

为海量视频处理
提供解决方案

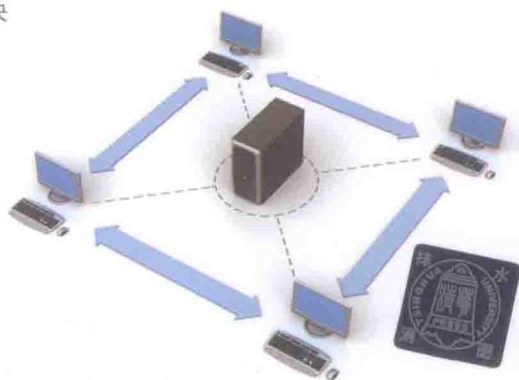
Visual Big Data Basis and Applications

视觉大数据

基础与应用

谢剑斌 等编著

- 结合海量视频分析与搜索在相关行业的应用，解决视频大数据处理中的诸多核心问题
- 详细阐述人脸搜索系统、车牌搜索系统、车标搜索系统、暴力行为检测系统、可疑行为检测系统、海量视频摘要系统和海量视频管控平台等实际案例

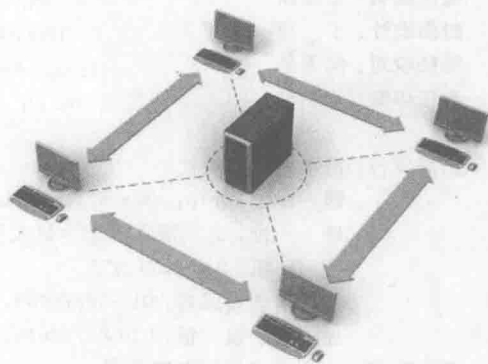


清华大学出版社

视觉大数据

基础与应用

谢剑斌 刘通 闫玮 编
李沛秦 王勇 谭筠



清华大学出版社
北京

内 容 简 介

本书是视频大数据处理领域的著作。为使读者全面了解海量视频分析与搜索的基础知识及应用方法,本书首先介绍海量视频概论、海量视频模型、海量视频管理和海量视频分析等相关基础知识,然后具体阐述面向大数据的大规模人脸搜索系统、面向高清卡口的车辆车牌与车标等信息搜索系统、暴力行为检测系统、可疑行为检测系统、海量视频摘要系统和海量视频管控平台等典型的海量视频分析与搜索实例,并将海量视频分析与搜索领域的新技术和新成果贯穿于全文的描述之中。

本书主要适用于从事海量视频分析与处理领域的应用开发和工程施工技术人员阅读。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

视觉大数据基础与应用/谢剑波编著. —北京:清华大学出版社,2015

ISBN 978-7-302-39112-7

I. ①视... II. ①谢... III. ①视频编辑软件 IV. TP394

中国版本图书馆CIP数据核字(2015)第012817号



责任编辑:王金柱

封面设计:王翔

责任校对:闫秀华

责任印制:宋林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印刷者:北京富博印刷有限公司

装订者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:180mm×230mm 印 张:16.5 字 数:410千字

版 次:2015年3月第1版 印 次:2015年3月第1次印刷

印 数:1~3000

定 价:49.00元

产品编号:056802-01

前言

常言道“百闻不如一见”，人类感官接受的各种信息约有 80% 来自视觉。视频和图像等可视化信息是对客观事物形象、生动的描述，是直观而具体的信息表达形式，是人类社会最重要的信息载体。随着光学成像、数字视频、计算机、信号处理等技术的快速发展，以及人类社会对信息获取、安全保卫、智能服务等应用的迫切需求，视频与图像日益受到人们的青睐。海量视频处理在视频图像与内容描述之间建立映射关系，通过视频图像分析来理解场景内容，如人脸识别、行为识别、车牌搜索、车标搜索、视频摘要等。

本书是作者十多年研究海量视频处理的心血结晶，可作为信息、计算机、自动化、电子与通信等学科专业高年级本科生和研究生的实践教材，也可以作为从事海量视频处理领域技术人员的参考设计资料。全书分为 10 章，首先详尽地介绍海量视频的模型、管理、分析的基本理论；然后深入地阐述大规模人脸搜索、高清卡口车辆信息搜索、暴力行为检测、可疑行为检测、海量视频摘要等系统的实施方案和实验仿真，并提供配套的源代码和视频库；最后以某个市级公安局应用为参考，深入浅出地描述海量视频管控平台。

本书由国防科技大学电子科学与工程学院数字视频课题组编著，谢剑斌教授主编，第 4 章、第 10 章由谢剑斌执笔，第 7 章、第 8 章由刘通执笔，第 1 章、第 9 章由闫玮执笔，第 2 章由谭筠执笔，第 3 章、第 6 章由李沛秦执笔，第 5 章由王勇执笔，全书由谢剑斌统稿、修改和完善。在编著过程中得到国防科技大学庄钊文教授、唐朝京教授的关心和指导，林成龙、刘双亚、穆春迪、许浩、戴超、李润华等为本书的编著做了大量工作，国家自然科学基金项目（61303188）对本书的相关研究工作进行了资助，在此一并致谢！由于时间有限，书中若有纰漏之处，敬请广大读者批评指正。

配套资源的网站为：www.kedachang.com。

编者

目录

第 1 章 海量视频概述	1
1.1 视觉大数据	1
1.2 关键技术	3
1.3 应用领域	4
1.4 挑战与发展	5
第 2 章 海量视频模型	8
2.1 HSV 颜色模型	8
2.2 肤色模型	12
2.3 形状模型	14
2.4 人体可变形模型	20
2.5 混合高斯模型	21
2.6 概率图模型	24
2.7 感兴趣区域模型 (ROI)	26
2.8 视觉显著性模型	28
2.9 多分辨率模型	31
2.10 视觉词袋模型	34
2.11 视频语义模型	37

第 3 章 海量视频管理	40
3.1 视频数据库	40
3.1.1 海量视频数据	40
3.1.2 面向对象的海量视频数据库	41
3.2 集中式视频数据库	42
3.3 分布式视频数据库	43
3.3.1 基于 Hadoop 的视频数据库	44
3.3.2 MapReduce 模型	47
3.4 博世视频管理系统	52
3.5 微博视频管理系统	53
3.6 VOD 视频点播及管理系统	55
第 4 章 海量视频分析	57
4.1 Harris 描述子	57
4.2 SIFT 描述子	61
4.3 K 均值聚类方法	68
4.4 K 近邻法	72
4.5 SVM 方法	73
4.6 BP 网络	83
4.7 多感知器模型	93
4.8 卷积神经网络 (CNN)	95
4.9 AdaBoost 方法	102
4.10 模拟退火方法	106
4.11 遗传方法	109
第 5 章 大规模人脸搜索系统	119
5.1 概述	119
5.2 人脸检测	124
5.2.1 人脸检测方法分类	124
5.2.2 基于 Adaboost 的人脸检测	126
5.3 人脸特征提取	130
5.3.1 PCA 方法	132
5.3.2 LDA 方法	134

5.3.3	Kernel 方法	136
5.4	人脸特征比对	138
5.4.1	典型的度量方法	139
5.4.2	典型的分类器	141
5.5	“大海捞针”人脸搜索系统	144
5.5.1	体系结构	144
5.5.2	关键技术	145
5.5.3	算法伪代码	145
5.5.4	性能评价	148
5.5.5	系统搜索效果	149
第 6 章	高清卡口车辆信息搜索系统	150
6.1	车辆信息搜索	150
6.2	车牌搜索子系统	151
6.2.1	车牌搜索概述	151
6.2.2	车牌区域定位	152
6.2.3	车牌字符分割	159
6.2.4	索车牌字符识别	163
6.3	车标搜索子系统	166
6.3.1	车标定位	167
6.3.2	车标搜索	170
第 7 章	暴力行为检测系统	174
7.1	暴力行为	174
7.2	暴力行为检测	176
7.2.1	系统框架	176
7.2.2	行为数据库	186
7.2.3	评价指标	187
7.3	基于对象层次的暴力行为检测系统	188
7.4	基于光流变化的暴力行为检测系统	192
7.5	基于运动着色的暴力行为检测系统	194

第 8 章 可疑行为检测系统.....	198
8.1 可疑行为	198
8.2 可疑行为检测	200
8.3 基于轨迹特征的可疑行为检测系统	200
8.3.1 系统结构	201
8.3.2 人体目标检测	201
8.3.3 轨迹建模	203
8.3.4 轨迹特征提取	206
8.3.5 轨迹特征分类	207
8.4 基于运动方向的可疑行为检测系统	208
8.4.1 系统流程	208
8.4.2 背景边缘模型	209
8.4.3 前景帧判断	209
8.4.4 行为特征描述	210
8.4.5 SVM 分类	211
8.5 基于形状特征的可疑行为检测系统	211
第 9 章 海量视频摘要系统.....	214
9.1 视频摘要	214
9.2 视频摘要过程	215
9.3 特征提取和表示	219
9.3.1 颜色特征提取	219
9.3.2 纹理特征提取	221
9.3.3 形状特征提取	223
9.3.4 运动特征提取	224
9.3.5 音频特征提取	227
9.4 典型系统	229
第 10 章 海量视频管控平台.....	234
10.1 平台要求	234
10.2 平台架构	235
10.3 平台组成	236
10.4 平台服务器	239

10.5 平台功能	240
10.5.1 视频监控与回放	240
10.5.2 视图无缝融合功能	242
10.5.3 大规模人脸等目标监测	243
10.5.4 异常行为检测	244
10.5.5 海量视频摘要	244
10.5.6 高清卡口车辆信息搜索	244
10.6 平台应用	246
参考文献	247

第 1 章

海量视频概述

2008 年 9 月, *Nature* 推出封面专栏“大数据 (Big Data)”, 阐述大数据在数学/物理/生物/信息等基础学科、工程技术、社会经济领域中扮演着非常重要的角色。随后 *Science*、《华尔街日报》、《求是》等权威媒体大篇幅介绍大数据, 大数据在 Google、百度、必应等成为搜索热点。大数据成为当代的标志符号, 海量视频是大数据的重要形态, 即视觉大数据。

1.1 视觉大数据

如图 1.1 所示, 符合 4 个 V 的数据为“大数据”, 海量视频就是视觉大数据。

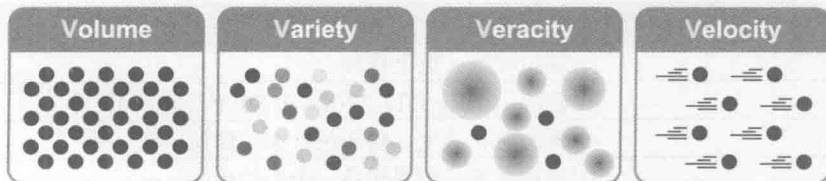


图 1.1 大数据的 4 个 V

1. Volume

Volume (大量) 指数据总量巨大。

以视频监视系统产生的视频数据为例,近年来随着各个城市联网视频监控系统以及高清摄像头的普及,视频数据快速增长。以某个部署 10 000 个标清摄像机的中等城市为例,每个摄像机每秒采集到的视频数据经压缩编码后的数据量约为 720 (画面宽度) $\times$ 576 (画面高度) $\times 2\text{B}$ (使用 4:2:2 采样) $\times 25$ (帧率) $\times 0.01$ (H.264 编码平均压缩比) $= 207360\text{B} \approx 0.2 \times 10^6 \text{B}$ (0.2MB),每天产生的视频数据量约为 $0.2 \times 10^6 \text{B} \times 60$ (秒) $\times 60$ (分) $\times 24$ (小时) $\times 10000$ (台) $= 172.8 \times 10^{12} \text{B}$ (172.8TB),每个月产生的视频数据量约为: $172.8 \times 10^{12} \text{B} \times 30$ (天) $= 5.184 \times 10^{15} \text{B}$ (5.184PB)。

在实际系统中,为了降低存储压力,通常仅存储关键事件(如人、车、物)的画面和描述信息,保守估计平均每个摄像机每 10 秒发生 1 个事件(即每秒 0.1 个事件),则每天产生的事件记录约为 0.1 (条/秒) $\times 60$ (秒) $\times 60$ (分) $\times 24$ (小时) $\times 10000$ (台) $= 86.4 \times 10^6$ 条,每年产生的事件记录约为 86.4×10^6 条 $\times 365$ (天) $\approx 31.5 \times 10^9$ 条;假设记录需保存 3 年,每条记录平均需要占用 0.4MB (2 秒视频)的存储空间,则所需的总存储空间约为 31.5×10^9 条 $\times 3$ (年) $\times 0.4\text{MB} \approx 37.8 \times 10^{15} \text{B}$ (37.8PB)。

视频分享网站产生的视频数据量同样巨大,据统计,2012 年 YouTube 网站上每分钟由用户上传的视频数据平均超过 40 小时,按标清视频数据量计算,每年产生视频数据量约为 0.2×10^6 (字节/秒) $\times 60$ (秒) $\times 60$ (分) $\times 40$ (小时) $\times 60$ (分) $\times 24$ (小时) $\times 365$ (天) $\approx 15.14 \times 10^{15} \text{B}$ (15.14PB)。近几年随着智能手机等具备视频采集功能设备的普及,视频上传量更是呈现爆发式增长。

2. Variety

Variety (多样)指数据种类很多。

如表 1.1 所示,海量视频数据的来源多种多样、内容包罗万象。

表 1.1 海量视频数据分类

分类依据	具体类别
信号形式	模拟视频、数字视频
分辨率	CIF、4CIF、D1、720P、1080I/P、2K、4K 等
色彩	真彩色、灰度、伪彩色(如热像仪成像)等
来源	监控系统、影视作品、个人拍摄等
场景	室内、室外,白天、夜晚
环境	晴天、阴雨、大雾等
摄像机姿态	固定式、运动式
编码格式	MPEG-1/2/4、H.261/263/264/265、AVS、SVAC 等

3. Veracity

Veracity（精确）指数据的数据总量大、价值密度低。

该特点在海量视频数据上尤为突出。以监视视频为例，在 1GB 的监视视频中，有用的数据总量可能仅仅只有 10MB。

4. Velocity

Velocity（速率）指数据的流通速度快、实时性强。

监控视频数据放映的是监控场景的实时情况，具有实时性；在对监视视频进行处理时，处理速度越快，实时性越高，数据所体现出的价值就越大。

1.2 关键技术

海量视频数据是由传统的分立多源视频数据形成的聚合体，不仅包含原始视频数据的全部数据量，而且通过分析多源视频的内在联系，还可以挖掘出单个视频数据无法提供的信息，实现 $1+1>2$ 的超越。

下面介绍与海量视频相关的关键技术。

1. 存储与管理

海量视频数据集记录数众多，容量巨大，导致在采集、传输、存储、处理、检索、共享、分析、显示数据集时产生巨大障碍，无法采用传统的基于单机或小规模服务器集群的数据库、文件存储、分布式处理技术，必须采用基于大规模计算集群或数据中心的可灵活扩展、可容错、大规模分布式并行处理技术。“云计算”（大规模集群分布式并行计算技术）被认为是当前的最佳解决方案，已经成为智慧城市物联网应用的组成部分。

2. 分析与识别

传统的视频监控系统依赖人工对监控视频进行实时查看和后期搜索。由于人眼并非可靠的观察者，人工值守容易忽视画面监控造成失误；同时人工搜索的效率异常低下，不能满足海量视频的应用需求。

视频分析与识别技术是解决该问题的关键。譬如：在前端高清网络摄像机中植入智能功能，通过视频分析，对高清监控场景中的人或物进行分析和识别，对异常现象产生提示或报警；通过网络将前端视频数据汇总到中心服务器，借助高性能硬件进行实时分析和识别；使用分布式计算和云计算技术，挖掘历史记录视频中的有用信息。

3. 摘要与搜索

海量视频数据的价值密度低，完整存储时会浪费巨大的存储空间。利用视频图像处理（如视频浓缩、摘要、增强等）、模式识别、海量数据分类存储以及视频搜索等技术，对海量的存储录像等原始信息进行分析和挖掘，对于目标的特征、行为、联接关系等信息内容，形成各种分类的特征信息库、元数据和索引等，提供统一接口供外部应用搜索，通过有限的线索，达到快速关联和可靠定位的功能。

1.3 应用领域

下面介绍海量视频的典型应用。

1. 情报侦察领域

在公开的媒体视频数据中有时会包含某个重要目标的局部特征片段，情报机构通过分析海量视频数据，将包含类似目标的视频数据进行提取和汇总，有可能挖掘出有价值的目标信息。

2. 公共安全领域

通过分析遍布大街小巷、车站码头、商场酒店等场所的摄像机数据，借助视频分析技术，安全部门可以及时发现异常情况，并在第一时间做出响应，搜寻事发现场的可疑目标及其去向；借助人脸搜索技术，通过和公安系统嫌疑人信息数据库对接，可以及时发现网上追逃的嫌疑人员等。

3. 智能交通领域

通过分析管辖范围内所有道路摄像机的监视数据，实时分析道路交通流量，交通主管部门可以综合分析和统计全城的交通状况；通过建立统一的车辆信息数据库，借助车牌识别、车型识别、车标识别技术，快速发现套牌车和假牌车，快速搜索并定位特定车辆的轨迹和位置。

4. 休闲娱乐领域

网络视频点播已经成为广播电视传播的重要方式，通过建立分布式云存储架构，用户在任何时间、任何地点，只要通过联网终端，就可以随时点播和观看喜欢的视频节目，以便更好地安排工作和休闲时间。

5. 个性广告领域

网络广告已经成为广告业的重要分支,从业者通过收集、分析用户与广告间的海量互动视频,可以分析出什么内容的广告更能吸引客户、什么长度的广告不会引起用户的反感、什么时段适合哪些广告的投放、什么网站的用户更倾向于哪些类型的广告等。

1.4 挑战与发展

1. 面临的挑战

海量视频数据具有庞大的数据量和信息量,相关领域的深入研究和有效应用面临巨大挑战。

(1) 高效存储

海量视频数据对传输、存储和计算的带宽要求巨大,由于海量视频数据量的急速扩大,大规模计算需求越来越多,处理技术尚未取得重大突破,堆砌高配硬件成为唯一选择,导致硬件投资增长迅猛。如何有效利用已有硬件、避免重复建设、改进硬件性能、降低系统成本是当前面临的重大问题。

摄像机每天 24 小时不停地工作,如实记录发生的一切,而对于用户来说可能大部分信息无效。为了提高海量视频数据的信息密度,亟需视频摘要与搜索。

(2) 快速分析

原始视频数据是非文本、非结构化的数据,对视频内容进行建模和数学表述决定提取性能。如何根据实际需求选用合适模型、如何优化已有模型以满足特定需求、如何衡量模型的表述性能是应用要面临的复杂问题。

在视频监控业务中,错看、漏看、来不及看等是常见困扰。海量视频数据的回溯给安全人员带来生理与心理的双重挑战,经常有看到吐、看到晕等无奈情况。

视频分析的效率决定价值,更低的延迟、更准确的分析是智慧城市的普遍需求。现有技术对 TB 级的数据进行分析和检索需要花费数小时的计算,不能胜任时效性需求。要深入研究适用于海量视频数据实时分析与识别的先进算法和计算模型,实现海量视频数据的模糊查询、快速检索和精准定位。

(3) 优化应用

在视频监控业务中,看只是信息采集方式之一,用才是业务拓展的根本。视频监控业务的效率问题成为阻碍产业发展的关键瓶颈。

随着摄像机覆盖广度、密度的增大,视频数据量呈指数级上升,而视频监控数据的

使用效率却在下降,大量的视频数据仍然是独立的、零散的,散布在各个行业与单位独立的系统中,没有联网共享。

在视频监控业务网络化之后,网络设备越来越多,但是设备利用率相对较低,很多计算资源处于闲置状态,没有实现资源的最大化利用,运算效率很低。

2. 发展方向

(1) 分布式存储

如果类比水库蓄水方式,典型的网络视频监控数据存储模型是一个由小溪汇聚河流、再汇聚到水库的蓄水方式。小溪数量增多、水量增大是水库蓄水量的保证,然而传统方式下蓄水量增大将提高水库建造成本和对蓄水安全性的要求。

采用分布式蓄水模式,在河流中游建立多个中间蓄水池,不仅可以减少主水库蓄水压力和成本,化整为零,还可以提高就近用水效率。

在大数据技术的支撑下,网络视频监控数据的存储模型可转向分布式的数据存储体系,提供高效、安全、廉价的存储方式。

(2) 并行计算

并行计算是指采用多台计算机的计算资源,并行处理分布到各个节点的海量数据,提高数据处理的整体效率,这是目前提高大规模数据处理效率的有效手段。

并行计算主要分为3类,即MPI、MapReduce和Dryad。

□ MPI

MPI(Message Passing Interface,消息传递接口)是目前国际上并行计算领域最流行的API规范,由多家单位共同设计完成,易用性好、可移植性强、异步通信功能完备,是计算机集群、多处理器计算机、超级计算机进行高性能计算的常用技术。

在基于MPI的实现中,对于一个计算任务,一般需要划分为一组独立的计算部分,在初始化时对应生成一组进程,每一个进程完成一个计算部分,在不同节点上运行,进程之间通过集合通信或点对点通信方式进行数据交互,各个节点的计算结果最终汇总到主计算节点,完成同一个计算任务。

□ MapReduce

MapReduce是进行大规模数据处理的并行计算模型,由Google在2004年提出,应用于大规模集群。

Map(映射)和Reduce(化简)是计算的两个阶段,前者通过调用Map函数实现一组键值到一组新键值的映射计算;后者采用Reduce函数对所有映射计算结果进行化简。

与 MPI 相比, MapReduce 在数据存储节点就地或就近完成 Map 或 Reduce 计算, 减少了数据的网络传输压力。

□ Dryad

Dryad 是微软在 2007 年提出的数据并行计算模型, 与 MapReduce 相同, Dryad 也是通过在数据存储节点就地或就近完成相关计算的方式, 减少数据的网络传输压力。

Dryad 采用 DAG (有向无环图) 表示单个任务, 按照 DAG 的方向依赖进行计算, 计算类型相对于 MapReduce 更加丰富, 计算结果可以通过 TCP Pipes、Shared-memory FIFOs 方式进行传输, 避免冗余磁盘 IO 操作, 传输手段更加高效。

海量视频模型是海量视频处理与分析的基础，本章针对海量视频模型，重点介绍其基本理论和使用方法，包括 HSV 颜色模型、肤色模型、形状模型、人体可变形模型、混合高斯模型、概率图模型、感兴趣区域模型、视觉显著性模型、多分辨率模型、视觉词袋模型、视频语义模型等。

2.1 HSV 颜色模型

颜色模型是采用数学方法表示和处理视频图像信息的第一步，将颜色定义为特定空间的坐标值，不同的颜色空间定义可以得到不同的颜色模型，不同行业常用的颜色模型有 HSV、RGB、HSI、CHL、LAB、CMY 等，海量视频处理常用 HSV 颜色模型。

1. HSV 颜色模型的定义

如图 2.1 所示，在 HSV 模型中，每种颜色由色调（Hue，H）、饱和度（Saturation，S）和明暗度（Value，V）表示，对应于圆柱坐标系中的一个圆锥形子集。

色调 H 由绕 V 轴的旋转角给定，红色对应 0° ，绿色对应 120° ，蓝色对应 240° ；每种颜色和它的补色相差 180° 。

饱和度 S 取值从 0 到 1，等于颜色点到 V 轴的距离。

明暗度 V 取值从 0 到 1，对应于颜色点在 V 轴的投影位置。