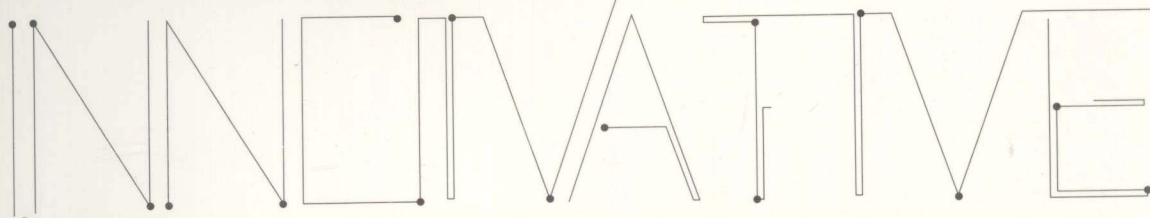




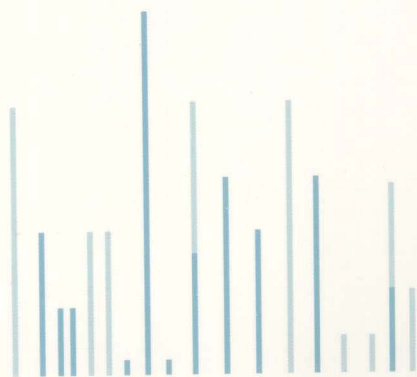
创新技术学术专著



复杂场景下 图像与视频分析

Image and Video Analysis
in Complex Scenes

龚声蓉 刘纯平 季怡 著



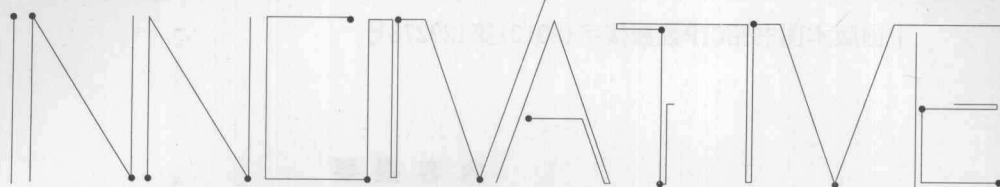
人民邮电出版社
POSTS & TELECOM PRESS

013162649

TN941.3
25



创新技术学术专著



复杂场景下 图像与视频分析

Image and Video Analysis
in Complex Scenes

龚声蓉 刘纯平 季怡 著



北航

C1670343

TN 941.3

25

人民邮电出版社

北京

图书在版编目(CIP)数据

复杂场景下图像与视频分析 / 龚声蓉, 刘纯平, 季怡著. — 北京: 人民邮电出版社, 2013. 9
ISBN 978-7-115-32239-5

I. ①复… II. ①龚… ②刘… ③季… III. ①数字视频系统—系统分析 IV. ①TN941.3

中国版本图书馆CIP数据核字(2013)第129273号

内 容 提 要

图像分析是指用模式识别和人工智能方法对景物进行分析、描述、分类和解释, 又称景物分析或图像理解。视频分析是指使用图像分析技术, 通过将场景中的背景和目标分离进而分析并追踪在摄像机场景内出现的目标。本书从基本概念入手, 系统地介绍了图像和视频分析的基本理论, 包括图像与视频分割、特征提取与描述、分类器设计等, 并结合理论重点介绍了文档图像版面分析、场景分类、运动目标检测和跟踪3个应用实例。全书突出了理论和实践相结合的特点。

本书可作为高校计算机科学、电子工程、自动化、生物医学、遥感、地质、矿业、通信、气象、农业等相关专业研究生和高年级本科生教材, 也可供相关领域的大学教师、科研人员和工程技术人员参考。

-
- ◆ 著 龚声蓉 刘纯平 季 怡
责任编辑 代晓丽
责任印制 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 26 2013年9月第1版
字数: 616千字 2013年9月北京第1次印刷
-

定价: 98.00 元

读者服务热线: (010)67119329 印装质量热线: (010)67129223

反盗版热线: (010)67171154

前 言

图像分析是指用模式识别和人工智能方法对景物进行分析、描述、分类和解释,又称景物分析或图像理解。它一般利用数学模型并结合图像处理的技术来分析底层特征和上层结构,倾向于对图像内容的分析、解释和识别。视频分析是指使用图像分析技术,通过将场景中的背景和目标分离进而分析并追踪在摄像机场景内出现的目标。由于视频由一帧帧的图像组成,因此,本书将更多篇幅放在图像分析上。

图像分析和视频分析包含 4 个基本过程。

(1) 图像输入。把实际物景转换为适合计算机处理的表达形式,即把三维物景转换成二维图像。

(2) 分割。从景物图像中分解出物体和它的组成部分。

(3) 识别或者分类。对图像中分割出来的物体给予相应的名称,如自然物景中的车辆、行人、建筑物等。

(4) 解释。结合目标识别建立物景的分级构造,说明物景中有什么目标、目标之间的相互关系、图像是什么场景等。

本书正是从基本概念入手,将理论与实践相结合,以图像和视频分析的基本过程为主线,系统地介绍图像和视频分析的基本理论应用。全书包括两个部分。第一部分为基本理论,包括图像与视频分割、特征提取与描述、分类器设计等;第二部分为实例,包括基于 SVM 的文档图像版面分析、基于显著性和 LDA 主题模型的图像场景分类、运动目标检测和跟踪 3 个应用实例,在实例部分,结合理论又分别介绍了多种处理方法。

本书适用于作为计算机技术、电子技术、网络和通信工程相关专业研究生和高年级本科生的教材或参考书,也可作为从事图像处理、模式识别研究和开发的工程技术人员的参考书。

本书受苏州大学研究生教育基金和 211 工程建设经费的资助,是作者开展国家自然科学基金(基于二型模糊概率图模型的多摄像头目标跟踪研究(61170124)和基于显著性和信任传递的动态场景主题发现(61272258))和多个江苏省自然科学基金及其他相关课题研究的结晶。全书由龚声蓉教授、刘纯平副教授、季怡博士共同完成,青年教师蒋德茂和陈宇也参与了部分撰写工作,邹羚老师和研究生林卉、刘龙等参与了文字录入、校对工作。同时,本书也参考国内出版的大量书籍和论文,



对所引用论文和书籍的作者深表感谢。人民邮电出版社为本书出版做了大量工作，在此表示衷心感谢。

由于作者学术水平的限制，书中难免有不足和不当之处，恳请读者批评指正。

作者

2013年5月于苏州大学

目 录

第 1 章 绪论	1
1.1 视觉感知与彩色模型	1
1.1.1 视觉感知	1
1.1.2 彩色模型	3
1.2 图像和视频的数字化	8
1.2.1 图像的数字化	8
1.2.2 视频的数字化	10
1.3 图像和视频分析的相关概念	10
1.3.1 数字图像处理	10
1.3.2 图像分析	11
1.3.3 视频分析	12
1.4 图像与视频分析的研究内容	13
1.4.1 图像与视频场景分割	13
1.4.2 图像与视频场景特征描述	14
1.4.3 图像/视频中的目标识别	15
1.4.4 场景描述与理解	16
1.5 图像与视频分析的应用	17
1.5.1 工业方面的应用	17
1.5.2 医学图像分析	18
1.5.3 遥感领域应用	18
1.5.4 军事公安领域的应用	19
1.5.5 交通	20
1.5.6 其他	21

1.6 小结	22
参考文献	22
第 2 章 图像分割技术	24
2.1 图像与视频分割概述	24
2.2 几种经典图像分割方法	25
2.2.1 基于边缘的分割	25
2.2.2 基于阈值的分割	32
2.2.3 基于区域的分割	37
2.3 基于形态学分水岭的分割	39
2.3.1 形态学图像处理基本概念和运算	39
2.3.2 基于分水岭的分割	43
2.4 基于聚类的分割	46
2.4.1 C-均值聚类方法	46
2.4.2 模糊 C-均值聚类方法	47
2.5 基于图论的图像分割方法	49
2.5.1 基本原理	49
2.5.2 GraphCut 及其改进图像分割方法	50
2.5.3 其他基于图的分割方法	55
2.6 偏微分图像分割	56
2.6.1 Snakes 及其改进模型	57
2.6.2 水平集 (Level Set) 方法	58
2.6.3 基于变分水平集的图像分割	62
2.7 多特征融合的图像分割	63



2.8 彩色图像分割	64	3.3.3 基于统计信息的特征 提取	184
2.8.1 直方图阈值法	64	3.4 图像中层语义描述	185
2.8.2 彩色空间聚类法	65	3.4.1 视觉词包 (Bag-of-Visiterms, BOV)	185
2.8.3 融合颜色和空间信息的 彩色图像分割	66	3.4.2 语义主题	188
2.9 视频对象分割	68	3.5 图像高层语义特征描述	189
2.9.1 基于时空的视频对象 分割	68	3.5.1 语义提取模型	190
2.9.2 基于运动的视频对象 分割	71	3.5.2 语义关联	192
2.9.3 交互式视频对象分割	79	3.5.3 高层语义描述推理	193
2.9.4 基于高斯混合模型的 自适应阴影检测	79	3.6 运动特征描述	199
2.10 小结	82	3.6.1 基于 MPEG-7 的运动 特征描述	201
参考文献	82	3.6.2 非参数模型	202
第 3 章 特征描述与提取	86	3.6.3 参数模型	206
3.1 概述	86	3.6.4 基于特征的运动估计	210
3.1.1 视觉特征	87	3.7 小结	210
3.1.2 图像内容	87	参考文献	211
3.1.3 图像特征	87	第 4 章 分类器设计	218
3.1.4 图像特征提取	89	4.1 概述	218
3.1.5 特征选择	89	4.1.1 生成模型	218
3.2 常用的低层视觉特征	90	4.1.2 判别模型	220
3.2.1 颜色特征描述	91	4.1.3 混合生成—判别模型	221
3.2.2 纹理特征描述	96	4.2 常见分类器	222
3.2.3 形状特征描述	127	4.2.1 贝叶斯分类器	222
3.2.4 局部特征描述	150	4.2.2 SVM 分类器	228
3.2.5 视觉特征的比较	182	4.2.3 强化学习分类器	235
3.3 其他低层特征提取	183	4.2.4 神经网络分类器	250
3.3.1 图像代数特征	183	4.2.5 基于点集 Voronoi 图的 分类器	256
3.3.2 图像变换系数特征	184	4.2.6 最近邻凸包分类器	258

4.2.7 遗传算法和免疫算法的 分类器	262	5.5 阅读顺序未知的纯文本图像 版面分析	306
4.2.8 基于稀疏表示的 分类器	265	5.5.1 已有算法分析	307
4.3 多分类器集成	267	5.5.2 基于 SVM 的复杂纯文本 图像版面分析算法	308
4.3.1 集成学习的有效性和 条件	268	5.6 小结	312
4.3.2 基分类器产生方法	269	参考文献	312
4.3.3 多分类器集成系统 结构	270	第 6 章 基于显著性和 LDA 主题模型 的图像场景分类	315
4.3.4 多分类器设计方法	272	6.1 基于语义生成模型的图像 场景分类概述	315
4.3.5 分类器的选择准则	273	6.2 基本术语	317
4.3.6 Adaboost 分类器	274	6.3 图像中的主题模型	318
4.4 小结	279	6.4 基于主题模型的场景分类 方法基本框架	319
参考文献	280	6.5 基于显著性的主题模型 场景分类	319
第 5 章 基于 SVM 的文档图像版面 分析	286	6.5.1 视觉词包表示	320
5.1 概述	286	6.5.2 视觉词包生成	321
5.1.1 版面分析发展历程	286	6.5.3 基于频域显著性的视觉 词包生成	322
5.1.2 版面分析方法分类	288	6.5.4 统计可视单词的词频表	327
5.2 文档图像倾斜检测	290	6.5.5 改进 LDA 模型在图像 上的实现	327
5.2.1 倾斜校正方法概述	290	6.6 本章小结	340
5.2.2 改进的最近邻链倾斜 检测算法	292	参考文献	340
5.3 简单背景下版面分割	298	第 7 章 运动目标检测与跟踪	342
5.3.1 基于边缘检测的连通区 构造	298	7.1 概述	342
5.3.2 结合连通区和游程平滑 版面分割	299	7.2 基于背景差分的目标检测	345
5.3.3 文本区域二值化	303	7.2.1 背景建模概述	345
5.4 复杂背景下版面分割	304	7.2.2 单高斯背景建模	347



7.2.3 高斯混合背景建模	347
7.2.4 引入二型模糊的混合高斯背景建模	351
7.3 非参数核密度估计的运动目标检测	356
7.3.1 经典方法	357
7.3.2 基于时间窗的核密度估计运动检测	358
7.3.3 聚类差分核密度估计算法	363
7.4 检测结果后处理	366
7.5 运动目标跟踪	367
7.5.1 Kalman 预测与全局特征	

匹配跟踪	368
7.5.2 基于 Mean Shift 的目标跟踪	373
7.5.3 改进的 Mean Shift 跟踪方法	377
7.5.4 基于粒子滤波的跟踪	378
7.5.5 抗遮挡目标跟踪	385
7.5.6 基于角点抽样的互遮挡多目标跟踪	392
7.6 小结	399
参考文献	399

名词索引	404
------	-----

第1章 绪论

1.1 视觉感知与彩色模型

人类视觉系统是人类获取外部图像和视频信息的工具。光辐射刺激人眼时,将会产生复杂的生理和心理变化,这种感觉就是视觉。视觉是人类最重要,同时也是最完美的感知手段。据统计,在人类接受的信息中,图像等视觉信息所占的比重达到75%。对人的视觉系统研究涉及光学、色度学、视觉心理学、神经科学和认知科学等许多领域。因此,本节首先回顾一下人的视觉感知机理及彩色空间。

1.1.1 视觉感知

光和人类的活动有着十分密切的关系。人的视觉离不开光,景物反射的光作用于人眼后,经过复杂的心理和生理过程,才能感觉到景物的存在。

描述人类色彩感觉的属性有两个:亮度和色度。亮度是被感知的光的明暗程度,与可视频带中的总能量成正比。色度描述被感知的光的颜色和深浅。人眼可以感觉到的光波长在380~780 nm范围内。光的彩色感觉取决于光谱成分,即它的波长组成。例如,能量集中在700 nm附近的光显示为红色。饱和度指的是颜色有多纯,它是由光谱的范围或带宽决定的。

人眼对光的的感觉始于视网膜的感光细胞。有两种类型的感光细胞:锥状细胞和杆状细胞。锥状细胞在亮光下起作用,能感受彩色色调;杆状细胞工作在低光环境下,只能提取亮度信息。当一束光照射到物体上时,在某个波长范围内的能量被吸收,而剩下的则被反射出来。反射出来的彩色取决于入射光的光谱成分和被吸收的波长范围。由于人们日常看到的图像一般是由目标物体上反射出的光量度而得到的,因此,一幅图像基本上由以下两部分构成^[1]。

(1) 入射到可见场景上光的量。

(2) 场景中目标物体反射光的量。

确切地说,它们分别称为照度成分和反射成分。由于图像与照度成分和反射成分都成正比,所以可以认为图像是由照度和反射量相乘得到的。照度是由光源决定的,而反射成分是由场景中的目标物体特性所决定的。因此,图像实际上记录的是物体辐射能量的空间分布,这个分布是空间坐标(x, y, z)、时间(t)和波长(λ)的函数,可以表示为

$$I=f(x, y, z, \lambda, t) \quad (1-1)$$



三维视频是时域离散的图像帧序列 $f(x, y, z, \lambda, t_n)$, $n=1, 2, \dots$ 在连续播放时人眼的主观感觉, 可以表示为^[2]:

$$I = V[f(x, y, z, \lambda, t_n)] \quad (1-2)$$

其中, $V[\cdot]$ 为人眼的视觉效应。

人们常见的视频, 如电视, 是二维平面活动图像, 可以表示为光的强度(I)随平面坐标(x, y)、波长(λ)的函数随时间(t)的变化, 即

$$I = f(x, y, \lambda, t) \quad (1-3)$$

当一幅图像为平面静止图像时, 空间坐标变量 z 和时间变量 t 可以从函数中去除。这样, 一幅图像可以用下面的二维函数 $f(x, y)$ 来表示, 即

$$I = f(x, y, \lambda) \quad (1-4)$$

在特定的坐标(x, y)处, f 的值或幅度是一个正的标量, 当图像为单色图像时, λ 可从函数中去除, 该标量就是图像值的灰度级, 它正比于光源的辐射能量。因此, $f(x, y)$ 一定是非零和有限的, 即

$$0 < f(x, y) < \infty \quad (1-5)$$

由上面分析可知, 函数 $f(x, y)$ 可由两个分量来表征。

(1) 入射到观察场景的光源总量, 称为入射分量, 表示为 $i(x, y)$ 。

(2) 场景中物体反射光的总量, 称为反射分量, 表示为 $r(x, y)$ 。

由此, 函数 $f(x, y)$ 可以表示为

$$f(x, y) = i(x, y)r(x, y) \quad (1-6)$$

其中,

$$0 < i(x, y) < \infty \quad (1-7)$$

$$0 < r(x, y) < 1 \quad (1-8)$$

式(1-8)指出反射分量限制在 0 (全吸收) 和 1 (全反射) 之间。 $i(x, y)$ 的性质取决于照射源, 而 $r(x, y)$ 取决于成像物体的特性。因此, 同一场景不同的光照, 所形成的图像会有明暗之分; 而同一光照条件但场景不同, 所形成的图像可能会完全不同。

显然, 对于单色图像, 设在任何坐标处的强度为图像在那一点的灰度级为 I , 则有

$$L_{\min} \leq I \leq L_{\max} \quad (1-9)$$

区间 $[L_{\min}, L_{\max}]$ 称为灰度级。实际常常令该区间为 $[0, L-1]$, 这里 $I=0$ 为黑, $I=L-1$ 为白。所有中间值是从黑到白的各种灰色调。

图像按空间坐标和明暗程度的连续性不同, 可分为模拟图像和数字图像。模拟图像指空间坐标和明暗程度都是连续变化的、计算机无法直接处理的图像。数字图像是一种空间坐标和灰度均不连续的、用离散的数字表示的图像。这样的图像才能被计算机处理。因此, 数字图像可以理解为图像的数字表示, 是时间和空间的非连续函数, 是为了便于计算机处理的一种图像表示形式。它是由一系列离散单元经过量化后形成的灰度值的集合, 即像素的集合。

按图像的维数划分, 可分为二维图像、三维图像和多维图像; 按图像内容的变化性质分, 可分为静态图像和活动图像, 活动图像也称为序列图像或视频。

视频信号是从动态的三维场景物投影到视频摄像机平面上的一个二维图像序列^[3]。因此，它实际是一个由在时间轴上的许许多多幅图像组成的图像序列，其中的每一幅图像称为一帧。一个视频帧中的任何一点的色彩值记录了所观察景物中一个特定的三维点所发出或反射的光。

1.1.2 彩色模型

要准确描述颜色特性需要选择一个合适的颜色空间，并用量化的方式将颜色特征表征为向量的形式^[4]。可以用亮度（Value）、色调（Hue）、饱和度（Saturation）来描述。色调指光的颜色，如红、橙、黄、绿、青、蓝、紫。饱和度指彩色的深浅程度，即与一定的色调的纯度相关。饱和度高表示颜色深，如深红；饱和度低，则表示颜色浅，如浅红。亮度指人眼感受到的光的明暗程度。人眼看到任一彩色光都是这3个特性的综合效果。自然界常见的各种颜色光都可由红（R）、绿（G）、蓝（B）3种颜色按不同比例相配而成。

表示数字图像的颜色空间（颜色模型）有很多种，每一种颜色空间都是在某种特定上下文中对颜色的特性和行为的解释。没有哪一种颜色空间可以解决所有的颜色问题。因此一个恰当的颜色空间是解决相应的颜色问题的基础。常见的颜色空间包括：RGB（Red, Green, Blue）颜色空间、HSV（Hue, Saturation, Value）颜色空间、CMYK 颜色空间、HSL 颜色空间、HSB 颜色空间、HIS 颜色空间、YCbCr 颜色空间、XYZ 颜色空间、Lab 颜色空间以及 LUV 颜色空间等。

1. RGB 颜色空间

RGB 颜色空间是最常用的颜色空间之一，是一种与设备相关的色彩空间。这种颜色空间在常见的一些硬件中使用非常广泛。计算机显示器系统采用这一颜色空间，如彩色阴极射线管、彩色光栅图形的显示器都使用 R、G、B 数值来驱动 R、G、B 电子枪发射电子，并分别激发荧光屏上的 R、G、B 3 种颜色的荧光粉发出不同亮度的光，并通过相加混合产生各种颜色；扫描仪也是通过吸收原稿经反射或透射而发送来的光线中的 R、G、B 成分，并用它来表示原稿的颜色。RGB 颜色模型所覆盖的颜色域取决于显示设备荧光点的颜色特性，是与硬件相关的。因此不同的扫描仪扫描同一幅图像，会得到不同色彩的图像数据；不同型号的显示器显示同一幅图像，也会有不同的色彩显示结果。

在 RGB 颜色空间中，每种颜色出现在红、绿、蓝的原色光谱分量中，这个基于三维笛卡尔坐标系统，所考虑的彩色子空间如图 1-1 所示的立方体。图 1-1 中，R、G、B 位于 3 个角上；青、品红和黄位于另外 3 个角上，黑色在原点处，白色位于离原点最远的角上。在该模型中，红、绿、蓝原色是加性原色，各个原色混合在一起可以产生复合色。不同的灰度等级沿着从原点的黑色（0, 0, 0）到点（1, 1, 1）的白色的主对角线分布，在主对角线上各原色的强度相等。

大部分的数字图像都是用 RGB 颜色空间表示的。在 RGB 颜色空间中，数字图像由 3 个颜色分量组成，每一幅红、绿、蓝图像都是一幅 8 bit 颜色深度的图像，所以 RGB 彩色图像的颜色深度为 24 bit。所说的真彩色图像常定义为 24 bit 的彩色图像。

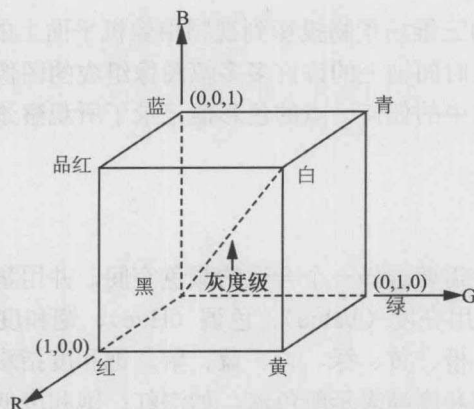


图 1-1 RGB 彩色立方体

RGB 颜色空间的主要缺点是不直观,从 RGB 颜色空间中很难知道该值所表示颜色的认知属性,因此 RGB 颜色空间不符合人对颜色的感知心理。其次,RGB 颜色空间是最不均匀的颜色空间之一,两种颜色之间的知觉差异不能采用该颜色空间中两个颜色点之间的距离来表示。由于 RGB 颜色空间是一种与硬件设备相关的色彩空间,与人类视觉系统对颜色相似性的主观判断不符合,因此在图像分析过程中经常会将 RGB 颜色空间转换到符合人类视觉系统的颜色空间中。

2. HSV 颜色空间

RGB 颜色空间是一种基于硬件的颜色空间,而人类视觉对色彩的感知则没法用这一颜色空间来表示,因此在图像与视频分析的过程中,为了能够更好地描述颜色特征,往往会选择一种符合人类视觉系统对颜色感知的表示方法,常用的符合视觉感知的颜色空间主要有面向用户的 HSV (Hue, Saturation and Value) 颜色空间、HSB 颜色空间、HSI 颜色空间、Lab 颜色空间等。

HSV 颜色空间对应于画家配色模型,能够较好地反映视觉系统对颜色的感知和鉴别能力,其中 3 个分量分别代表色彩 (Hue)、饱和度 (Saturation) 和亮度 (Value)。

在画家配色模型中,画家用改变色泽和色深的方法来从某种纯色获得不同色调的颜色。其做法是:具有 $S=1$ 和 $V=1$ 的任何一种颜色相当于画家使用的纯颜色,在一种纯色中加入白色 (相当于降低 S 值,而 V 值不变) 以改变色泽,加入黑色 (相当于降低 V 值,而 S 值不变) 以改变色深,同时加入不同比例的白色、黑色 (同时降低 S 和 V) 即可得到不同色调的颜色 (图 1-2)。

对用户来说,HSV 颜色空间是一种直观的面向视觉感知的颜色表示,是直方图最常用的颜色空间。

HSV 颜色空间主要利用色调、饱和度和亮度 3 个要素来描述人眼的色彩感知。在这个模型中包含彩色信息的两个参数是色调 (H) 和饱和度 (S),图 1-3 给出了描述色调和饱和度这两个参数的六边形、圆形和三角形这 3 种色平面表示形式。在这 3 种平面中可以看到,原色是按 120° 分隔的,二次色与原色

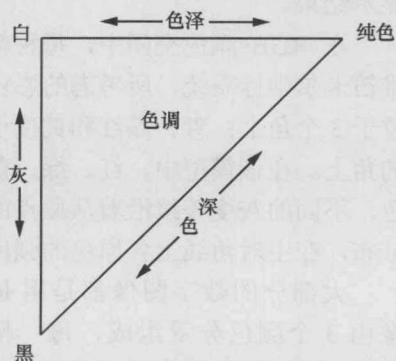


图 1-2 画家配色模型配色示意

相隔 60° ，也就是说二次色之间也相隔 120° 。一般假定 0° 的彩色为红色， 120° 的为绿色， 240° 的为蓝色。色调从 0° 变到 240° 覆盖了所有可见光谱的彩色。在 240° 到 360° 之间是人眼可见的非光谱色（紫色）（图 1-4 (c)）。设想从 RGB 沿立方体对角线的白色顶点向黑色顶点观察，就可以看到立方体的六边形外形（图 1-3 (a)）。六边形边界表示色彩，水平轴表示饱和度，从六边形中心到边的过渡表示饱和度逐渐增加。图 1-3 (b) 圆周上各点代表不同的色调，从圆心向圆周过渡表示饱和度逐渐增加。图 1-3 (c) 中从三角形的中心到三角形边的过渡表示饱和度逐渐增加，三角形边界表示色彩。

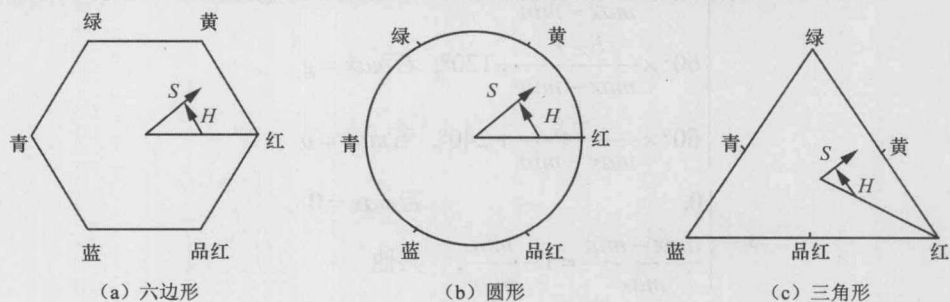


图 1-3 三种色平面表示

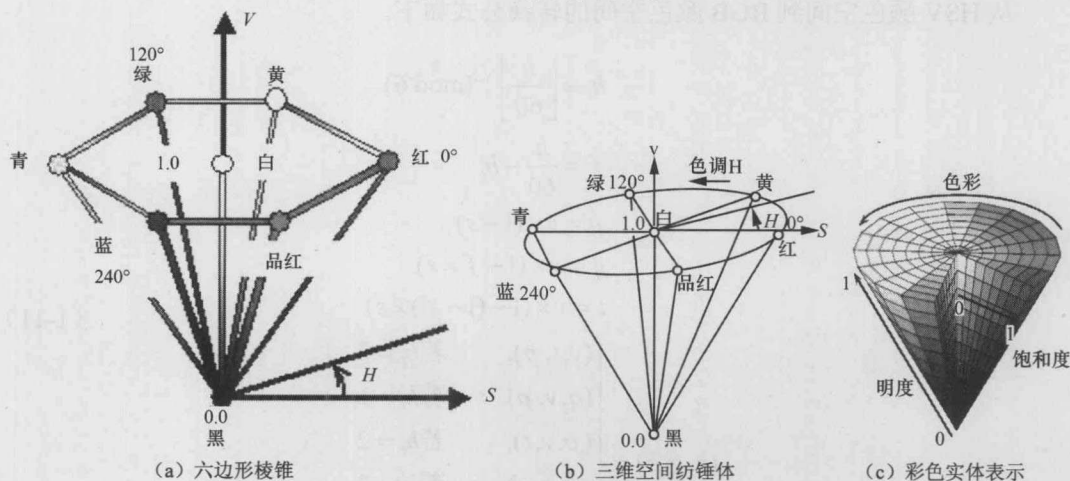


图 1-4 HSV 颜色空间示意图

饱和度参数给出一种纯色被白光稀释的程度的度量，是色平面的原点（圆心）到彩色点的半径的长度。在色平面的外围圆周是纯的或称饱和的颜色，其饱和度值为 1。在中心是中性（灰色）阴影，即饱和度为 0。因此，HSV 颜色空间实际上选择什么形状的色平面并没有关系，因为这些形状中的任何一个都可以通过几何变换转换为其他形式。

HSV 颜色空间的三维表示从 RGB 立方体演化而来。HSV 颜色空间对应于圆柱坐标系的一个圆锥形子集。在 HSV 颜色空间中，其明度（值）都是沿垂直轴测量（图 1-4）。明度沿六棱锥中心轴 V 从上至下变化。中心轴顶端，即圆锥的顶面呈白色（ $V=1$ ），它包含 RGB 模型中的 $R=1, G=1, B=1$ 3 个面，所代表的颜色较亮。中心轴底端，即圆锥的顶点（即原点）呈黑色（ $V=0$ ）。从中心轴的顶端点到原点代表亮度渐暗的灰色，即具有不同灰度的灰色。它们表示无彩色系的灰度颜色。因此可以说，HSV 颜色空间中的 V 轴对应于 RGB 颜色

空间中的主对角线。因此 RGB 颜色空间和 HSV 颜色空间之间存在着转换关系。

从 RGB 颜色空间到 HSV 颜色空间的转换公式如下。

$$h = \begin{cases} 0^\circ, & \text{若 } \max = \min \\ 60^\circ \times \frac{g-b}{\max-\min} + 0^\circ, & \text{若 } \max = r \text{ 且 } g \geq b \\ 60^\circ \times \frac{g-b}{\max-\min} + 360^\circ, & \text{若 } \max = r \text{ 且 } g < b \\ 60^\circ \times \frac{b-r}{\max-\min} + 120^\circ, & \text{若 } \max = g \\ 60^\circ \times \frac{r-g}{\max-\min} + 240^\circ, & \text{若 } \max = b \end{cases} \quad (1-10)$$

$$s = \begin{cases} 0, & \text{若 } \max = 0 \\ \frac{\max-\min}{\max} = 1 - \frac{\min}{\max}, & \text{其他} \end{cases}$$

$$v = \max$$

从 HSV 颜色空间到 RGB 颜色空间的转换公式如下。

$$h_i = \left\lfloor \frac{h}{60} \right\rfloor \pmod{6}$$

$$f = \frac{h}{60} - h_i$$

$$p = v \times (1 - s)$$

$$q = v \times (1 - f \times s)$$

$$t = v \times (1 - (1 - f) \times s)$$

$$(r, g, b) = \begin{cases} (v, t, p), & \text{若 } h_i = 0 \\ (q, v, p), & \text{若 } h_i = 1 \\ (p, v, t), & \text{若 } h_i = 2 \\ (p, q, v), & \text{若 } h_i = 3 \\ (t, p, v), & \text{若 } h_i = 4 \\ (v, p, q), & \text{若 } h_i = 5 \end{cases} \quad (1-11)$$

HSV 颜色空间有两个特点：(1) 亮度分量与图像的彩色信息无关；(2) 色调和饱和度分量与人感受颜色的方式是紧密相连的，这些特点使 HSV 颜色空间非常适合于借助人的视觉系统来感知彩色特性的图像处理算法。HSV 颜色空间直接对应于人眼色彩视觉特征的三要素，且通道之间各自独立，因此可以独立感知各颜色分量的变化。其中，色调尤其影响人的视觉判断，因此应用这种彩色模型会更适合用户的视觉判断。由于 HSV 颜色空间使得饱和度具有更大的动态取值范围，适合处理色度和饱和度，因此在图像分割、特征提取等领域都有重要的应用。

3. CMYK 颜色空间

彩色印刷或彩色打印的纸张是不能发射光线的，因而印刷机或彩色打印机就只能使

用一些能够吸收特定光波而反射其他光波的油墨或颜料。油墨或颜料的三基色是青 (Cyan)、品红 (Magenta) 和黄 (Yellow), 简称为 CMY。青色对应蓝绿色, 品红对应紫红色。理论上说, 任何一种由颜料表现的颜色都可以用这 3 种基色按不同的比例混合而成, 这种颜色表示方法称为 CMY 颜色空间表示法。彩色打印机和彩色印刷系统都采用 CMY 颜色空间。

用 CMY 模型产生的颜色被称为相减色, 是因为它减少了为视觉系统识别颜色所需要的反射光。在 CMY 相减混色中, 三基色等量相减时得到黑色; 等量黄色 (Y) 和品红 (M) 相减而青色 (C) 为 0 时, 得到红色 (R); 等量青色 (C) 和品红 (M) 相减而黄色 (Y) 为 0 时, 得到蓝色 (B); 等量黄色 (Y) 和青色 (C) 相减而品红 (M) 为 0 时, 得到绿色 (G)。

CMY 空间正好与 RGB 空间互补,即用白色减去 RGB 空间中的某一颜色值就等于同样颜色在 CMY 空间中的值。根据这个原理,很容易把 RGB 空间转换成 CMY 空间。由于彩色墨水和颜料的化学特性,用等量的 CMY 三基色得到的黑色不是真正的黑色,因此在印刷术中常加一种真正的黑色 (Black Ink),所以 CMY 又写成 CMYK。

4. YUV 颜色空间

在现代彩色电视系统中,通常采用三管彩色摄像机或彩色 CCD (电荷耦合器件) 摄像机,它把摄得的彩色图像信号,经分色,分别放大校正得到 RGB,再经过矩阵变换电路得到亮度信号 Y 和两个色差信号 $R-Y$ 、 $B-Y$,最后发送端将亮度和色差 3 个信号分别进行编码,用同一信道发送出去。这就是常用的 YUV 色彩空间。采用 YUV 色彩空间的重要性是它的亮度信号 Y 和色度信号 U 、 V 是分离的。如果只有 Y 信号分量而没有 U 、 V 分量,那么这样表示的图就是黑白灰度图。彩色电视采用 YUV 空间正是为了用亮度信号 Y 解决彩色电视机与黑白电视机的兼容问题,使黑白电视机也能接收彩色信号。

YUV 是 PAL 和 SECAM 模拟彩色电视制式采用的颜色空间, Y 表示亮度, UV 用来表示色差, U 、 V 是构成彩色的两个分量。

YUV 表示法的重要性在于它的亮度信号 (Y) 和色度信号 (U, V) 是相互独立的, 也就是 Y 信号分量构成的黑白灰度图与用 U, V 信号构成的两幅单色图是相互独立的。由于 Y, U 和 V 是独立的, 所以可以对这些单色图分别进行编码。此外, 黑白电视能接收彩色电视信号也就是利用了 YUV 分量之间的独立性。YUV 表示法的另一个优点是可以利用人眼的特性来降低数字彩色图像所需要的存储容量。人眼对彩色细节的分辨能力远比对亮度细节的分辨能力低。若把人眼刚刚能分辨出的黑白相间的条纹换成不同颜色的彩色条纹, 那么眼睛就不再能分辨出条纹来。由于这个原因, 就可以把彩色分量的分辨率降低而不明显影响图像的质量, 因而就可以把几个相邻像素不同的彩色值当作相同的彩色值来处理, 从而减少所需的存储容量。例如, 要存储 RGB 8:8:8 的彩色图像, 即 R, G 和 B 分量都用 8 位二进制数表示, 图像的大小为 640 像素×480 像素, 需要的存储容量为 921 600 byte。如果用 YUV 来表示同一幅彩色图像, Y 分量仍然为 640×480, 并且 Y 分量仍然用 8 bit 表示, 而对每 4 个相邻像素 (2×2) 的 U, V 值分别用相同的一个值表示, 那么存储同样的一幅图像所需的存储空间就减少到



460 800 byte。

根据美国国家电视制式委员会 NTSC 制式的标准,当白光的亮度用 Y 来表示时,它和红、绿、蓝三色光的关系可用如下式的方程描述。

$$Y = 0.3R + 0.59G + 0.11B \quad (1-12)$$

这就是常用的亮度公式。色差 U 、 V 是由 $B-Y$ 、 $R-Y$ 按不同比例压缩而成的。YUV 颜色空间与 RGB 颜色空间的转换关系如下。

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.3 & 0.59 & 0.11 \\ -0.15 & -0.29 & 0.44 \\ 0.61 & -0.52 & -0.096 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1-13)$$

如果要由 YUV 空间转化成 RGB 空间,只要进行相应的逆运算即可。

与 YUV 颜色空间类似的还有 Lab 颜色空间,它也是用亮度和色差来描述颜色分量,其中 L 为亮度、 a 和 b 分别为各色差分量。

YIQ 模型与 YUV 模型非常类似,它是在彩色电视制式中使用的另一种重要的颜色模型,在 NTSC 彩色电视制式中使用它。其中, Y 表示亮度, I 、 Q 是两个彩色分量。YIQ 和 RGB 的对应关系用下面的方程式表示。

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ I &= 0.596R - 0.275G - 0.321B \\ Q &= 0.212R - 0.523G + 0.311B \end{aligned} \quad (1-14)$$

或者写成如下的矩阵形式。

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1-15)$$

1.2 图像和视频的数字化

1.2.1 图像的数字化的

数字图像可以理解为对二维函数 $f(x, y)$ 进行抽样和量化(即离散处理)后得到的图像,因此,通常用二维矩阵来表示一幅数字图像。数字化过程包括 3 个步骤:扫描、抽样和量化。

抽样(Sampling)是对图像空间坐标的离散化,它决定了图像的空间分辨率。简单地讲,就是用一个网格(图 1-5)把待处理的图像覆盖,然后把每一小格上模拟图像的各个亮度取平均值,作为该小方格中点的值;或者把方格的交叉点处模拟图像的亮度值作为该方格交叉点上的值。这样,一幅模拟图像变成只用小方格中点的值来代表的离散值图像,或者只用方格交叉点的值表示的离散值图像。这个网格称为抽样网格,其意义是以网格为基础,采用某种形式抽取模拟图像代表点的值,即抽样。抽样后形成的图像称为数字图像。