

新型显示技术及应用集成系列丛书

张兆杨 杨高波 刘 志 等 著

# 视频对象分割提取的原理与应用



新型显示技术及应用集成系列丛书

# 视频对象分割提取的 原理与应用

科学出版社

北 京

## 内 容 简 介

本书主要论述视频对象分割提取的原理,像素域和压缩域视频对象分割的各种方法,以及视频对象分割提取在多个领域中的典型应用。主要内容包括:时域、空域和时空融合分割的基本原理及视频对象分割性能的客观评价准则;以分割准确度为应用目标的像素域视频对象的分割提取,重点论述出现多个视频对象、发生相互遮挡时的分割方法和基于智能剪的交互式视频对象分割及其自动跟踪方法;以实时性为应用目标的像素域和压缩域视频对象的分割提取,前者主要讨论基于细胞神经网络(CNN)开发工具的实时分割方法,后者阐述基于 H. 264 和 MPEG 压缩域与视觉关注度的对象分割方法;以及视频对象分割提取在各个方面的典型应用等。

本书可作为通信、信息、计算机、传媒类等研究机构和企事业单位的研究人员从事研究和应用的参考书,也可作为上述专业的高年级本科生、硕士生、博士生和教师用于教学与科研的参考书。

### 图书在版编目(CIP)数据

视频对象分割提取的原理与应用/张兆杨等著. —北京:科学出版社,2009  
(新型显示技术及应用集成系列丛书)

ISBN 978-7-03-024185-6

I. 视… II. 张… III. 视频信号-信号处理 IV. TN941.1

中国版本图书馆 CIP 数据核字(2009)第 029108 号

责任编辑:刘宝莉 汤 枫 / 责任校对:钟 洋

责任印制:赵 博 / 封面设计:嘉华永盛

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>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2009 年 3 月第 一 版 开本: B5(720×1000)

2009 年 3 月第一次印刷 印张: 14

印数: 1—3 000 字数: 270 000

定价: 40.00 元

(如有印装质量问题, 我社负责调换〈双青〉)

# 前 言

人们对多媒体信息的需求不断增长,其中视频信息是最重要的部分。为便于视频信息的传输和存储,需要进行高效的压缩编码。自基于视频对象分割提取和交互为特点的第二代压缩标准 MPEG-4 公布以来,视频对象的分割提取不仅可用于显著提高编码效率,还在基于对象的视频索引、视频检索、智能视频监控系统、可视化通信、虚拟现实、医学和遥感及军事应用,以及工农业生产中的监控检测和对影视资料的损伤修复与交互式电视、拟人化电子游戏等领域得到越来越多的应用。

视频对象分割提取的理论、方法和技术涉及数学、物理学、生理/心理学、电子学和计算机科学等多个学科的交叉,正在发展成为一门综合学科。目前在高级语义性对象分割的理论上仍有很大的发展空间,在面向应用上尚需进一步提高其实用性、实时性和系统性。

本书是在“基于压缩域的二维/三维视频对象提取(60572127)”、“基于视觉关注度的语义对象分割(60602012)”和“任意形状视频对象的自动提取及其形状编码(60172020)”三项国家自然科学基金项目资助下所积累的研究成果的基础上完成的,并融入了国内外的一些最新研究成果。本书内容上既包含视频对象分割提取的基本理论,又在此基础上既广又深地给出了像素域和压缩域视频对象分割提取的机理、方法和具体的算法步骤以及实验环境、工具、过程与结果,便于读者根据各自的应用要求方便地实现和进一步深入扩展。

全书分四大部分。第一部分包括第 1 章和第 2 章,阐述了视频对象分割提取的概念、发展概况和基本理论。第二部分包括第 3 章和第 4 章,主要针对像素域视频素材的,前者以提高对象分割的准确度为应用目标,重点讨论时空信息的融合分割方法、存在背景遮挡和前景对象互遮挡时的分割方法以及以人机交互方式进一步提高分割精度和对后续帧对象进行自动跟踪的各种方案;后者以提高对象分割的实时性为应用目标,重点是以细胞神经网络(CNN)开发工具和各类 CNN 模板实现视频对象的实时分割。第三部分是针对压缩域视频素材的第 5 章,讨论了从已经过 MPEG 或 H. 264 标准压缩后的视频码流中分割视频对象的方法和算法。由于码流中已有分割前景运动对象所需的运动矢量场,无需十分耗时的运动估计而具有快速实时分割的优势。但该矢量场是稀疏的,因此重点是讨论如何通过多帧累加和内插使其稠密化的基础上再给出各种有代表性的对象分割方案。第四部分即第 6 章,介绍了视频对象分割的多个典型应用案例。

本书由张兆杨教授拟订全书的大纲和撰写第 1 章,并对全书进行统稿、修改和

定稿;由杨高波教授撰写第2~4章;由刘志副教授和韩军副教授分别执笔第5章和第6章,其中5.3节、5.4节部分的内容和6.5节的内容分别由陆宇博士和石旭利副教授撰写。此外,本书在著述过程中得到了上海市经济和信息化委员会、上海大学通信学院和教育部新型显示技术与应用重点实验室的大力支持,在此表示衷心的感谢。

视频对象分割提取是尚在发展中的新技术,加之限于作者水平,本书难免存在不妥之处,谨请读者指正。

张兆杨

2008年11月于上海大学

# 目 录

## 前言

|                      |    |
|----------------------|----|
| 第 1 章 视频对象分割提取概述     | 1  |
| 1.1 视频对象分割提取的基本概念    | 1  |
| 1.1.1 视频对象的定义        | 1  |
| 1.1.2 视频分割与图像分割的关系   | 2  |
| 1.2 视频对象分割方法的分类与应用概况 | 2  |
| 1.2.1 分割方法的分类        | 2  |
| 1.2.2 应用概况           | 4  |
| 1.3 本书的结构            | 4  |
| 参考文献                 | 5  |
| 第 2 章 像素域视频对象分割基础    | 7  |
| 2.1 图像的预处理和后处理技术     | 7  |
| 2.1.1 图像的滤波处理        | 7  |
| 2.1.2 梯度算子           | 8  |
| 2.1.3 数学形态学预/后处理     | 9  |
| 2.2 基于空间域的分割         | 15 |
| 2.2.1 基于区域的分割        | 15 |
| 2.2.2 基于像素聚类的分割      | 19 |
| 2.2.3 基于分水岭变换的分割     | 20 |
| 2.3 基于时间域的分割         | 21 |
| 2.3.1 光流场法           | 22 |
| 2.3.2 帧差法            | 23 |
| 2.3.3 矢量场估计法         | 23 |
| 2.4 基于时空融合的对象分割      | 26 |
| 2.4.1 基于时空聚类的分割方法    | 26 |
| 2.4.2 基于光流的运动对象分割方法  | 27 |
| 2.4.3 基于对象跟踪的分割方法    | 29 |
| 2.5 视频对象分割的性能评价      | 32 |
| 2.5.1 空间准确度评价        | 34 |

|                                        |            |
|----------------------------------------|------------|
| 2.5.2 时间一致性评价 .....                    | 36         |
| 参考文献 .....                             | 36         |
| <b>第3章 以分割准确度为目标的像素域视频对象分割方法 .....</b> | <b>38</b>  |
| 3.1 基于背景记录和重建的 VO 自动分割 .....           | 38         |
| 3.1.1 基于背景记录 and 变化检测的 VO 分割 .....     | 38         |
| 3.1.2 基于背景重建的 VO 提取 .....              | 44         |
| 3.2 时空融合 VO 分割的典型方法 .....              | 52         |
| 3.2.1 融入时域信息的分水岭 VO 分割的方案组成 .....      | 52         |
| 3.2.2 时间分割 .....                       | 53         |
| 3.2.3 空间分割 .....                       | 55         |
| 3.2.4 时/空融合分割 .....                    | 57         |
| 3.2.5 实验结果 .....                       | 58         |
| 3.3 存在多个视频对象时的分割方法 .....               | 60         |
| 3.3.1 基于贝叶斯估计的多视频对象分割 .....            | 60         |
| 3.3.2 时空曲线演化的多个 VO 的分割 .....           | 73         |
| 3.4 用户辅助的交互式视频对象分割 .....               | 86         |
| 3.4.1 智能剪及其改进方法的视频对象分割与跟踪 .....        | 89         |
| 3.4.2 基于种子区域合并的交互式视频对象分割 .....         | 105        |
| 参考文献 .....                             | 112        |
| <b>第4章 以实时应用为目标的像素域视频对象分割方法 .....</b>  | <b>116</b> |
| 4.1 细胞神经网络基础 .....                     | 116        |
| 4.1.1 细胞神经网络模型及其特点和结构 .....            | 117        |
| 4.1.2 细胞神经网络的开发工具 .....                | 119        |
| 4.1.3 CNN 模板的设计方法简介 .....              | 122        |
| 4.1.4 CNN 在图像和视频处理方面的研究现状 .....        | 123        |
| 4.2 适合头肩序列的基于 CNN 模板的 VO 分割 .....      | 124        |
| 4.2.1 头肩序列的特点 .....                    | 124        |
| 4.2.2 视频对象分割算法 .....                   | 125        |
| 4.2.3 分割算法的 CNN 实现及实验结果 .....          | 127        |
| 4.3 人脸提取算法 CNN 实现 .....                | 133        |
| 4.3.1 算法概述 .....                       | 133        |
| 4.3.2 模板结构 .....                       | 136        |
| 4.3.3 实验结果 .....                       | 141        |
| 4.4 基于光流和改进分水岭分割算法的 CNN 实现 .....       | 143        |

---

|                                   |            |
|-----------------------------------|------------|
| 4.4.1 以 CNN 实现算法的方案 .....         | 143        |
| 4.4.2 CNN 模板设计 .....              | 145        |
| 4.4.3 实验结果 .....                  | 148        |
| 参考文献 .....                        | 149        |
| <b>第 5 章 压缩域视频对象分割</b> .....      | <b>153</b> |
| 5.1 基于压缩域视频对象分割的基本思路 .....        | 153        |
| 5.2 基于 H.264 压缩域的视频对象分割方法 .....   | 154        |
| 5.2.1 运动矢量场归一化和累积 .....           | 155        |
| 5.2.2 全局运动补偿 .....                | 157        |
| 5.2.3 累积运动矢量场分割 .....             | 157        |
| 5.2.4 基于匹配矩阵的时空分割 .....           | 158        |
| 5.2.5 实验结果 .....                  | 164        |
| 5.3 基于 MPEG 压缩域的视频对象分割方法 .....    | 167        |
| 5.3.1 提取 DC+2AC 图的轮廓特征 .....      | 167        |
| 5.3.2 基于运动场的分割 .....              | 169        |
| 5.3.3 时空信息的融合 .....               | 170        |
| 5.3.4 对象边缘的精细化 .....              | 170        |
| 5.3.5 实验结果 .....                  | 171        |
| 5.4 MPEG 压缩域视觉关注度对象分割 .....       | 172        |
| 5.4.1 场景纹理分析 .....                | 173        |
| 5.4.2 I 帧运动矢量场处理 .....            | 175        |
| 5.4.3 基于 DCT 系数和运动矢量的统计区域生长 ..... | 176        |
| 5.4.4 关注度对象提取 .....               | 176        |
| 5.4.5 实验结果 .....                  | 177        |
| 参考文献 .....                        | 180        |
| <b>第 6 章 视频对象分割技术的应用</b> .....    | <b>182</b> |
| 6.1 基于视频对象的查询与检索系统 .....          | 184        |
| 6.1.1 视频场景的分割 .....               | 185        |
| 6.1.2 聚类与关键帧的提取 .....             | 186        |
| 6.1.3 视频检索数据库的建立 .....            | 187        |
| 6.1.4 基于音频的浏览系统 .....             | 188        |
| 6.2 视频对象分割在智能监控系统中的应用 .....       | 188        |
| 6.2.1 智能监控系统的组成 .....             | 188        |
| 6.2.2 运动对象检测 .....                | 190        |

---

|                             |     |
|-----------------------------|-----|
| 6.2.3 基于场景内容的查询 .....       | 193 |
| 6.3 可视化通信中的人脸对象分割技术 .....   | 195 |
| 6.3.1 基于二叉划分树的人脸分割方法 .....  | 195 |
| 6.3.2 人脸分割算法步骤 .....        | 196 |
| 6.4 视频对象分割在影视资料修复中的应用 ..... | 201 |
| 6.4.1 斑点损伤的修复算法 .....       | 202 |
| 6.4.2 基于时空结合的斑点损伤修复模型 ..... | 203 |
| 6.4.3 斑点修复的实验结果 .....       | 207 |
| 6.5 压缩视频中运动交通车辆的检测 .....    | 208 |
| 6.5.1 视频对象平面检测和检索算法 .....   | 209 |
| 6.5.2 车辆检测系统的实验结果 .....     | 213 |
| 参考文献 .....                  | 213 |

# 第 1 章 视频对象分割提取概述

进入 21 世纪以来,随着社会信息化的发展,人们对多媒体信息的需求不断增长,其中视觉信息是极其重要的部分。视觉信息中视频是指一个被观察系统(如摄像机)所记录的运动图像序列,是人类直接从外界获得的重要动态信息,但由于其数据量十分巨大,为便于传输和存储,需要进行高效的压缩编码。

现有的视频压缩标准,可分为两类。第一类压缩标准以 JPEG、MPEG1/2、H. 261/3/4 等为代表,其主要特点是将像素和像素块作为基本的编码单元,这类编码标准已在如数字电视、视频通信和 VCD/DVD 等方面应用成功。以基于对象(内容)编码为特点的第二类压缩标准 MPEG-4 除能提供高效的压缩编码效率外,还能提供基于对象的交互功能,使用户能够访问(搜索、浏览)和操作(剪贴、移动)场景中的各个对象,可更广泛地延拓应用范围,因此也被称为第二代编码标准。

基于对象的编码和交互功能首先需要将场景或视频序列中的各类对象(如运动的汽车、人等前景对象和静止的房屋、树木等背景对象)分割提取出来,但 MPEG-4 并没有规定从视频序列中分割出此类具有语义意义的视频对象(video object, VO)的方法,而是对用户开放,其目的是便于用户针对具体应用来设计特定的视频对象分割算法<sup>[1]</sup>。然而,语义视频对象分割与早期的图像分割相比更是一项挑战性的难题<sup>[2]</sup>,为此自 MPEG-4 标准诞生后的 10 年来,国内外包括各大公司、高校和各类研究机构在内的学者和研究人员已进行深入、广泛的研究。目前,尽管还不很完善,但已进入应用阶段,而且应用领域已远远超越了原先仅作为便于高效编码和对象交互功能的范围。

## 1.1 视频对象分割提取的基本概念

### 1.1.1 视频对象的定义

在 MPEG-4 标准中,视频对象被定义为“在景物中的一个单元,允许用户存取(搜索、浏览)和操作(剪贴、移动)”,即视频对象是区域的聚集,且至少有一个特征出现在视频序列中<sup>[1]</sup>,此描述较为抽象。在实际的视频场景中,视频对象是指某些具有语义意义的实体或区域,如行驶的汽车、人等,都可被视为语义视频对象。但对同一视频场景,不同的应用所感兴趣的视频对象是不同的。例如,对于图 1.1 所示场景,尽管除了人以外还有其他内容,但是如果我们关注的是人,则两个玩耍的

小孩和他们正在玩耍的球就是我们感兴趣的拟分割提取的视频对象。语义视频对象的提取过程本质上是一个分割问题,因此有时称为提取,有时也称为分割。

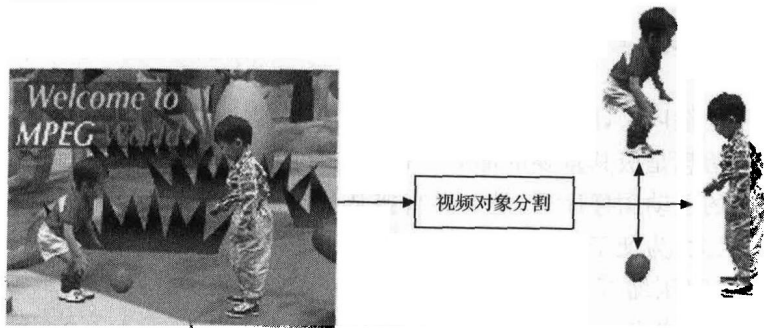


图 1.1 从视频场景中分割关注的视频对象

### 1.1.2 视频分割与图像分割的关系

视频是一连串的图像序列,视频中的每一帧(某一时刻的采样)即可视为一幅图像。因此,视频实际上是图像序列在时间维的扩展,但视频处理并不是图像处理在时间维的简单扩展,这是因为视频序列中相邻图像之间存在大量的时间冗余和空间冗余,相关性很强。因此,视频分割与图像分割之间既有区别又存在联系:

- ① 图像分割常利用空间信息进行分割,视频分割同时还利用时间信息进行分割。
- ② 图像分割依赖于像素级的低级特征,如灰度、颜色和纹理的一致性,往往是相似性区域的聚类,通常不涉及语义意义,因此也被人称为“低级”分割。视频对象分割由于对象包含着运动信息,常会发生遮挡区和显露区等问题以及存在非刚体对象的形变问题,很难仅用上述的低级特征来刻画,还需借助于语义的一致性。由此而言,视频分割比图像分割要复杂得多。
- ③ 图像分割是一幅静态图像单独的分割。视频分割先要进行初始帧分割,分割初始帧对象时既可沿用图像分割的某些方法,又常运用后续帧的相关信息,获得初始帧视频对象后,通过对象跟踪算法,持续地分割出后续帧的视频对象,因此视频分割并不是单独一帧图像的分割,而是一种前后关联的多线索融合分割。此外,图像分割已研究了数十年,目前已有千余种分割算法<sup>[3]</sup>,而视频对象分割起步相对较晚,但由于有广泛的应用前景,发展甚为迅速。

## 1.2 视频对象分割方法的分类与应用概况

### 1.2.1 分割方法的分类

常见的视频对象分割方法可有下面四种分类。

### 1. 应用目标不同

按照应用目标的不同来分类可分为两类:一类是要求得到准确的视频对象轮廓<sup>[4]</sup>,但不追求实时应用;另一类是要求能实时处理、在线应用,但对所分割得到的对象轮廓的准确性要求得并不很严格。

### 2. 是否需要人工参与或人机交互

按照是否需要人工参与或人机交互来分类可分为两类:一类是无需人工参与或人机交互的自动分割,包括对初始帧视频对象分割和对后续帧的对象跟踪都可以自动进行<sup>[5]</sup>;另一类需借助人机交互的半自动分割,即借助人工参与或人机交互来定义语义视频对象的轮廓形状和位置,进而来分割初始帧的视频对象,后续帧的对象跟踪则自动地进行(有时需对跟踪对象的区域边界按预先定义的语义特征稍做修正,以减少跟踪误差<sup>[6]</sup>)。

### 3. 分割过程中所用信息不同

按照分割过程中所用信息的不同来分类可分为空间分割、时间分割和时空分割三类。空间分割先按传统的图像分割将该帧图像划分为区域,其中某些区域具有相似特征而与其他区域明显不同,然后将这些具有相似特征的一致性区域,按照一定的空间信息(颜色、灰度、边沿、纹理)、变换域信息(DCT 系数)、统计信息<sup>[7]</sup>和先验知识<sup>[8]</sup>进行分割和聚类成语义视频对象;时间分割通常是利用前后帧之间视频对象的运动信息进行分割,也可结合颜色、纹理和边沿等特征<sup>[9]</sup>;时空分割一般先通过时间分割标识出运动对象,然后与空间分割得到的对象边界融合在一起,以得到更精确的分割结果<sup>[10]</sup>,这是目前的主流分割方法。

### 4. 所提供视频素材的不同

根据所提供的视频素材的不同来分类可分为像素域分割和压缩域分割两类。以前提供的视频素材较多的是未经过压缩编码的原始数字视频,每帧图像是以像素集的形式出现的,建立在对这类视频素材的分割方法称为基于像素域视频对象的分割;但随着多媒体信息的逐步普及,用户获得的视频素材越来越多是经过压缩编码的码流,如果仍沿用基于像素域的分割方法,就需先对压缩流进行解压缩,这就会使处理时间大量增加。因此,近几年来发展了不经解压缩、直接对压缩流进行分割的方法,称为基于压缩域的视频对象分割<sup>[7]</sup>。由于压缩的码流中已有可作为视频分割主要线索的运动矢量,因此具有快速分割的特点。尽管给出的运动矢量场是稀疏的,需经过多帧运动矢量的累加和内插以得到稠密的运动矢量场,但视频对象分割速度仍远高于基于像素域的分割,完全可满足在线实时分割的要求<sup>[8]</sup>。

鉴于早期的压缩流大多是 MPEG-2 码流,近期的压缩流较多是 H. 264 码流,因此,目前的压缩域视频对象分割大都是基于 MPEG 压缩域<sup>[9]</sup>和 H. 264 压缩域<sup>[10]</sup>。

本书对视频对象分割方法的分类主要采用上述的第 4 种分类方法,即分为像素域视频对象分割和压缩域视频对象分割方法。对于前者又按照第 1 种分类方法,即分为以分割准确度为目标的像素域分割方法和以实时应用为目标的像素域视频对象分割方法。至于上述的第 2 种分类和第 3 种分类,其实分别是使用工具的不同或所用信息特征的不同,可包含在分割过程中。

### 1.2.2 应用概况

在 MPEG-4 视频编码标准中,视频对象分割是用于基于对象的编码的,即将视频场景分割为若干个运动的前景对象和静止或存在全局运动的背景对象,对其中所关注的前景对象在压缩编码时分配较多的比特,而对那些不重要的背景对象分配很少的比特,从而达到既增强了所关注视频对象的图像质量,又使编码的比特率不增加,甚至减少。而且,它可以实现由用户操作的对象间的交互功能。

目前,视频对象分割的应用早已超越了上述基于对象编码的范围,归纳起来有以下的延伸应用:

(1) 结合 MPEG-7<sup>[11]</sup>,进行基于对象(内容)的视频索引和检索等<sup>[12]</sup>。

(2) 基于对象的智能视频监控系统,包括用于公安、交通、金融、大楼和社区等各种不同用途的智能监控。

(3) 可视化通信,包括可视电话、手机电视以及视频会议系统等。

(4) 虚拟现实,包括电子游戏、虚拟演播室、虚拟博物馆以及商业上的虚拟专卖店、虚拟试衣间和虚拟眼镜店等。

(5) 医学应用,如脑部图像的分割、各种器官的病灶分割等。

(6) 遥感应用,如分割合成孔径雷达图像中的目标或遥感云图中的不同云系和背景分布等。

(7) 工农业中的监控检测,如分割燃烧炉中的火焰以便监测燃烧过程并将图像分割用于户外植物的检测等。

(8) 各类修复中的应用,如影视资料中斑点区域和损伤区域的分割提取以及布匹生产线中疵点的分割等,以便进行在线或离线的修复。

上述的应用将在本书的第 6 章中给予较详细的介绍。

## 1.3 本书的结构

本书共分为 6 章,其结构安排如下。第 1 章概括性介绍了视频对象分割的基本概念,包括视频对象的定义、视频对象分割与传统图像分割的区别和联系,各种

视频对象分割方法的分类和目前视频对象分割的应用概况。考虑到针对性,将国内外研究动态和发展现状放在其后的各个章节中。

第2章分析讨论了视频对象分割算法的基本理论及其所采用的主要技术。着重阐述空间分割、时间分割和时空融合分割的基本原理,也给出了对象分割之前的各种预处理和后处理方法,这对提高所分割、提取的视频对象的质量颇为重要。对于视频对象分割算法的性能评价,除了主观评价外,也给出了客观性能评价的准则。这样,既可定量地确定所提取的视频对象是否符合用户需要应用场合的性能要求,还可在线判断所用分割方法的优劣,以便及时调整分割参数或选择新的分割方法。

在第2章基本理论的基础上,第3~5章对当前有代表性应用目标视频对象分割算法进行了深入和较详细的分析讨论。其中第3章和第4章均是基于像素域视频素材的,前者以提高对象分割的准确度为目标,后者以提高分割处理的实时性为目标。第3章中对于如何进行高质量的时空信息的融合分割、如何解决背景遮挡问题,以及对存在多个运动视频对象且有互遮挡的情况下如何进行有效分割等难题,给出了合理的解决方案;此外,还对通过人机交互进一步提高所分割对象轮廓的精度和提高后续帧对象跟踪的准确性和完备性给出了可行方案。通常的像素域视频对象分割方法尚难以达到应用意义上的实时分割(即在帧间隔40ms内完成分割),第4章和第5章给出了实现实时分割的各种解决方案。其中第4章虽仍是基于像素域的,但用细胞神经网络(CNN)来完成,介绍了CNN的开发工具和CNN模板的设计方法以及举例说明如何将通常的像素域分割方法移植到以CNN工具来实现。第5章是基于压缩域视频素材的,是近几年才发展起来的可实现实时分割的新方法。由于目前压缩域视频素材大多是H.264和MPEG-2的码流,因此主要讨论基于这两类压缩域的VO分割方法。由于从压缩码流中直接提取的运动矢量场是稀疏的,所得到空间信息仅是DCT的直流系数(DC),也是粗糙的,为此,第5章结合具体的分割算法较深入地讨论了如何使稀疏的运动矢量场稠密化和使粗糙的空间信息精细化问题,在此基础上再进一步给出从压缩域提取视频对象的具体方法和算法。上述三章大都是在完成三项国家自然科学基金项目过程中所取得的研究成果,故每种VO的分割提取方法都给出了实验结果,以供广大科技工作者进行拓展研究或应用时作参考和便于进行性能比较。

第6章介绍了视频对象分割的应用。因篇幅所限,不可能面面俱到,这里给出了不同应用领域的5个案例,供读者参考,便于举一反三。

## 参考文献

- [1] Sidora T. The MPEG-4 video standard verification model. IEEE Transactions on Circuits System for Video Technology, 1997, 7(1): 19—31.

- 
- [2] Zhang D S, Lu G J. Segmentation of moving object in image sequence: A review. *Circuits System Signal Processing*, 2001, 20(2): 143—183.
  - [3] 章毓晋. 图像分割. 北京:科学出版社, 2001:10.
  - [4] Daras P, Kompatsiaris I, Grinias I, et al. MPEG-4 authoring tool using moving object segmentation and tracking in video shots. *EURASIP Journal on Applied Signal Processing*, 2003, 9: 1—18.
  - [5] Yang G, Zhang Z. Objective performance evaluation of video object segmentation algorithms with ground truth. *Journal of Shanghai University*, 2004, 8(1): 70—74.
  - [6] Kim Y R, Kim J H, Ko S J. Semiautomatic segmentation using spatio-temporal gradual region merging for MPEG-4. *IEICE Transactions on Fundamentals*, 2003, E86-A(10): 2526—2543.
  - [7] Babu R V, Remakrishnan K R. Video object segmentation: A compressed domain approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004, 14(4): 462—475.
  - [8] Mezaris V, Kompatsiaris I, Strintzis M G. Real-time compressed domain spatiotemporal segmentation and ontologies for video indexing and retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004, 14(5): 606—621.
  - [9] Shi X, Zhang Z, Shen L. Multi-resolution of video objects in the compression domain. *Optical Engineering*, 2007, 46(9): 097401—097406.
  - [10] Liu Z, Lu Y, Zhang Z. Real-time spatiotemporal segmentation of video objects in the H. 264 compressed domain. *Journal of Visual Communication and Image Representation*, 2007, 18(2): 275—290.
  - [11] Guo J, Kuo C-C J. *Semantic Video Object Segmentation for Content-based Multimedia Application*. Berlin:Springer, 2001:17.
  - [12] Lin J, Kung S Y. *Video Object Extraction and Representation: Theory and Application*. Berlin:Springer, 2000:155.

## 第2章 像素域视频对象分割基础

数字图像处理技术是数字视频处理的基础。在视频对象分割提取中,需要使用到众多的数字图像处理技术,尤其是预处理技术,如滤波、去噪等。本章先简要介绍视频对象分割所涉及的图像预处理和后处理技术,然后分别讨论基于空间域、时间域和基于时/空融合的视频对象分割的基础知识,最后给出对于视频对象分割性能的客观评价准则。

### 2.1 图像的预处理和后处理技术

#### 2.1.1 图像的滤波处理

图像滤波是数字图像处理过程中最经常使用的、也是最重要的处理过程,因为图像在摄取、传输等过程中经常会受到噪声的干扰,反映在原本均匀和连续变化的灰度突然变大或者变小,形成一些虚假的物体边缘或轮廓,使得图像的后续处理容易引入误差。因此,需要重视图像的滤波处理。以下是几种常见的图像滤波处理方法<sup>[1]</sup>。

##### 1. 均值滤波法

假设图像由许多灰度恒定的小块组成,相邻像素间有很强的空间相关性,而噪声是统计独立地叠加在图像上的,其均值为零。因此,可以用像素邻域内的各像素灰度值的平均值代表原有的灰度值,实现图像的平滑。在灰度图像  $f$  中以像素值  $(x, y)$  为中心的  $N \times N$  窗口 ( $N = 3, 5, 7, \dots$ ) 内,若平均灰度值为  $\alpha$  时,无条件地令  $f(x, y) = \alpha$ 。这种方法就是对一个噪声点进行模糊,把被处理点的某一领域中所有像素灰度值的平均值作为该点灰度的估计值。由于参加平均的像素在原始无噪声图像中灰度值是不等的,但在去噪过程中邻域内所有像素都进行了平均,均值滤波法可能会使边界模糊。

##### 2. 中值滤波器

中值滤波法是以局部中值代替局部平均值。在灰度图像  $f$  中以像素值  $(x, y)$  为中心的  $N \times N$  窗口 ( $N = 3, 5, 7, \dots$ ) 内,首先把这  $N \times N$  个像素点的灰度值按大小进行排序,然后选取值的大小为处于正中位置的那个灰度值  $\alpha$ ,使  $f(x, y) = \alpha$ 。这样,把被处理点的某一邻域中像素灰度中值作为该点的灰度的估计值。由于中

值滤波不影响阶跃函数和阶梯函数,而当宽度小于窗口的一半时,冲击函数趋于消失,三角形函数的顶部则将被削平,因此中值滤波一般不会引起边缘模糊,而能够达到减小离散的冲击噪声的目的。

### 3. 低通滤波器法

从频谱上看,噪声特别是随机噪声是一种具有较高频率分量的信号。平滑的目的就是通过一定的手段滤除这类信号。一个很自然的想法就是使图像经过一个二维的低通数字滤波器,让高频信号得到较大的衰减。在空间域上进行的这种滤波实际上就是对图像和滤波器的冲击函数进行卷积。

设图像为  $f(x, y)$ , 滤波器的冲击响应函数为  $H(x, y)$ , 则卷积表达式为

$$g(u, v) = \sum_x \sum_y f(x, y) H(u - x + 1, v - y + 1) \quad (2.1)$$

常见的冲击响应函数有

$$\mathbf{H}_1 = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{H}_2 = \frac{1}{10} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{H}_3 = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.2)$$

显然,采用  $\mathbf{H}_1$  作为滤波器,其效果将与  $3 \times 3$  窗口下的均值滤波法得到的效果类似。不同的冲击响应函数之间的区别在于计算窗口内像素值的加权平均值的加权系数不同。

#### 2.1.2 梯度算子

图像边缘对图像识别和计算机分析十分有用。边缘能勾画出目标物体,蕴含了丰富的内在信息(如方向、形状、阶跃性质等),是图像识别中抽取图像特征的重要属性。

图像边缘是由于相邻像素间灰度值剧烈变化引起的。对于图像中变化比较平坦的区域,因相邻像素的灰度变化不大,因而其梯度幅值较小(趋于零),而图像的边缘地带,因相邻像素的灰度值变化剧烈,所以梯度幅值较大。因此,用一阶导数幅值的大小可以判断图像中是否有边缘以及边缘的位置,二阶导数的符号则可以用于判断一个边缘像素是在边缘亮的一边还是暗的一边,而且过零点的位置就是边缘的位置。

梯度对应一阶导数,对于一个连续图像函数  $f(x, y)$ ,它在点  $(x, y)$  上的梯度是一个矢量,定义为

$$\nabla f(x, y) = [G_x, G_y]^T = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]^T \quad (2.3)$$

式中,  $G_x$  和  $G_y$  分别为沿  $x$  和  $y$  方向的梯度。梯度的幅度  $|\nabla f(x, y)|$  和方向角  $\phi(x, y)$  分别为