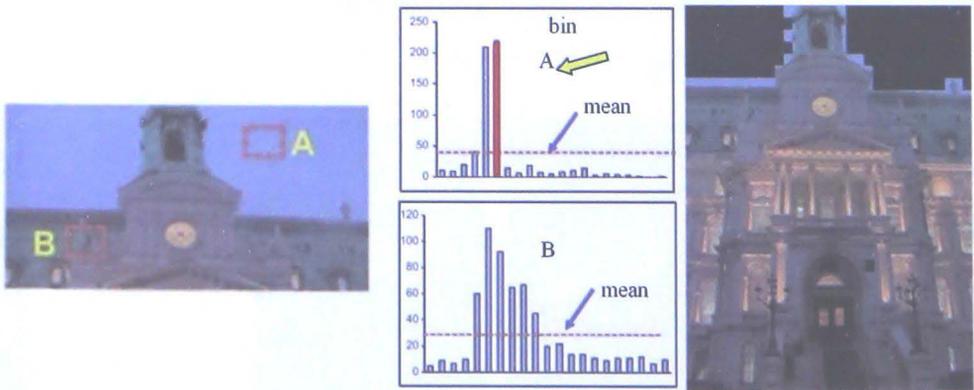


图7-10 在每一块上面做HALEQ后所得到的结果



图7-12 使用权重平均的方法去除边界效应后的结果



(a) 原始图像

(b) 平滑区域和非平滑区域的直方图对比

(c) 平滑区的检测结果

图7-13 平滑区域和非平滑区域的效果对比



图7-14 在考虑平滑区域之前和之后的对比



$\sigma_s=0.1$

$\sigma_s=0.05$

图7-15 σ_s 取不同值对应的效果图

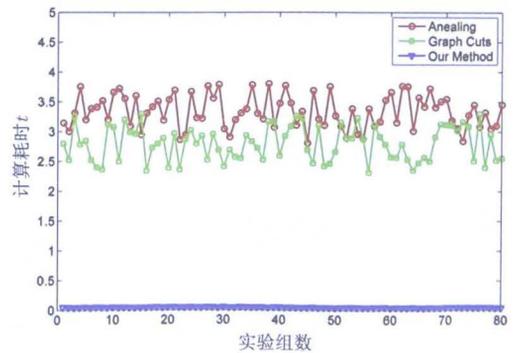
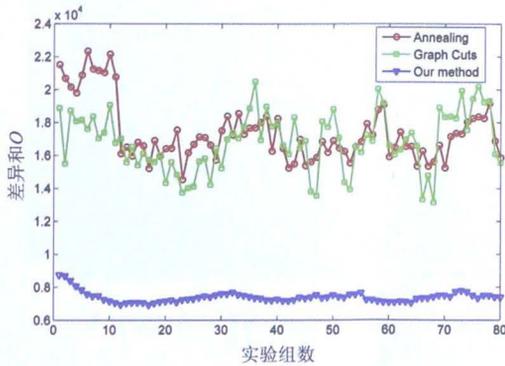


图9-11 80组图像的缝合线最优性指标 O 与计算耗时 t 数据统计图

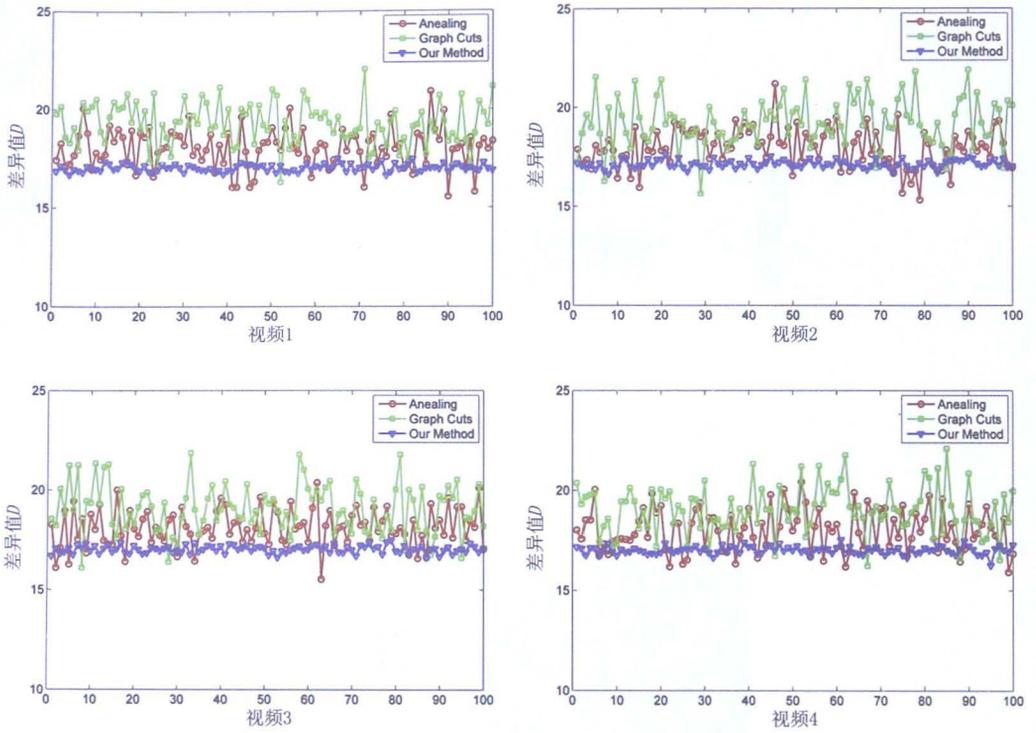


图9-13 利用三种不同算法拼接四路视频数据, 得到的视频图像前后帧差异统计图



图12-1 稀疏表示去噪方法效果



图12-11 稀疏表示去噪效果

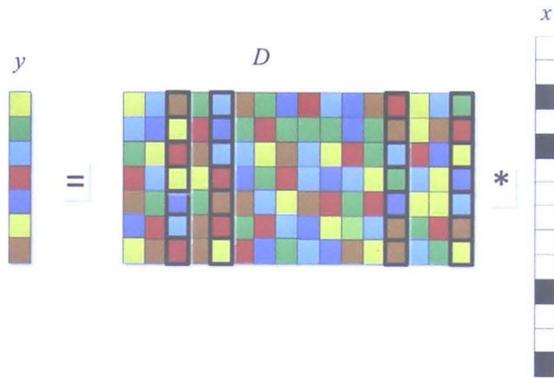
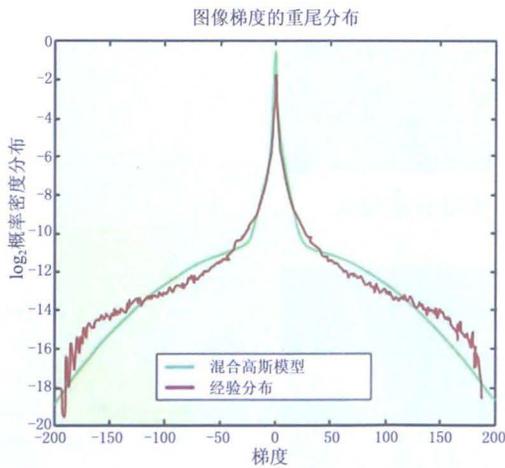


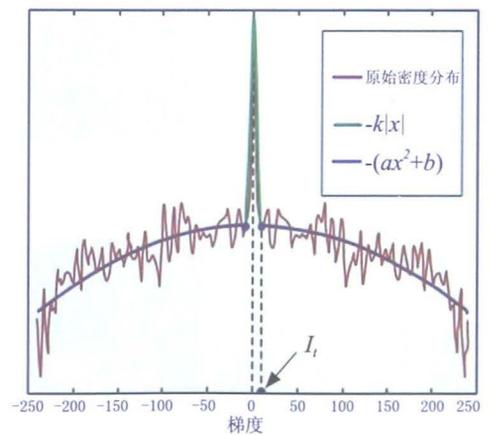
图12-12 过完备稀疏表示模型示意



(a) 一个自然场景



(b) 图像梯度的对数密度曲线



(c) 设计一个参数模型来近似对数密度曲线

图14-7 采用图像梯度先验的核估计