

基于内容的 视频编码与传输控制技术



武汉大学学术丛书
WUHAN UNIVERSITY ACADEMIC LIBRARY

贺贵明 吴元保 蔡朝晖 著
蒋 旻 刘振盛 刘志雄



全国优秀出版社
武汉大学出版社

图 书 在 册 编 目 (CIP)

出 版 大 学 汉 武 大 学 出 版 社 出 版

2005.4

(书 丛 学 学 大 汉 武)

ISBN 7-307-04387-4



武 汉 大 学 学 术 丛 书

8 1919.8

基于内容的 视频编码与传输控制技术

贺贵明 吴元保 蔡朝晖 著
蒋 旻 刘振盛 刘志雄

支 持 计 算 机 图 像 处 理 技 术 的 应 用

出 版 大 学 汉 武 大 学 出 版 社 出 版

(电 子 邮 件: wqbq@whu.edu.cn 网 址: www.whu.edu.cn)

武 汉 大 学 出 版 社

8 页 册

2005 年 1 月 第 1 版

定 价: 40.00 元

凡 购 书 者 均 附 赠 试 阅 样 本 一 份, 如 有 需 要 者 请 向 本 社 索 取。 本 书 内 容 详 实, 图 文 并 茂, 是 研 究 本 领 域 的 必 读 之 作。

图书在版编目(CIP)数据

基于内容的视频编码与传输控制技术/贺贵明等著. —武汉: 武汉大学出版社, 2005. 4

(武汉大学学术丛书)

ISBN 7-307-04387-4

I. 基… I. 贺[等]… II. ①图象编码 ②图象通信—数据传输
IV. TN919.8

中国版本图书馆 CIP 数据核字(2004)第 117878 号

责任编辑: 黄金文 杨 华 责任校对: 程小宜 版式设计: 支 笛

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: wdp4@whu.edu.cn 网址: www.wdp.whu.edu.cn)

印刷: 湖北新华印务有限公司

开本: 787×1092 1/16 印张: 23.25 字数: 555 千字 插页: 3

版次: 2005 年 4 月第 1 版 2005 年 4 月第 1 次印刷

ISBN 7-307-04387-4/TN·18 定价: 40.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有缺页、倒页、脱页等质量问题, 请与当地图书销售部门联系调换。

内 容 简 介

本书讨论当前多媒体技术领域极度关注并代表发展方向的基于内容处理技术。全书分两部分:第一部分为基于内容的视频编码技术,共分六章,主要讨论视频对象分割、视频对象形状编码、对象的纹理编码、对象的运动编码以及整体编码程序结构;第二部分为视频流传输 QoS 控制技术,共分四章,主要讨论基于因特网的传输控制技术和网络应用层 QoS 控制技术。

本书可供相关专业大学生、研究生阅读,也可供从事多媒体处理与传输控制的相关工程技术人员参考。

前 言

在国际多媒体技术领域,继 MPEG-4 于 1999 年 12 月被推出后,2001 年 9 月又推出了 MPEG-7,同时做好了制定 MPEG-21 标准的计划,该计划在 2000 年 6 月即被批准,该标准正在加紧制定中。这些标准推动了全世界多媒体技术按面向对象和基于内容处理发展,这也是本书题目和内容确定的初衷。

MPEG-4 是多媒体数据压缩编码国际标准,强调基于对象编码,为了便于实现对视频内容的访问,引入了视频对象(Video Object)的概念,采用基于对象的压缩编码,支持基于内容的交互处理。

MPEG-7 是多媒体内容描述接口的国际标准,为实现基于内容的多媒体检索提供支持,故郑重地把多媒体应用和处理按基于内容的框架固定下来。

MPEG-21 将是多媒体框架(Multimedia Framework)的国际标准,其力图使不同的多媒体技术和处理标准结合在一起,使解决查询和获取多媒体信息能根据不同内容采用不同方法。

所有这些为我们面向对象、基于内容处理多媒体指明了方向、奠定了基础,并锁定了决心。

现在的问题是如何施行基于内容的多媒体处理,特别是基于内容的视频处理。视频处理类型主要包括:视频数据压缩编码、传输、存储、检索等,而压缩编码是前提和基础。所以本书把基于内容的视频编码作为第一部分主要研究和讨论的内容。其中:

第一章是基于内容处理概述;

第二章讨论视频对象分割;

第三章讨论视频对象形状编码;

第四章讨论对象的纹理编码;

第五章讨论对象的运动编码;

第六章就 MPEG-4 编码程序设计讨论其流程结构和参数设置。

在 Internet 快速发展进步的现今世界,多媒体应用决不局限于个体,面向广域范围、面向全社会交互这是必然,那么基于 Internet 传输成了首选,未来 Internet 上运行的主要内容也将是不同应用的多种媒体流。

在过去的年代里,传送话音是使用专用市话网或长途电话网,传送会议电视也使用专用设备,所以应用起来很单纯,没有其他的影响。但在 Internet 上大不一样,数据、文本、视频、音频等多种媒体都向其中充斥,资源的有限、设备的异质、管理的多层使得多媒体在其中传输延时和丢失不可避免,给应用造成限制,给从事多媒体技术的人们带来许多挑战。所以本书第二部分选定了这个棘手的问题展开研究和讨论。其中:

第七章讨论几种常规传输协议带来的控制措施;

第八章讨论网络设备支持的 QoS 控制技术;

第九章研究在网络应用层的 QoS 控制技术；

第十章综合讨论基于 Internet 的视频内容传输控制,其中还结合了适当的与传输控制相适应的编码措施。

全书由贺贵明主笔,吴元保、蔡朝晖、蒋旻、刘振盛、刘志雄都参与了一定的工作。在此特地指出,贺贵明指导的研究生贾振堂、李凌娟、向晓英、许蓉、周捷、罗敏、李国怀的论文和研究工作成果不同程度地在文中得以体现,在此一并致谢。

本书在写作过程中还得到本实验室许多成员的大力支持,在此致以衷心的感谢。

由于多媒体技术正处于全面蓬勃发展时期,作者必有未跟上步伐之处,加之学识水平有限,使书中错误和不足之处难免,恳请读者批评指正。

作 者

2004 年 5 月

目 录

第一部分 基于内容的视频编码	1
第一章 基于内容处理概述	2
1.1 信息与内容	2
1.1.1 基于内容处理的优越性	2
1.1.2 视频内容的表示	3
1.2 基于内容编码概述	4
1.2.1 功能特点 ^{[2][3]}	4
1.2.2 MPEG-4 编码特点	5
1.2.3 基于内容编码的体系结构 ^{[1][4]}	5
1.2.4 基于内容处理的逻辑层次 ^[5]	7
1.3 编码终端层次结构 ^{[11][13]}	8
1.4 传送多媒体集成框架 DMIF ^[13]	9
1.4.1 DMIF 参考体系结构	9
1.4.2 DMIF 应用接口(DAI)	10
1.4.3 DMIF 信令协议	11
1.4.4 基于 DMIF 的媒体内容传送	11
第二章 视频对象分割	14
2.1 视频分割技术基础	14
2.1.1 视频分割概述	14
2.1.2 图像预处理技术	17
2.1.3 基于边缘的分割 ^[23]	21
2.1.4 基于区域的分割 ^{[51][75]}	31
2.1.5 基于像素聚类的分割	36
2.1.6 基于运动的分割 ^{[73][51]}	40
2.2 视频对象提取技术	45
2.2.1 视频活动对象提取 ^[32]	47
2.2.2 基于边缘检测的视频对象提取	55
2.3 基于颜色的对象提取	74
2.3.1 图像中的彩色信息	74
2.3.2 颜色区域分割	77
2.3.3 基于颜色的人脸检测 ^[31]	78

第三章 视频对象形状编码	82
3.1 形状编码概述	82
3.1.1 关于对象的形状	82
3.1.2 形状编码的作用和影响	83
3.2 形状编码原理	84
3.2.1 位图形状编码	84
3.2.2 数学形态学算法 ^{[12][23]}	89
3.2.3 轮廓形状编码	91
3.3 形状编码的实施	100
第四章 对象的纹理编码	106
4.1 纹理编码概念	106
4.2 纹理编码原理	107
4.2.1 变换压缩概述	107
4.2.2 变换的快速计算	109
4.2.3 离散余弦变换 DCT	111
4.2.4 为节省变换的措施	121
4.3 纹理编码技术	122
4.3.1 背景纹理编码	122
4.3.2 任意形状区域的纹理编码	123
4.3.3 形状自适应 DCT 算法	124
4.3.4 视频对象 VOP 纹理编码流程	128
4.3.5 纹理分层编码	129
4.4 量化编码	135
4.4.1 量化原理	135
4.4.2 常用量化表	141
4.4.3 量化编码实施	150
4.4.4 量化与方块效应	153
4.4.5 量化与码率控制	156
4.5 变长编码(VLC)技术	158
4.5.1 哈夫曼(Huffman)编码	158
4.5.2 行程编码(RLC)	158
4.5.3 MPEG-4 中的可变长编码	159
4.6 视频内容可扩展分层编码	161
4.6.1 分层扩展性编码应用需求	162
4.6.2 分层可扩展性编码的分类	162
4.6.3 精细的可扩展性分层编码	164
4.6.4 改进的精细可扩展性编码	167
4.6.5 渐进的精细可扩展编码	169
第五章 对象运动编码	172

5.1 常规方块运动编码	172
5.1.1 运动编码概述	172
5.1.2 传统基于块的运动估计算法	173
5.2 面向对象的运动编码原理	177
5.2.1 重复填充技术	178
5.2.2 运动检测	179
5.2.3 整像素搜索	179
5.2.4 半像素搜索	184
5.2.5 考虑对象运动的搜索	188
第六章 基于内容的视频编码程序结构	189
6.1 基于内容的视频编码参数	189
6.1.1 VO 和 VOL	189
6.1.2 VOP	190
6.2 基于内容的视频编码主要流程	191
6.2.1 基本层 VOP 编码主流程图	191
6.2.2 INTRA MB 纹理编码	191
6.2.3 INTER MB 纹理编码	193
6.2.4 MB Binary Shape 编码	195
6.3 码流示例	195
6.3.1 视觉对象层	195
6.3.2 视频对象平面	196
6.3.3 宏块层	196
第二部分 视频流传输 QoS 控制技术	198
第七章 基本网络协议对传输的控制	201
7.1 几种传输协议的控制措施	201
7.1.1 IPV4 协议的控制方式 ^{[47][43]}	201
7.1.2 IPV6 协议及其控制措施 ^[47]	203
7.1.3 TCP 协议控制方式 ^{[42][46]}	210
7.1.4 UDP 协议及控制方式	215
7.1.5 RTP、RTCP 协议及控制方式 ^{[18][48][49]}	217
7.1.6 实时流协议 RTSP(Real Time Streaming Protocol) ^{[9][18]}	219
7.2 多媒体 IP 组播的实现 ^[18]	222
7.2.1 组播分布树	223
7.2.2 MPLS 中的 IP 组播	225
7.2.3 组播 TCP(MTCP-Multicast TCP)	227
第八章 网络设备支持的 QoS 控制技术	228
8.1 QoS 的定义及相关参数	228
8.2 尽力而为(Best Effort)模式 ^{[47][48][76]}	229

8.3 集成服务与资源预留模型 ^{[47][76]}	229
8.4 区分服务模型(Diff-Serv) ^[76]	231
8.4.1 区分服务体系结构 ^[47]	231
8.4.2 边缘路由器中的业务量调整单元	233
8.4.3 DSCP 和每一跳行为方式组	234
8.4.4 PHB 与编码点的映射	235
8.5 多协议标记交换(MPLS) ^{[47][76]}	236
8.5.1 MPLS 概述	236
8.5.2 MPLS 工作原理 ^[47]	238
8.5.3 MPLS 流量工程 ^[47]	242
8.5.4 MPLS Diff-Serv 解决方案	254
8.5.5 改进 MPLS Diff-Serv 解决方案	255
第九章 应用层 QoS 控制技术	259
9.1 基于 TCP 友好方式的控制技术 ^{[18][24]}	259
9.1.1 TCP-Friendly 协议 ^[18]	259
9.1.2 基于模型的 TCP 友好速率控制协议 TFRCP	263
9.2 基于终端的拥塞控制技术	264
9.2.1 码率控制技术 ^{[11][22]}	264
9.2.2 码率整形技术	267
9.3 高层差错控制技术	268
9.3.1 传输层差错控制 ^[11]	268
9.3.2 基于编码的差错恢复技术 ^{[10][11][18]}	269
9.3.3 解法器错误隐藏技术 ^{[15][18]}	272
9.3.4 编、解码结合的差错控制方法 ^{[15][22]}	273
第十章 基于因特网的视频内容传输 ^{[19][17]}	275
10.1 视频内容基于因特网传输概述	275
10.1.1 视频内容基于因特网传输体系	275
10.1.2 传输协议	277
10.1.3 系统的反馈控制	278
10.2 视频发送端的编码结构	279
10.2.1 编码模式选择	280
10.2.2 视频数据打包算法	280
10.3 自适应数据率控制编码	281
10.3.1 ARC 的二次 R-D 模式	282
10.3.2 初始化阶段的工作	283
10.3.3 预编码阶段的工作	283
10.3.4 编码阶段	284
10.3.5 编码后的处理工作	285
10.4 基于 RTCP 的端-端反馈控制	286

10.5 接收端解码	287
10.5.1 视频解码整体功能结构	287
10.5.2 解码错误隐藏	288
10.6 传输信道模型探测	288
附 录 实 时 流 协 议^[9]	290
1 简 介	290
2 符号约定	296
3 协议相关参数 (Protocol Parameters)	297
4 RTSP 消息 (RTSP Message)	300
5 一般头字段 (General Header Fields)	301
6 请求 (Request)	301
7 应答	302
8 实体 (Entity)	306
9 连接 (Connections)	307
10 方法 (Method) 定义	308
11 状态码 (Status Code) 的定义	317
12 头域的定义	319
13 Caching	332
14 例子	333
15 语法	341
16 安全考虑	342
附录 A RTSP 协议状态机	344
附录 B 用 RTP 进行交互	346
附录 C 针对 RTSP 会话描述的 SDP 的使用	347
附录 D 最小 RTSP 实现	350
参 考 文 献	353

第一部分 基于内容的视频编码

引 言

基于内容的视频编码区别于过去一般视频编码的主要点是,它将所表现的内容(尤其是动态内容)从视频中提取出来分别编码,而过去一般编码的做法是主要针对时间序列的视频帧进行编码,帧似乎就是对象。

视频内容主要指视频所表现的对象,首先是活动对象,其次也包括一些相对静止的对象;进一步的视频内容则是围绕对象发生的内容,表现对象各个方面信息的内容,主要有:对象的形状、对象的运动、对象的颜色、对象的纹理等。

基于这样对内容的分类,所以基于内容的视频编码要研究的东西也就确定了,首先研究视频对象的分割提取,将视频对象(本书中限于活动对象)一个一个地分开以后,接着研究对象的形状编码、研究对象的运动编码、研究对象的纹理编码,也研究对象的颜色编码,还包括背景的纹理和颜色编码。这些形成本书第一部分的内容:

第一章 基于内容处理概述,简述内容处理的特点和基于内容编码的框架。

第二章 视频对象分割,主要讲述分割的基本原理和技术,这是只有基于内容的视频编码才具有的技术。

第三章 视频对象形状编码,讲述形状编码的技术和方法;这是基于内容编码特有的技术和处理。

第四章 对象的纹理编码,讲述变换和量化等原理和技术;它在基于内容方面的特色是,要顾及对象的不规则形状而采取一系列必要的措施。

第五章 对象运动编码,讲述运动估计和运动补偿,同时顾及对象的特点,重视整体对象多宏块同时运动的情况。

第六章 MPEG-4 编码程序结构,试图说明基于内容编码的程序设计特点。

在视频按内容编码以后,就有了按内容存储、按内容检索的前提和基础,就能实现规范性的基于内容检索和应用。这是全世界多媒体领域众望所归的事情。

第一章 基于内容处理概述

本章从分析基于内容处理的优越性入手,阐述了基于内容处理做法是发展趋势;接着分析了视频内容的表现形式和视频内容的分类,这样就有了基于内容编码的认识基础。本章的内容着重在第二节,首先简介了基于内容编码的功能特点和技术特点,然后概述了基于内容编码的体系结构和逻辑层次,使读者能在基于内容编码方面建立一个整体概念和思路。

1.1 信息与内容

1.1.1 基于内容处理的优越性

信息是由对具体的事或物的描述形式演变升华形成的,它是对具体事和物的反映,而这些具体的事和物就是内容,就是信息所表达的本质内容。

在任何场合,总是有要刻意描写及与其配套的内容,内容既定,反映内容的信息则随之确定。当然描述内容的新信息仍会不断发生,这有两方面原因:一是原有信息不完整,对既定内容的有些侧面没有认识到,信息原本缺少对内容的表达或描述;二是内容本身的事或物在发展变化,导致新信息产生。

信息随内容而确定,但表达和反映信息的方式(或曰关于信息的载体)是可以有多种的,用不同的载体传达信息其导致的数据量会有很大差异。例如对“珞珈山”这座山,可以用照片反映,可以用写在纸上的文字描述,可以用某个人讲话的语音进行介绍,可以用一段视频录像来说明,还可以创作一段惟妙惟肖的动画来表达,它们使用了不同的信息载体,都传达了关于珞珈山的信息,但针对的对象就是珞珈山,而且所有这些载体用不同形式传达的关于珞珈山的信息都应是准确的、相互没有信息偏差的,只是使用数据多寡的不同和描述的优劣程度不等而已。这就是多媒体,它们表达同样的内容,但使用不同的形式和不等量的数据。

上帝赋予人们一双眼睛、一对耳朵,使人们从小即习惯于眼观六路、耳听八方,天生具有用视频、音频从外界获取信息的本领;只是在上学以后慢慢学会用文字表达,或者学会用多种文字表达;在掌握计算机以后则更学会了图形、动画、超文本等更丰富的多种表现形式。但人是有理智的,不会被这多种形式所迷惑,人们仍会回到这些载体所传达信息的本质——内容上来,或者说人们总应紧紧围绕着内容来描述、传达。

人们终于发现,在信息表达时如果充分利用内容本身,不仅直观,而且数据量会小得多。

例如声音,姑且不说它所讲述的事或物是什么,仅就它是对文字的口头传达而言,它是表达文字的声波,可以说其中传达的文字就是声音讲述的内容。一个标准速率的文字发声约每分钟 120 个汉字,但就声音而言,1 分钟时间长度声音的数据量,按 24kHz 速率采样,按单字节的精度进行量化,则所需要的数据量是:

$$60 \text{ 秒} \times 24\,000 \text{ Hz/秒} \times 8 \text{ bit/Hz} = 11.52 \text{ Mbit} = 1\,440 \text{ K 字节};$$

若按高质量要求进行声音处理,则应使采样速率为 48kHz,采样精度按双字节,那么 1 分钟下来数据量就达到 46.08Mbit。

但如果按声音所表达的内容——汉字来计算,每个汉字占 2 个字节,120 个汉字共占 240 个字节,也就是 1 分钟下来的数据量仅有 1.12Kbit。

两方面在数据量上相差 4 万 1 千余倍。

又例如对视频基于内容处理情况。

设想一段长 1 分钟的视频图像,记录两个小孩在青青的大草原上牧马以及他们俩在草地上摔跤的情况,这其中的对象可以划分为:草原背景、放牧的马群、两个小孩及摔跤过程。

如果对视频图像按 25 帧/秒(PAL 制式)捕获,帧分辨率按 720×576 ,颜色分辨率按 3 色单字节,则 1 分钟的数据量为:

$$60 \text{ 秒} \times 25 \text{ 帧/秒} \times 720 \times 576 \text{ 像素/帧} \times 3 \text{ 字节/像素} = 1\,866.24 \text{ M 字节};$$

折合成数据则为 14.92992Gbit。

对此数据量存储(未压缩)时需要占据两张光盘。

按照基于内容处理的方式则是:将活动对象“两个小孩和放牧的马群”分别提取出来,草原成为图像的背景,小孩、马群、背景就是该段视频图像的内容。由于背景基本不变,在这 1 分钟的 1 500 帧过程中,只需第一帧传送,其他帧不再重复传送,节省了大量数据量;对马群提取以后单独编码传送,在不同帧的编码中只记录马的形态变化,而在本段视频中马群不是主要关心的对象,所以对其编码只达到常规活动可辨,而不刻意加强描述信息;对小孩也单独编码,但小孩摔跤过程是本段视频中最被关心的内容,所以拟对其加强时间域扩展编码和空间域扩展编码,时间域扩展编码是在关键动作时刻增加插入一些中间帧,使看起来其动作更加连贯清晰;空间域扩展编码是对诸如面部表情等部位扩展分辨率予以细致的描述。

这样的基于内容视频图像编码使数据量压缩效果比其他压缩方法扩大几倍,即压缩后数据率更少,但对图像的表现效果却好得多,因为这种方法刻意加强了对人们最关心内容的描述和表达。

1.1.2 视频内容的表示

内容——从字面上看,它是对事物实质内涵的表达。

对于视频来说,其表达的内容可以分为活动对象、背景和明显区别于背景的静止对象;图像是瞬间锁定的视频,是视频的特定帧。

对于对象来说,反映它的内容包括形状、颜色和纹理,它们组成了反映视频和图像最基本的内容。

纹理——对象(无论活动对象还是静止对象)表面所具有的明确特性,反映纹理特征的

有:粗糙度、对比度、方向性、线性度、规整度。

颜色——包括彩色和灰度,都用于表示对象的内容。彩色被分解成红、绿、蓝三种基色,每种基色的不同大小数据又反映出不同深浅的颜色,不同基色数据的组合又形成不同的颜色,所以造就了颜色内容非常丰富。

形状——对象的形状描述可分成两种方式:边界的形状和区域的形状。边界是指形状的外轮廓;区域则是指整个二维平面形状区域。对于形状的描述,在有些应用场合希望能在形状平移、旋转、缩放时保持不变。从这个意义上来说傅立叶描述子和不变矩描述方法是相对突出的描述方法,傅立叶描述子是对边界作形状描述,不变矩则还可对区域进行矩的描述。

1.2 基于内容编码概述

1.2.1 功能特点^{[2][3]}

在多媒体技术系列国际标准中,从 MPEG-4 开始就基本遵循基于内容的处理模式,MPEG-4 编码主体上就属于基于内容的编码,与其他原已成形的编码标准相比,有以下几个方面突出的特点。

(1) 基于内容的处理特性

基于内容即针对视频或图像中具体的活动对象和静止对象进行处理,可为对象进行单独的操作或数据流编辑,进行增强或抽取,特别进行分层和可扩展性编码。

(2) 在时间域对内容的处理

在一段或一定的时间间隔内可按对象进行存取,对一个时间段的视、音频序列以视、音频对象为目标进行检索或“快进”搜索。

(3) 直接对自然视频合成数据

即在按对象特征对自然视频分段的帧序列里增加合成数据(如字幕、文本、图形等),使其有效结合,并不交互性操作。

(4) 增加对内容的多描述性处理

在对某一对象或某一景物的处理中,可增加对其多视角描述内容的处理,也可添加多声道伴音的编码并进行相应的视听同步。在这个意义上的处理,意味着可对 MPEG-4 流数据进行三维自然景物或活动对象的描述和处理,在音频上增加多声道的听觉效果。

(5) 增高的数据编码压缩效率

由于 MPEG-4 基于内容处理的特性,使背景与对象分开编码,使对象具有形状编码,所以可比其他处理标准有更高的数据压缩比率和压缩效果;同时它具备对内容局部仔细描述增加扩展分层的处理机理而又能达到更好主观视觉质量的效果。

(6) 基于内容的分段可变性

由于该系统是基于内容可分开处理的,所以可根据需要对不同内容、不同对象分配不同优先级,对重要的或特定的对象可用较高的空间分辨率和时间分辨率表示,对第二级的或更低级的内容和对象用允许低一些的空间分辨率和时间分辨率表示,使系统整体数据率能适应带宽变化,以便更有效地利用传输资源和存储资源。

1.2.2 MPEG-4 编码特点

(1) 编码码率

MPEG-4 对视音信息编码的结果数据率支持在一个带宽范围内可适应性变化,使总带宽为 5Kbps ~ 4Mbps;其中超低码率范围 5Kbps ~ 64Kbps 适合于 CIF 格式以下分辨率且帧率在 15FPS 以下的应用场合,64Kbps ~ 4Mbps 适合于其他大多数应用场合,且与 MPEG-1, MPEG-2 等自然兼容;可以看出其上限 4Mbps 远低于 MPEG-2 标准的上限,但图像质量却不低且更能灵活控制。

(2) 编码技术

MPEG-4 实现对视频和图片基于内容的编码,适应视频、图片基于内容的可伸缩性,在空域和时域对图像质量的可伸缩性,可实现对视频序列和各种图片可扩展操作处理,也可实现针对内容对象的随机访问。

MPEG-4 算法工具的积累经历了大量的选择和试验:运动估计方面,使用了全局运动补偿、2D 三角网格预测和亚像素预测;纹理编码方面,使用了小波变换、3D DCT、高级帧内编码、重叠变换;任意形状区域纹理编码方面,使用了形状自适应 DCT、贴补 DCT、中值替换 DCT、延拓/内插 DCT、小波子带编码;形状编码方面,使用了自适应形状区域分割、可变块大小分割、几何变换等方法。

MPEG-4 编码技术着重体现在形状编码、运动编码、纹理编码、分级编码、基于 VOP 编码等几个方面,所有这些技术在后续章节中都将进行详细讨论。

1.2.3 基于内容编码的体系结构^{[1][4]}

为了实现对视频图像基于内容的表示,引入视频对象的概念,在一定长度的视频段中,很有可能描述的是几个相同的对象,可以对这几个对象分别予以分割提取,得到关于各个对象的视频段,但实际分割过程并非如此。

图 1.1 是基于内容的视频编码框图,图中示出对每一个视频帧进行若干个 VOP 提取, VOP 的个数应该是该帧中可能划分的对象的个数,但人们一般只对活动的对象提取,静态对象常被合并与背景同样对待,这样做的原因是有利于对象提取。

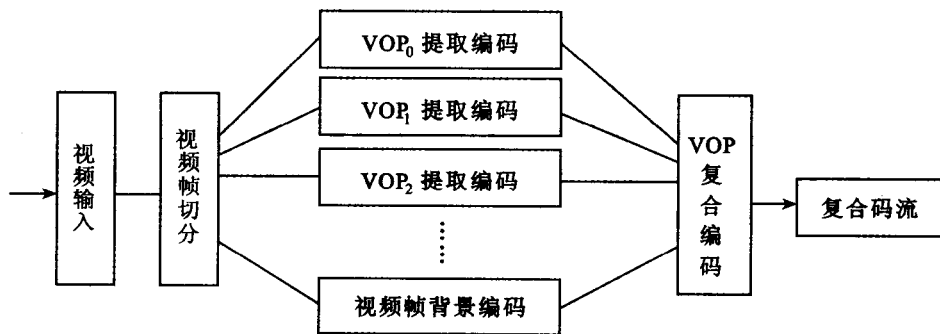


图 1.1 基于内容的视频编码框图

这是由于人们在实际成像时是利用摄像头一次又一次曝光而得到一幅一幅视频帧的,

对这些图像帧按时间顺序连续播放出(或连续捕获)才得到视频。所以人们要对视频进行基于内容的表示,只能立足于一幅一幅的视频帧,在每一帧中分割提取的只能是视频对象平面(VOP),对连续视频中关于同一对象的VOP再组成对该对象的动态描述即视频对象(VO)。

在对视频帧中各个VOP提取后进行编码表示时,包括VOP的运动信息编码、形状信息编码、纹理信息编码。如果能对视频图像进行分析,提取关于对象的视频图像模型,使得关于该对象的形状、纹理和运动能遵循该模型的规则,则编码会变得容易,编码结果的数据量也会进一步减小。

图1.2是基于内容的视频解码框图,解码器从收到的视频数据流中解读出关于每个 VOP_i 的数据,由运动规则及其形状和纹理的描述恢复出一个个对象平面,再将这些对象平面与同步的背景按位置分布组合即形成完整的视频帧,对这些视频帧按时间顺序展示即恢复原视频序列,完成视频解码。

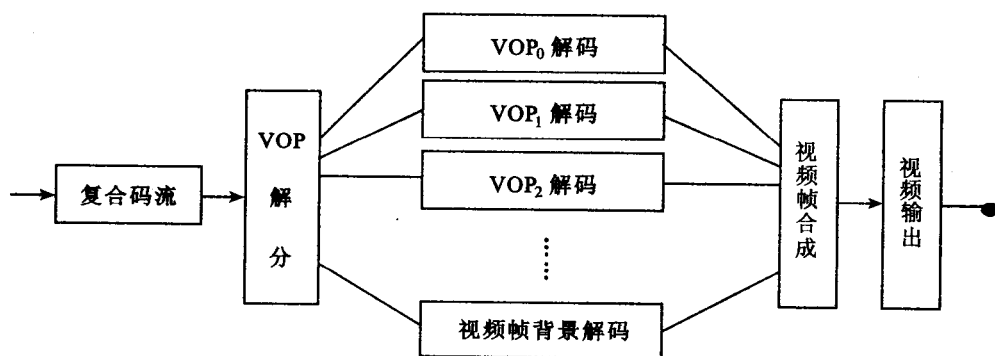


图 1.2 基于内容的视频解码框图

对视频序列编码所得到的视频流数据,其数据结构可以分以下层次进行理解,结合图1.3予以说明。

常规的完整视频序列可以按镜头切分情况划分成不同的视频段 $VS_0, VS_1, \dots$ 等;每个视频段由若干个视频帧组成,如 $VF_{01}, VF_{02}, \dots, VF_{0m}$;每个视频段中所含视频对象的个数常是不变的;

如 VS_0 视频段中被提取出 $VO_0, VO_1, \dots, VO_e$ 共 e 个视频对象VO;而视频段中每个视频帧 $VF_{01}, \dots, VF_{0m}$ 等,均被分割提取出若干个视频对象平面,如视频帧 VF_{01} 被分割出 $VOP_{01}, \dots, VOP_{e1}$ 等,而同一段视频各个帧中关于同一视频对象的VOP联合组织成VO,描述了该对象的独立活动情况,如图1.3中 $VOP_{01} \sim VOP_{0m}$ 联合描述了 VO_0 , $VOP_{e1} \sim VOP_{em}$ 联合描述了 VO_e 等。

在扩展视频编码时,会增加对其中某些视频对象细节的描述,得到多于常规划分的视频对象平面数目,如图中对 VO_0 扩展到 M 个VOP,对 VO_1 扩展到 N 个VOP,对 VO_e 扩展到 P 个VOP,这里 M, N, P 等都大于 m ,但 M, N, P 相互之间可以不等。

如为了增加对某个视频对象的运动细节描述,可以在时域扩展插入一些关于该对象的平面。为了增加对某个视频对象的纹理细节描述(如对视频会议系统中会议主席面部表情