

图像 / 视频检索 与图像融合

陈秀新 邢素霞 著



机械工业出版社
CHINA MACHINE PRESS



图像/视频检索与 图像融合

陈秀新 邢素霞 著

机械工业出版社

基于内容的图像/视频检索与图像融合是近年来多媒体信息处理领域的研究热点。本书对基于内容的图像/视频检索与图像融合技术进行了详细介绍。主要内容包括基于内容的图像/视频检索的产生、发展及目前的研究现状、常用图像和视频特征的提取及匹配方法；此外还有图像融合技术的发展、基于特征的图像配准技术、多传感器图像融合技术、彩色图像融合技术以及图像融合质量的评价方法。

本书可作为计算机科学与技术、电子信息工程等专业的研究生、高年级本科生的参考教材，也可作为从事图像/视频处理、计算机视觉等领域的研究人员的参考书。

图书在版编目 (CIP) 数据

图像/视频检索与图像融合/陈秀新, 邢素霞著. —北京:
机械工业出版社, 2011. 12
ISBN 978 - 7 - 111 - 36507 - 5

I. ①图… II. ①陈…②邢… III. ①数字图像处理
IV. ①TP391. 41

中国版本图书馆 CIP 数据核字 (2011) 第 238521 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 牛新国 责任编辑: 牛新国

版式设计: 常天培 责任校对: 陈秀丽

封面设计: 陈沛 责任印制: 杨曦

北京京丰印刷厂印刷

2012 年 1 月第 1 版·第 1 次印刷

169mm × 239mm · 14. 75 印张 · 292 千字

0 001—3 000 册

标准书号: ISBN 978 - 7 - 111 - 36507 - 5

定价: 38. 00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

电话服务

网络服务

社服务中心: (010) 88361066

门户网: <http://www.cmpbook.com>

销售一部: (010) 68326294

教材网: <http://www.cmpedu.com>

销售二部: (010) 88379649

读者购书热线: (010) 88379203

封面无防伪标均为盗版

前 言

随着电子技术和计算机网络技术的飞速发展，图像和视频等多媒体信息的获取和传播越来越方便快捷，基于图像和视频等多媒体内容的应用也越来越广泛，其中，基于内容的图像和视频检索对于在海量的图像和视频数据中检索到需要的图像和视频具有重要意义，而图像融合技术也在各种应用中具有重要的作用，这些都是当前数字图像和视频等多媒体信息处理领域的研究热点。

数字化后的媒体信息，尤其是数字化后的图像/视频和音频信息具有数据海量性，它给信息的管理、存储和传输造成极大的困难，成为阻碍人们有效获取和使用信息的主要瓶颈。按照传统的文本数据分析、管理、检索方法来管理图像/视频信息既要耗费大量的人力，又不能充分利用图像和视频中包含的丰富内容信息。因此，传统的文本信息内容描述、分析、管理、检索等相关技术已不适合处理大量图像/视频数据。基于图像/视频内容的数据分析、过滤、概要、检索等相关技术研究已受到广泛关注。为了对含有丰富信息的图像/视频数据进行高效快捷的过滤、浏览和检索，人们提出了基于内容的图像/视频检索技术。目前基于内容的图像/视频检索已经取得了一些成就，但仍然不能满足人们日益增长的检索需求，仍有很多问题需要研究和探讨。

图像融合是为了弥补单一传感器的不足应运而生的。图像融合技术最早于20世纪70年代后期提出，通过多源成像传感器获得同一目标或背景的图像信息，经过去噪、时间配准、空间配准和重采样后，再运用某种融合技术得到一幅合成图像，最终在显示设备上输出，以供进一步观测或进行决策和估计。多传感器所提供的信息和单一传感器所提供的数据相比，具有可靠性、冗余性、互补性。利用冗余信息可以改善信噪比，并可获得更可靠的结果；利用互补信息可获得更全面、更丰富、包含更多细节的信息。通过对多传感器图像的融合，可克服单一传感器图像在几何、光谱和空间分辨率等方面存在的局限性和差异性，提高图像的质量，从而有利于对物理现象和事件进行定位、识别和解释。因此，图像融合技术发展至今，一直受到国内外普遍关注，并在实际应用中取得了明显效果，其应用范围已遍及军事、遥感、天气预报、医学影像、安全监控等领域。

本书是一本基于内容的图像/视频检索和图像融合领域的学术专著，重点介绍基于内容的图像/视频检索的基础知识、典型方法、几种有效的检索方法

以及图像融合的相关知识和方法。本书第1篇介绍数字图像处理的基础知识,包括数字图像处理概述、常用方法、相关研究领域、常用图像数据集以及OpenCV简介等。第2篇介绍基于内容的图像检索的相关知识与方法,包括基于内容的图像检索概述、典型的方法(基于颜色特征的方法、基于纹理特征的方法、基于形状特征的方法、基于空间关系的方法、基于局部不变特征的方法以及其他方法等),并详细介绍了两种有效的图像检索方法。第3篇介绍基于内容的视频检索技术,包括基于内容的视频检索概述、视频结构化分析方法(镜头分割、关键帧提取、场景和镜头组的构造以及视频特征提取),并介绍了一种有效的视频检索方法以及视频语义分析与提取技术。第4篇介绍图像配准与图像融合技术,包括多传感器图像融合技术的发展、多传感器图像配准技术、多传感器图像融合技术、彩色图像融合技术以及图像融合质量的评价方法。

本书第1~9章由陈秀新编写,第10~14章由邢素霞编写。本书是在北京工商大学的北京市教委科技创新平台—信号检测与智能信息处理研究平台项目(编号:201151)资助下完成的。本书内容是在北京市教委科技规划项目(编号:KM20101001002)和北京市优秀人才培养资助项目(编号2009D005003000006)的研究基础上整理完成。在此一并感谢。本书在编写过程中得到了廉小亲教授的亲切关心和支持,在此表示诚挚的谢意!此外,本书第8章的部分内容采用了北京工业大学电子信息与控制工程学院数字多媒体信息处理及通信系统研究室贾克斌教授的研究成果,在此声明并表示衷心的感谢!另外,本书得到了南京理工大学常本康教授的帮助和支持,在此表示衷心的感谢!感谢北京工商大学计算机与信息工程学院电子信息系全体老师的关心与帮助!

由于作者水平、时间和精力有限,本书还有不尽如人意的地方,尽管我们做了很大的努力,但缺点、错误在所难免,希望得到读者的批评指正。

陈秀新 邢素霞
2011年9月

目 录

前言

第1篇 基础知识

第1章 绪论	1
1.1 数字图像处理概述	1
1.2 数字图像处理常用方法	2
1.3 数字图像处理相关研究领域	3
1.4 常用图像数据集	5
1.4.1 Corel 图像集	5
1.4.2 Caltech 101 图像集	5
1.4.3 Caltech 256 图像集	7
1.4.4 COIL 图像集	7
1.4.5 其他图像集	9
1.5 OpenCV 简介	9
参考文献	11

第2篇 基于内容的图像检索

第2章 基于内容的图像检索概述	12
2.1 基于内容的图像检索的产生与发展	12
2.2 图像检索系统的构成	16
2.3 图像检索系统的评价标准	16
参考文献	18
第3章 图像检索典型方法	19
3.1 基于颜色特征的图像检索	19
3.1.1 图像的颜色空间	19
3.1.2 典型的图像颜色特征	23
3.1.3 颜色直方图匹配	25
3.2 基于纹理特征的图像检索	26
3.2.1 灰度共生矩阵	27
3.2.2 Tamura 纹理特征	28
3.2.3 小波变换	29
3.3 基于形状特征的图像检索	31

3.3.1 边缘方向直方图	32
3.3.2 傅里叶形状描述符	33
3.3.3 Hu 不变矩	33
3.3.4 Jan Flusser 不变矩	34
3.4 基于空间关系的图像检索	35
3.5 基于局部不变特征的图像检索	35
3.5.1 Moravec's 角点检测子	38
3.5.2 Harris 角点检测子	38
3.5.3 SUSAN 算子	42
3.5.4 SIFT 算子	44
3.5.5 MSER 区域检测子	49
3.5.6 局部不变特征的性质	51
3.6 其他图像检索方法	52
参考文献	52
第4章 基于连通区域的图像仿射不变区域提取及描述	54
4.1 基于连通区域的仿射不变区域提取	55
4.1.1 连通区域提取方法及其实现	55
4.1.2 仿射不变区域的确定	56
4.1.3 去除不满足条件的区域	56
4.2 图像仿射不变区域的描述	57
4.3 实验结果与分析	58
参考文献	60
第5章 基于三维颜色直方图的图像检索	61
5.1 颜色空间量化	61
5.1.1 图像颜色量化理论	61
5.1.2 颜色矢量量化方案的选择	62
5.2 提取三维颜色直方图	65
5.3 颜色直方图匹配	66
5.4 实验结果与分析	66
参考文献	68
第3篇 基于内容的视频检索	
第6章 基于内容的视频检索概述	69
6.1 基于内容的视频检索的产生与发展	69
6.2 视频检索系统的构成	71
6.3 TRECVID 简介	71
6.4 视频检索系统评价标准	72
参考文献	72

第7章 视频结构化分析	73
7.1 镜头分割	74
7.1.1 基于直方图的镜头分割方法	74
7.1.2 像素法	75
7.1.3 基于边缘探测的算法	76
7.1.4 基于模型的镜头分割方法	77
7.1.5 面向压缩视频的镜头分割方法	78
7.2 关键帧提取	80
7.2.1 基于采样的关键帧提取方法	81
7.2.2 基于帧间差的关键帧提取方法	81
7.2.3 基于聚类的关键帧提取方法	81
7.2.4 基于视频单元分类的关键帧提取方法	81
7.2.5 基于累积帧间差的关键帧提取方法	82
7.2.6 基于运动信息的关键帧提取方法	82
7.2.7 基于文字和图像信息的关键帧提取方法	82
7.3 场景和镜头组的构造	83
7.4 视频特征提取	83
7.4.1 MPEG-7 标准中的视觉特征描述子	84
7.4.2 关键帧特征提取	89
7.4.3 视频特征提取	91
参考文献	106
第8章 综合时序特征曲线和关键帧特征的视频检索	109
8.1 视频时序特征的提取与匹配	110
8.1.1 视频特征曲线的提取	111
8.1.2 视频特征曲线的匹配	112
8.1.3 实验结果及分析	113
8.2 关键帧特征的提取与匹配	114
8.2.1 子片段划分与关键帧提取	114
8.2.2 关键帧特征提取和匹配	114
8.3 实验结果与分析	115
参考文献	116
第9章 视频语义分析与提取	117
9.1 视频语义分析概述	117
9.1.1 视频语义分析的意义	117
9.1.2 视频语义分析发展现状	119
9.2 特定领域的视频语义分析	120
9.2.1 智能监控	120
9.2.2 视频标注	124

9.2.3 交通事件检测 126
参考文献 128

第4篇 图像配准与图像融合

第10章 多传感器图像融合发展与概况 130

10.1 多传感器图像融合技术 130

10.2 多传感器图像融合步骤 134

10.3 多传感器图像融合的应用与发展趋势 135

参考文献 139

第11章 多传感器图像配准 141

11.1 图像配准原理 142

11.1.1 变换空间 142

11.1.2 图像配准过程 143

11.2 基于图像灰度的配准方法 144

11.2.1 互相关法 144

11.2.2 序贯相似性检测算法 145

11.3 基于变换域的图像配准方法 146

11.4 基于特征的图像配准方法 147

11.4.1 特征选择与提取 149

11.4.2 特征匹配 152

11.5 基于SIFT算法的图像配准 154

11.5.1 检测尺度空间极值点 154

11.5.2 精确定位关键点 156

11.5.3 关键点方向分配 156

11.5.4 特征点描述子生成 157

11.5.5 SIFT算法改进及其结果 158

参考文献 161

第12章 多传感器图像融合 165

12.1 多传感器图像融合规则 165

12.1.1 基于像素的融合规则 165

12.1.2 基于区域的融合规则 166

12.2 基于金字塔的图像融合算法 167

12.2.1 基于拉普拉斯金字塔的图像融合 167

12.2.2 基于对比度金字塔分解的图像融合 169

12.2.3 实验结果与分析 170

12.3 基于小波变换的图像融合算法 174

12.3.1 基于离散小波变换的图像融合方法 174

12.3.2 图像的小波分解层数 176

12.3.3 实验结果与分析	176
12.4 基于 NSCT 变换的图像融合	183
12.4.1 Contourlet 变换	183
12.4.2 NSCT 变换	186
12.4.3 基于 NSCT 的图像融合	189
12.4.4 实验结果与分析	189
参考文献	191
第 13 章 彩色图像融合技术	194
13.1 典型的色彩空间	194
13.1.1 RGB 颜色空间	194
13.1.2 HSI 颜色空间	195
13.1.3 YUV (Lab) 颜色空间	196
13.1.4 YCbCr 模型	197
13.1.5 YIQ 模型	197
13.1.6 CIE 色度模型	198
13.1.7 CMY 色度空间	199
13.2 彩色图像融合技术的发展	200
13.3 经典彩色图像融合方法	202
13.3.1 直接映射法	202
13.3.2 MIT 融合法	202
13.3.3 TNO 融合法	203
13.3.4 基于 $\alpha\beta$ 空间色彩传递的图像融合方法	204
13.3.5 基于 $YCbCr$ 空间色彩传递的图像融合方法	206
13.4 实验结果与分析	207
参考文献	208
第 14 章 多传感器图像融合质量评价	210
14.1 主观评价方法	210
14.2 基于单幅图像的评价方法	211
14.2.1 灰度均值和标准差	211
14.2.2 信息熵	212
14.2.3 平均梯度	212
14.2.4 边缘保持度客观评价指标	212
14.3 基于参考图像的评价方法	212
14.3.1 联合熵	213
14.3.2 交互信息量	213
14.3.3 空间频率	214
14.3.4 偏差度	214
14.3.5 对比度	214

14.3.6 相关系数	214
14.4 基于结构相似度的评价方法	215
14.5 基于视觉特性的评价方法	216
14.5.1 基于视觉感知的评价方法	216
14.5.2 基于视觉兴趣的评价方法	217
14.6 实验与结果分析	217
14.6.1 金字塔融合方法评价与讨论	217
14.6.2 小波分解层数评价与分析	219
14.6.3 基于区域的融合规则质量评价与讨论	220
14.6.4 NSCT 变化融合质量评价与讨论	221
参考文献	223

第 1 篇 基础知识

第 1 章 绪 论

1.1 数字图像处理概述

随着图像获取技术的快速发展,数字图像在人们的生活中起着越来越重要的作用,数字图像处理技术已经深入到我们生活中的方方面面。数字图像处理的产生和迅速发展主要受三个因素的影响:一是计算机的发展,二是数学的发展(特别是离散数学理论的创立和完善),三是广泛的航空航天、林业、环境、军事、工业、医学以及人们生活等各方面的应用需求的增长。

数字图像处理(Digital Image Processing)是指将图像信号转换成数字信号并利用计算机等设备对其进行处理的过程。数字图像处理最早出现于20世纪50年代,当时的电子计算机已经发展到一定水平,人们开始利用计算机来处理图形和图像信息。数字图像处理作为一门学科大约形成于20世纪60年代初期。早期的图像处理的目的是改善图像的质量,它以人为对象,以改善人的视觉效果为目的。随着数字图像处理技术的发展以及实际需求的不断增长,图像处理技术已不仅仅是改善图像的质量。目前,常用的图像处理方法有图像增强、图像分割、图像变换、图像复原、图像编码压缩等。

图像处理技术在许多应用领域受到广泛重视并取得了重大的开拓性成就,这些领域包括航空航天、生物医学工程、工业检测、机器人视觉、公安司法、军事制导、文化艺术等,使图像处理成为一门引人注目、前景远大的新型学科。从20世纪70年代中期开始,随着计算机技术、人工智能和思维科学研究的迅速发展,数字图像处理向更高、更深层次发展。人们已开始研究如何用计算机系统解释图像,实现类似人类视觉的系统理解外部世界,这被称为图像理解或计算机视觉。很多国家,特别是发达国家投入许多的人力、物力到这个研究领域,取得了不少重要的研究成果。其中代表性的成果是20世纪70年代末美国麻省理工学院(MIT)的Marr提出的视觉计算理论,这个理论成为计算机视觉领域其后10多年的主导思想。图像理解虽然在理论方法研究上已取得不小的进展,但它本身是一个比较难的研究领

域,存在不少困难,由于人类本身对自己的视觉过程的了解甚少,因此计算机视觉是一个有待人们进一步探索的新领域。

1.2 数字图像处理常用方法

1. 图像增强和复原

图像增强的目的是为了提髙图像的质量,如去除噪声、提高图像的清晰度等。图像增强不考虑图像降质的原因,为的是突出图像中感兴趣的部分。如强化图像高频分量,可使图像中物体轮廓清晰,细节明显;如强化低频分量可减少图像中噪声影响。图像复原要求对图像降质的原因有一定的了解,一般讲应根据降质过程建立“降质模型”,再采用某种滤波方法,恢复或重建原来的图像。

2. 图像变换

由于图像阵列很大,直接在空间域中进行处理,涉及的计算量很大。因此,往往采用各种图像变换的方法,如傅里叶变换、沃尔什变换、离散余弦变换等间接处理技术,将空间域的处理转换为变换域的处理,不仅可以减少计算量,而且可更有效地处理(如傅里叶变换可在频域中进行数字滤波处理)。目前,由于小波变换在时域和频域中都具有良好的局部化特性,它在图像处理中也有着广泛而有效的应用。

3. 图像编码压缩

图像编码压缩技术可以减少描述图像的数据量(即比特数),以便节省图像传输、处理时间并减少所占用的存储器容量。压缩可以在不失真的前提下获得,也可以在允许的失真条件下进行。编码是压缩技术中最重要方法,它在图像处理技术中是发展最早且比较成熟的技术。图像编码压缩技术在图像的传输中具有重要意义。

4. 图像分割

图像分割是数字图像处理中的关键技术之一。图像分割是将图像中有意义的特征部分提取出来,有意义的特征包括图像中的边缘、区域等,这是进一步进行图像识别、分析和理解的基础。虽然目前已研究出不少边缘提取、区域分割的方法,但还没有一种普遍适用于各种图像的有效方法。因此,对图像分割的研究还在不断深入之中,图像分割是目前图像处理中研究的热点之一。

5. 图像描述

图像描述是图像识别和理解的必要前提。最简单的二值图像可采用其几何特性描述物体的特性,一般图像的描述方法采用二维形状描述,它有边界描述和区域描述两类方法。对于特殊的纹理图像可采用二维纹理特征描述。随着数字图像处理研究的深入发展,已经开始进行三维物体描述的研究了,提出了体积描述、表面描述、广义圆柱体描述等方法。可用于图像描述的特征可以包括很多方面,如频域特

征、灰度或颜色特征、边界特征、区域特征、纹理特征、形状特征、拓扑特征和关系结构等。

6. 图像分类（识别）

图像分类（识别）属于模式识别的范畴，其主要内容是图像经过某些预处理（增强、复原、压缩）后，进行图像分割和特征提取，从而进行判决分类。图像分类常采用经典的模式识别方法，有统计模式分类和句法（结构）模式分类，近年来新发展起来的模糊模式识别和人工神经网络模式分类在图像识别中也越来越受到重视。

1.3 数字图像处理相关研究领域

图像是人类获取和交换信息的主要手段，因此，图像处理的应用领域必然涉及人类生活和工作的方方面面。随着人类活动范围的不断扩大，图像处理的应用领域也随之不断扩大。

1. 航天和航空技术方面

数字图像处理技术在航天和航空技术方面的应用，很大程度体现在飞机遥感和卫星遥感技术中。许多国家每天派出很多侦察飞机或者通过卫星，对地球上感兴趣的地区进行大量的空中摄影。对由此得来的照片进行处理分析，以前需要雇用几千人，而现在改用配备有高级计算机的图像处理系统来判读分析，既节省人力，又加快了速度，还可以从照片中提取人工所不能发现的大量有用情报。现在世界各国都在利用陆地卫星所获取的图像进行资源调查（如森林调查、海洋泥沙和渔业调查、水资源调查等）、灾害检测（如病虫害检测、水火检测、环境污染检测等）、资源勘察（如石油勘察、矿产量探测、大型工程地理位置勘探分析等）、农业规划（如土壤营养、水分和农作物生长、产量的估算等）和城市规划（如地质结构、水源及环境分析等）。我国也陆续开展了以上诸方面的一些实际应用，并获得了良好的效果。此外，在气象预报和对太空其他星球的研究方面，数字图像处理技术也发挥了相当大的作用。

2. 生物医学工程方面

数字图像处理在生物医学工程方面的应用十分广泛，而且很有成效。除了常用的CT处理技术之外，还有对医用显微图像的处理分析，如红细胞、白细胞分类，染色体分析，癌细胞识别等。此外，在X射线肺部图像增晰、超声波图像处理、心电图分析、立体定向放射治疗等医学诊断方面都在广泛地应用图像处理技术。

3. 通信工程方面

当前通信的主要发展方向是声音、文字、图像和数据结合的多媒体通信。具体地讲是将电话、电视和计算机以三网合一的方式在数字通信网上传输。其中以图像

通信最为复杂和困难，因为图像的数据量十分巨大，如传送彩色电视信号的速率达100Mbit/s以上。要将这样高速率的数据实时传送出去，必须采用编码技术来压缩信息的比特量。在一定意义上讲，编码压缩是这些技术成败的关键。除了已应用较广泛的熵编码、DPCM (Differential Pulse Code Modulation, 差分脉冲编码调制) 编码、变换编码外，目前国内外正在大力开发研究新的编码方法，如分行编码、自适应网络编码、小波变换图像压缩编码等。

4. 工业和工程方面

在工业和工程领域中图像处理技术有着广泛的应用，如自动装配线中检测零件的质量并对零件进行分类，印制电路板疵病检查，弹性力学照片的应力分析，流体力学图片的阻力和升力分析，邮政信件的自动分拣，在一些有毒、放射性环境内识别工件及物体的形状和排列状态，在先进的设计和制造技术中采用工业视觉等。其中值得一提的是研制具备视觉、听觉和触觉功能的智能机器人，它将会给工农业生产带来新的激励，目前它已在工业生产中的喷漆、焊接、装配中得到有效的利用。

5. 军事公安方面

在军事方面，图像处理和识别主要用于导弹的精确末制导，各种侦察照片的判读，具有图像传输、存储和显示的军事自动化指挥系统，飞机、坦克和军舰模拟训练系统等。在公安方面，图像处理和识别主要用于公安业务图片的判读分析、指纹识别、人脸鉴别、不完整图片的复原，以及交通监控、事故分析等。目前已投入运行的高速公路不停车自动收费系统中的车辆和车牌的自动识别都是图像处理技术成功应用的例子。

6. 文化艺术方面

目前这类应用有电视画面的数字编辑，动画的制作，电子图像游戏，纺织工艺品设计，服装设计与管理，发型设计，文物资料照片的复制和修复，运动员动作分析和评分等，现在已逐渐形成一门新的艺术——计算机美术。

7. 机器人视觉

机器视觉作为智能机器人的重要感觉器官，主要进行三维景物理解和识别，是目前处于研究之中的开放课题。机器视觉主要用于军事侦察、危险环境的自主机器人，邮政、医院和家庭服务的智能机器人，装配线工件识别、定位，太空机器人的自动操作等。

8. 视频和多媒体系统

目前，电视制作系统广泛使用图像处理、变换、合成，多媒体系统中静止图像和动态图像的采集、压缩、处理、存储和传输等也都基于数字图像处理技术。

9. 科学可视化

图像处理和图形学紧密结合，形成了科学研究各个领域新型的研究工具。

10. 电子商务

在当前呼声甚高的电子商务中,图像处理技术也大有可为,如身份认证、产品防伪、水印技术等。

总之,图像处理技术应用领域相当广泛,正在国家安全、经济发展、日常生活中充当越来越重要的角色,对国计民生的作用不可低估。

1.4 常用图像数据集

随着数字图像处理技术的发展,各种用于图像处理算法实验及研究的图像库/图像集应运而生,这就使得各种方法有了统一的分析对象和数据源,便于不同的理论方法优劣性的比较。本节将对现有的各种图像库进行简单介绍,便于读者根据自己的需要选择合适的图像集。

1.4.1 Corel 图像集

Corel 是不同类型图像的图像集,广泛应用于基于内容的图像检索领域。Corel 图像集有近 14 000 幅图像,共 10 类,不同的研究者通常选择使用其不同子集进行实验。图 1-1 所示为 Corel 图像集的几个实例图像。

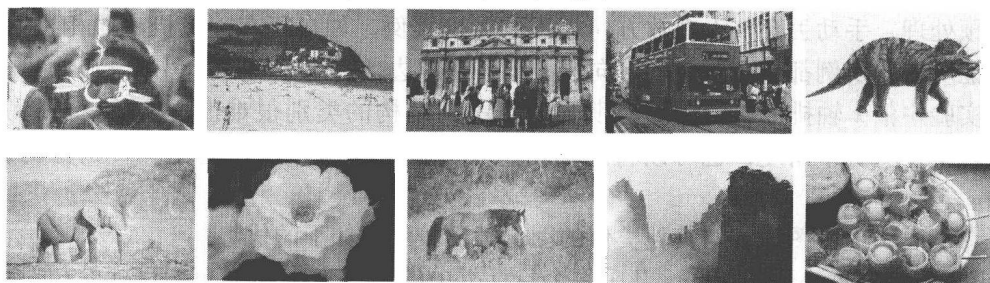


图 1-1 Corel 图像集实例图像

另外也有一些学者从 Corel 图像集中选择一些图像组成小的图像库,如宾夕法尼亚大学的 James Z. Wang 教授所在的研究所以及 Deselaers 等。

1.4.2 Caltech 101 图像集

Caltech 101 在 2003 年 9 月份由 California Institute of Technology 的 Fei-Fei Li、Marco Andreetto、Marc'Aurelio Ranzato 和 Pietro Perona 创建,可用于计算机视觉的研究,如图像识别、分类等。Caltech 101 包括 9 146 幅图像,分为 101 类前景物体(包括脸、手表、蚂蚁、钢琴等)和一个背景杂乱类别,总共 102 类。此外,Caltech101 图像库对每幅图像都进行了注释,每个注释包括两种信息,一种是目标

位置的边界盒，为黄色，一种是人工描绘的目标轮廓，为红色。另外，它还提供了查看注释的 Matlab 脚本。Caltech 101 图像库中的每个目标类别平均包括 40 ~ 800 幅图像，常用的类别图像较多，如人脸等。每幅图像的大小均为大约 300×200 个像素。图 1-2 所示为 Caltech 101 图像库中的图像实例。



图 1-2 Caltech 101 图像集中的图像实例

该图像集通过 Google 图像搜索引擎收集得到，同时，这些图像都经过了较小的预处理，手动丢弃与大多数方向相反的目标实例，如自行车、飞机、照相机等，保证所有的实例都具有相同的方向，如车头在左边，车尾在右边。另外，因为目标在实验中依 x 轴排序，所以对有明显垂直纵向结构的类别很难进行实验，可将其旋转至任意一个角度，也可以将 y 轴信息作为排序参考，从而避免旋转图像的影响。

Caltech 101 图像集与其他类似的图像集比较具有如下的优势：

1) 图像大小近似相同，几乎每个目标类中的图像大小和感兴趣目标的相对位置都是相同的，这意味着不需要花时间去裁剪图像就能够进行试验；

2) 图像的杂乱或遮挡部分比较少，甚至没有，与识别相关的算法可以存储待识别目标的唯一特征；

3) 对目标轮廓的细节进行了注释是该数据集的另一个优势。

尽管如此，Caltech 101 也存在如下问题：

1) 真实世界实际上拥有近 10 000 个目标类别，但 Caltech 101 却只能表示其中的一小部分，一些类别可能没有图像，这对识别来说图像集不够充分；

2) 由于在实际情况中，通常在目标相对位置和感兴趣目标方向上有更多的杂乱、遮挡和变化，而该图像集中的图像过于简单，表达方式比较单一，因此这些图像不总能代表实际的输入图像。

尽管如此，Caltech 101 图像集仍然是评判多类识别算法的客观实际标准。