

MPEG视频编码II



讲师：肖俊

软件与技术学院

1. MPEG-4

1.1 概述

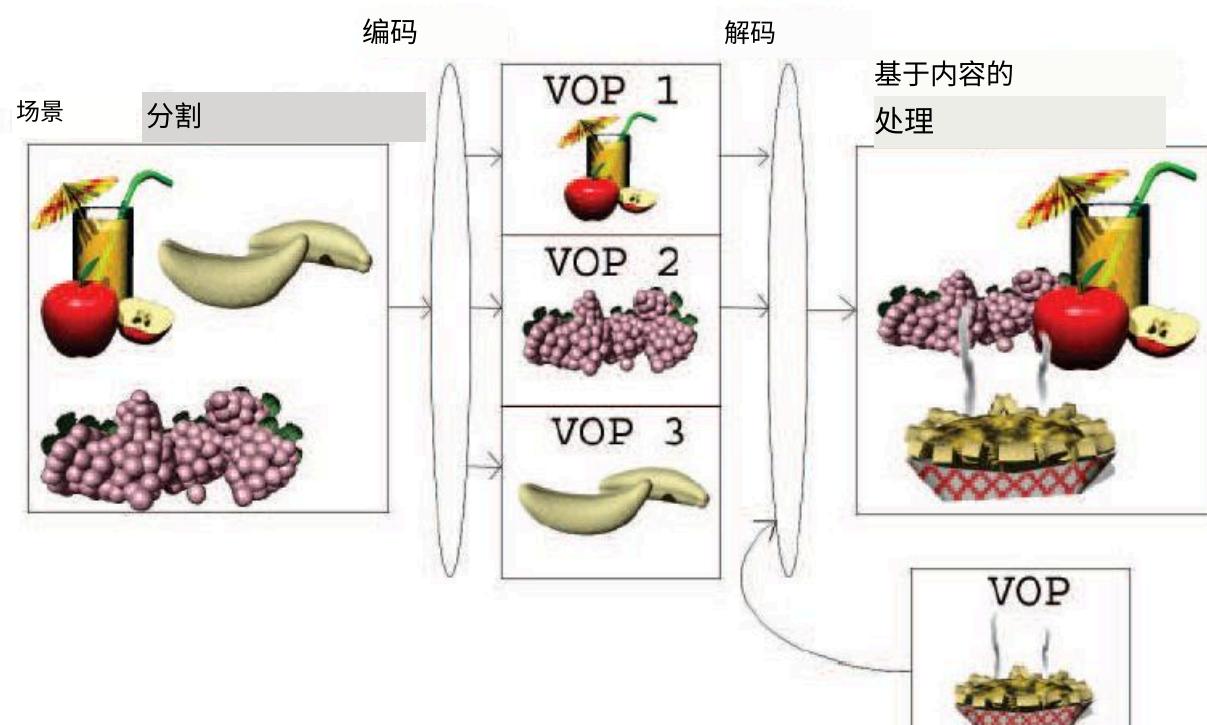


图12.1：MPEG - 4视频的组成与处理。

内容

- MPEG-4

- MPEG - 4概述
- MPEG - 4中的基于对象的视觉编码
- MPEG - 4中的合成对象编码
- 对象类型、配置文件和级别
- MPEG - 4第10部分/H.264

- MPEG-7

- MPEG-21

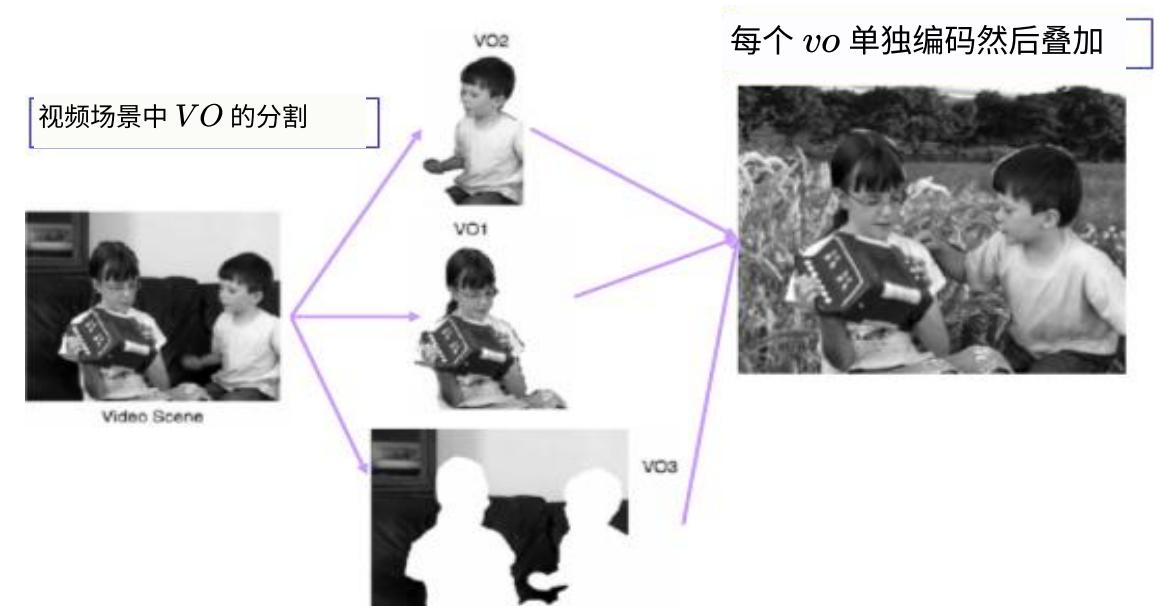
Fundamentals of Multimedia —— MPEG Video Coding II (2025 Spring)

2

1.1 概述

- MPEG - 4：一个较新的标准。除了压缩之外，还非常关注用户交互性问题。
- MPEG - 4在采用一种新的基于对象的编码方式上与其前身不同：
 - 提供更高的压缩比，也有利于数字视频合成、处理、索引和检索。
 - 任意形状编码
 - 静态纹理编码
 - 面部对象编码与动画
 - 身体对象编码与动画
- MPEG - 4视频的比特率范围现在涵盖了从5 kbps到10 Mbps的较宽区间。多媒体基础 - MPEG视频编码II（2025年春季）

1.1 概述



MPEG - 4 基于对象的处理

1.1 概述

- MPEG - 4 (图12.2(b)) 是一个全新的标准，用于：

(a) 组合媒体对象以创建理想的视听场景。

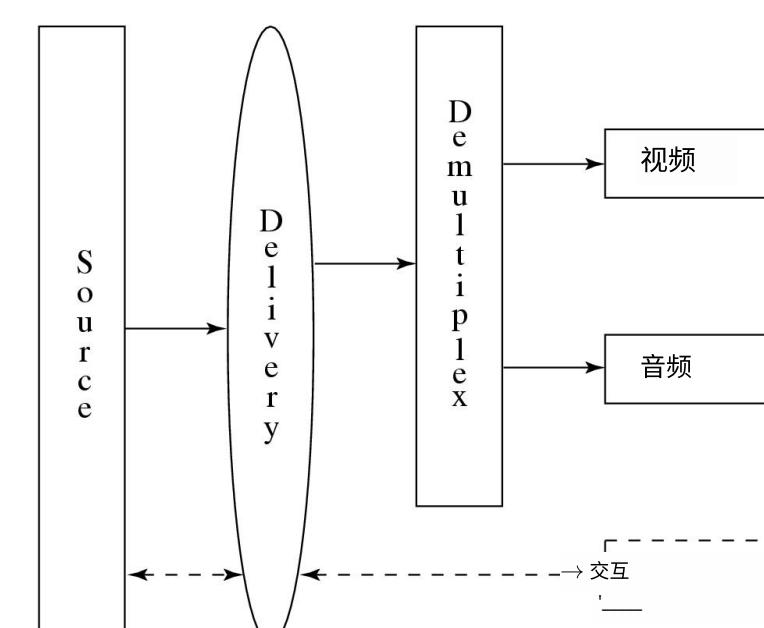
(b) 对这些媒体数据实体的比特流进行复用和同步，以便它们能够以有保证的服务质量 (QoS) 进行传输。

(c) 在接收端与视听场景进行交互——为音频和视频压缩提供一套高级编码模块和算法工具箱。

1.1 概述

- MPEG标准中交互性的比较。

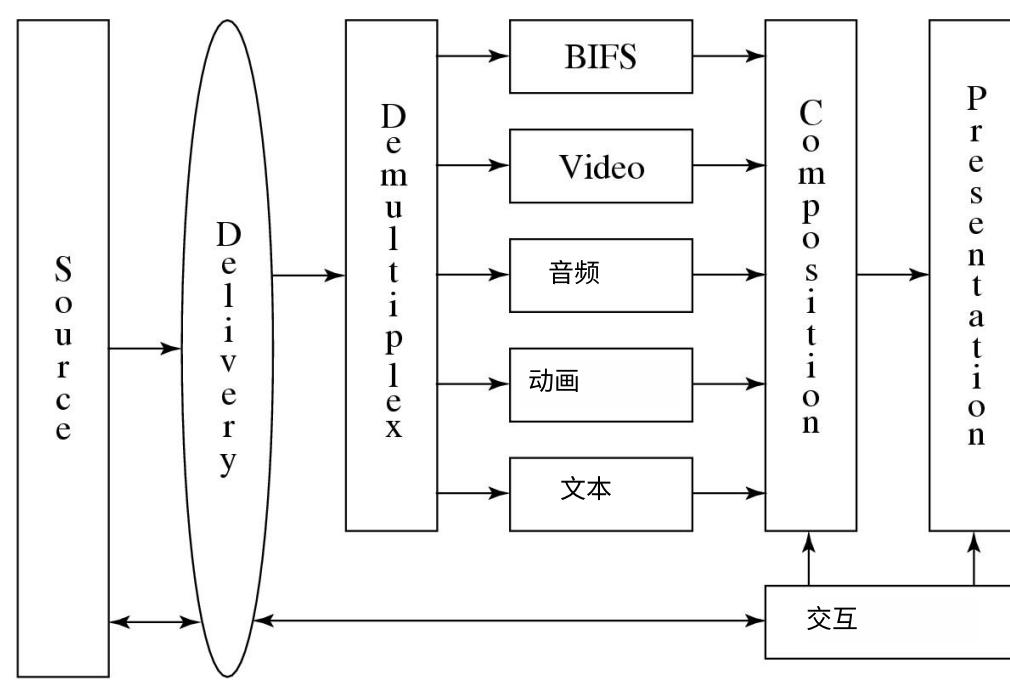
- MPEG - 1和2中的参考模型



虚线交互仅MPEG - 2支持

1.1 概述

- MPEG标准中交互性的比较。- MPEG - 4参考模型



1.1 概述

- MPEG - 4视觉比特流的层次结构与MPEG - 1和MPEG - 2有很大不同，它非常面向视频对象。



图12.3：MPEG - 4视觉比特流中场景的面向视频对象的分层描述。
多媒体基础 - MPEG视频编码II（2025年春季）

1.1 概述

1. 视频对象序列 (VS) ——呈现完整的MPEG - 4视觉场景，该场景可能包含二维或三维自然或合成对象。
2. 视频对象 (VO) ——场景中的特定对象，其形状可以是任意的 (非矩形)，对应于场景中的一个对象或背景。
3. 视频对象层 (VOL) ——提供了一种支持 (多层) 可伸缩编码的方式。在可伸缩编码下，一个视频对象 (VO) 可以有多个视频对象层；在非可伸缩编码下，一个视频对象可以只有一个视频对象层。
4. 视频对象平面组 (GOV) ——将视频对象平面组合在一起 (可选层级)。
5. 视频对象平面 (VOP) ——视频对象在特定时刻的快照。

1.2 基于对象的视觉编码

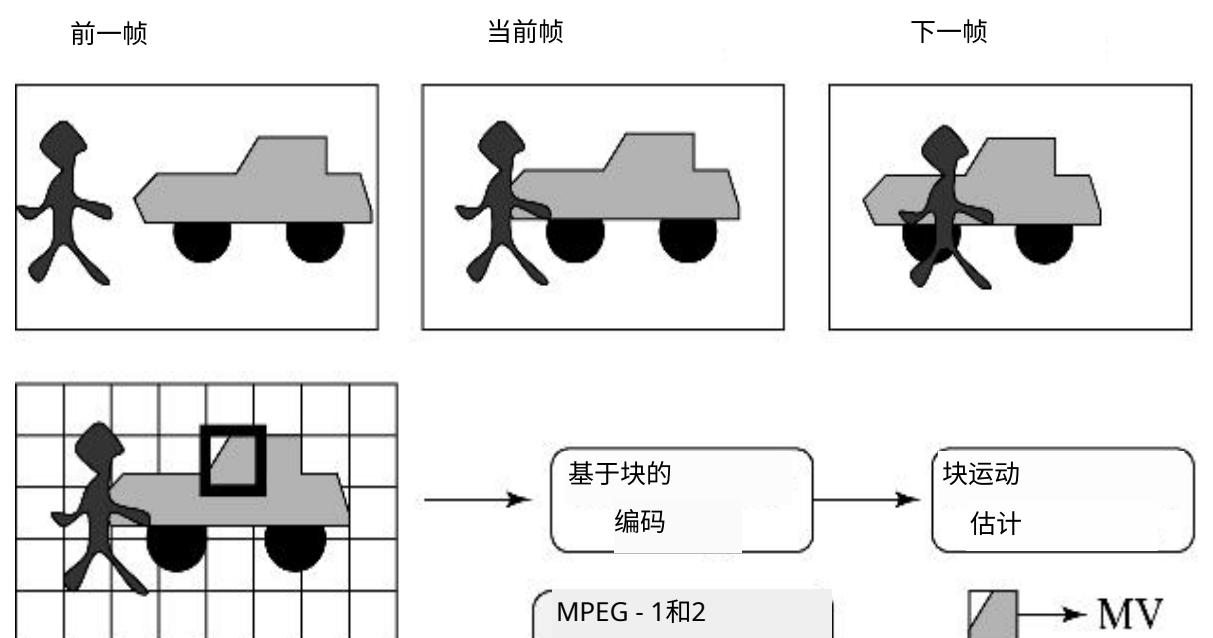
- 基于VOP的编码与基于帧的编码
- 运动补偿
- 纹理编码
- 形状编码
- 静态纹理编码
- 精灵编码
- 全局运动补偿

1.2.1 基于视频对象平面 (VOP) 的编码与基于帧的编码

- MPEG - 1和MPEG - 2不支持视频对象平面 (VOP) 概念，因此它们的编码方法被称为基于帧的编码 (也称为基于块的编码)。

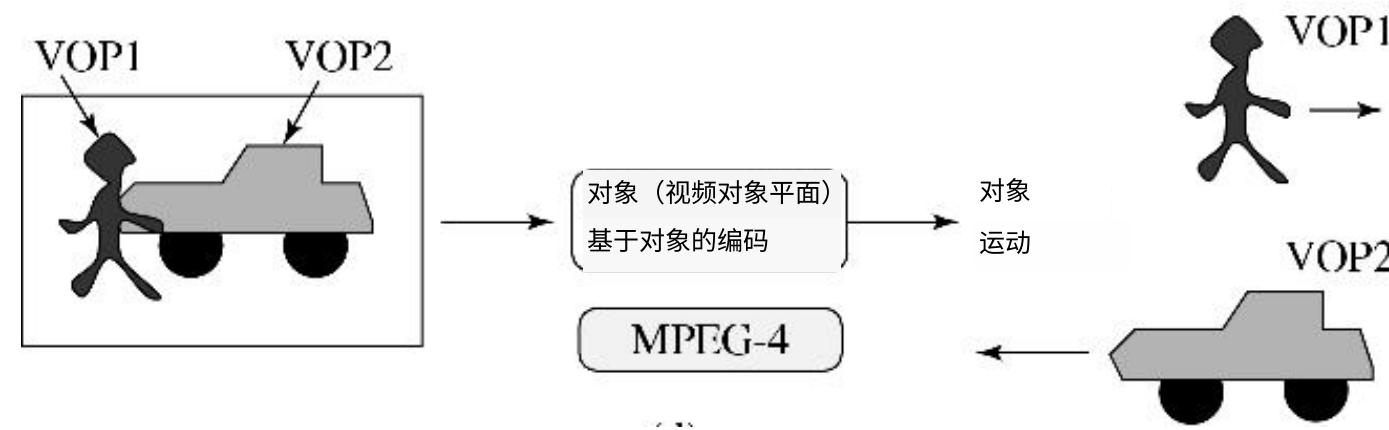
1.2.1 基于视频对象平面 (VOP) 的编码与基于帧的编码

图12.4 (c)展示了一个可能的示例，其中对于基于块的编码，两个潜在匹配都产生了较小的预测误差。



1.2.1 基于视频对象平面 (VOP) 的编码与基于帧的编码

图12.4 (d)表明，每个视频对象平面 (VOP) 具有任意形状，并且理想情况下将获得与实际对象运动一致的唯一运动矢量。



基于视频对象平面 (VOP) 的编码

- 基于MPEG - 4视频对象平面 (VOP) 的编码也采用了运动补偿技术：- 帧内编码的视频对象平面称为I - VOP。

•如果仅采用前向预测，帧间编码的视频对象平面称为P - VOPs；如果采用双向预测，则称为B - VOP。

•视频对象平面 (VOP) 面临的新难题：可能具有任意形状，除了视频对象平面的纹理信息外，还必须对形状信息进行编码。

注意：此处的纹理实际上指的是视觉内容，即VOP中像素的灰度级 (或色度) 值。

1.2.2 运动补偿

• MPEG - 4 中基于运动补偿 (MC) 的视频对象平面 (VOP) 编码再次涉及三个步骤:

- (a) 运动估计。
- (b) 基于运动补偿的预测。
- (c) 预测误差编码。

• 在运动补偿中, 仅考虑当前 (目标) 视频对象平面内的像素进行匹配。

• 为便于进行运动补偿 (MC), 每个视频对象平面 (VOP) 被划分为多个宏块 (MB)。默认情况下, 亮度图像中的宏块大小为 16×16 , 色度图像中的宏块大小为 8×8 。

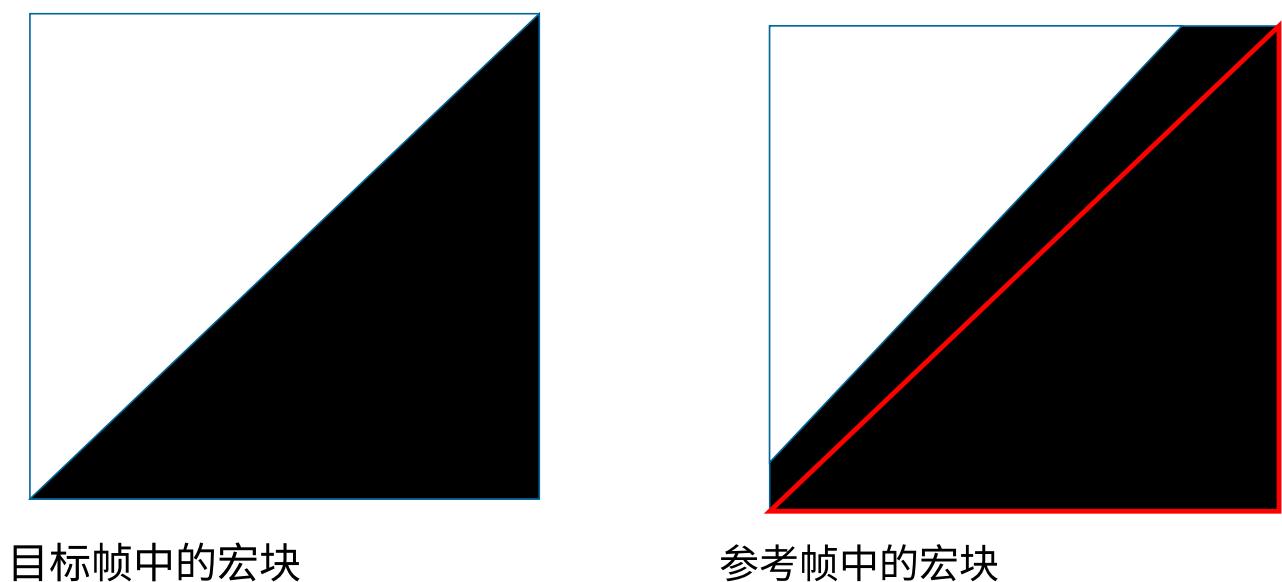
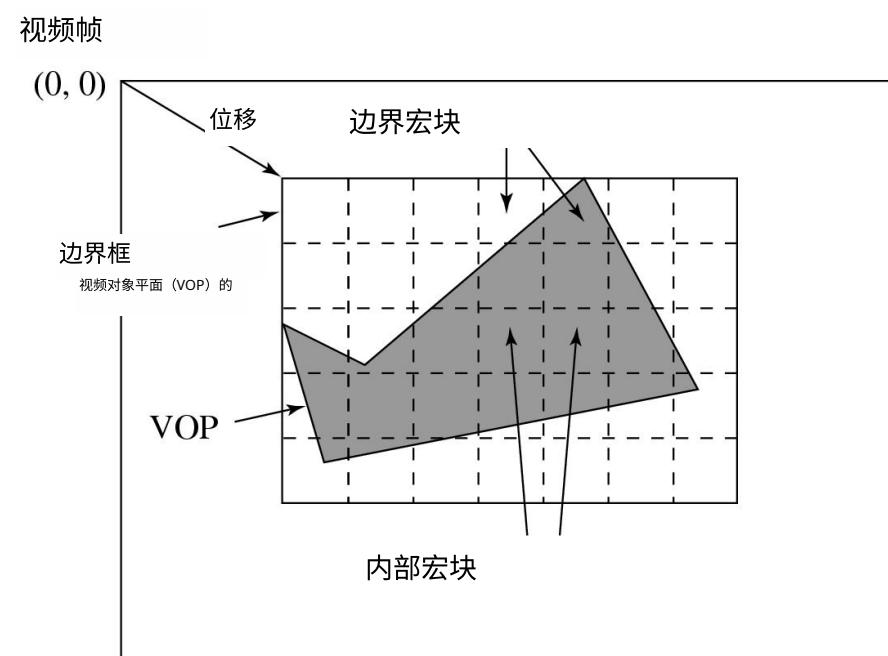
1.2.2 运动补偿

• MPEG - 4 为每个视频对象平面 (VOP) 定义了一个矩形边界框 (详情见图 12.5)。

• 完全位于 VOP 内部的宏块称为内部宏块。跨越 VOP 边界的宏块称为边界宏块。 - 为了帮助匹配目标 VOP 中的每个像素, 并满足变换编码 (如离散余弦变换, DCT) 中矩形块的强制要求, 在进行运动估计之前, 会对参考 VOP 进行填充预处理步骤。注意: 填充仅在参考 VOP 中进行。

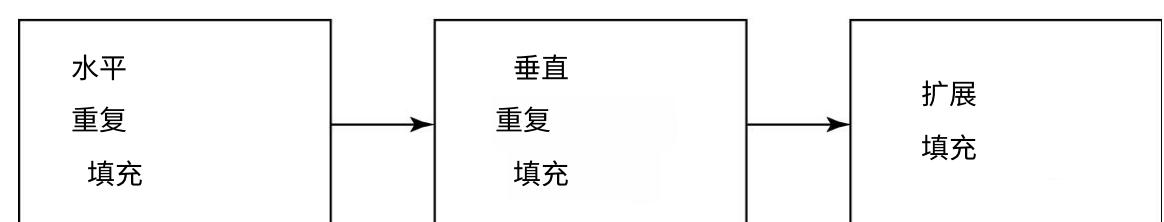
1.2.2 运动补偿

• 内部宏块的运动补偿采用与 MPEG - 1 和 2 相同的方式进行



1.2.2 运动补偿

• 在对边界宏块进行运动补偿之前进行填充。



• 水平填充会检查每一行, 每个边界像素会被复制到左侧和/或右侧, 以填充 VOP 外部的值。 - 如果区间由两个边界像素界定, 则采用它们的平均值;

• 垂直填充的工作方式类似。

1.2.2 运动补偿

算法12.1 水平重复填充: 开始对于参考VOP中边界宏块的所有行如果行中存在 \exists (边界像素) 对于VOP外部的所有区间如果区间仅由一个边界像素 b 界定将 b 的值赋给区间内的所有像素否则 // 区间由两个边界像素 b_1 和 b_2 界定将 $(b_1 + b_2)/2$ 的值赋给区间内的所有像素结束 - 后续的垂直重复填充算法以类似的方式工作。

• 填充 (示例)

(a)	(b)	(c)

• 扩展填充: 紧邻边界宏块的外部宏块通过复制边界宏块边界像素的值来填充。 - 使用的宏块遵循优先级列表: 左、上、右、下

1.2.2 运动补偿

- 运动矢量编码

- 运动估计

$$SAD(i, j) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \cdot Map(x+k, y+l)$$

$Map(p, q) = 1$ 如果 $C(p, q)$ 是视频对象平面 (VOP) 中的像素
否则 $Map(p, q) = 0$

运动矢量 $MV : (u, v) = |(i, j)| SAD(i, j)$ 最小
 $i, j \in [-p, p]$, p 是 u 和 v 中的最大值

•允许亮度分量具有四分之一像素的精度。

•运动矢量 (MV) 可以指向参考视频对象平面 (VOP) 边界之外, VOP 外部的像素在填充步骤中定义。多媒体基础 - MPEG 视频编码 II (2025 年春季)

25

1.2.3 纹理编码

- I - VOP 的编码方式与 JPEG 类似 - 对于 P - VOP 和 B - VOP

预测误差被发送到离散余弦变换 (DCT) 和变长编码 (VLC) - 基于 DCT 的纹理编码 - 对于 VOP 外部的边界宏块部分, 填充零 - 直流 (DC) 分量的量化步长为 8 - 交流 (AC) 系数可以采用两种方法 - H.263 方法, 所有系数使用相同的量化器 - MPEG - 2 方法, 同一宏块中的 DCT 系数可以使用不同的量化器。

1.2.3 纹理编码

二、基于 SA - DCT 的边界宏块编码——形状自适应离散余弦变换 (SA - DCT) 是另一种用于边界宏块的纹理编码方法。

—SA - DCT 是一种二维离散余弦变换, 它通过两次一维离散余弦变换迭代计算得到, 可作为一种可分离的二维变换。

N. 1D ($DCT - N$)

$$F(u) = \sqrt{\frac{2}{N}} C(u) \sum_{i=0}^{N-1} \cos \frac{(2i+1)u\pi}{2N} f(i)$$

一维 (IDCT - N)

$$\tilde{f}(i) = \sum_{u=0}^{N-1} \sqrt{\frac{2}{N}} C(u) \cos \frac{(2i+1)u\pi}{2N} F(u)$$

其中 $i = 0, 1, \dots, N-1; u = 0, 1, \dots, N-1$

$$C(u) = \begin{cases} \frac{\sqrt{2}}{2} & \text{if } u = 0 \\ 1 & \text{otherwise} \end{cases}$$

多媒体基础——MPEG视频编码II (2025年春季)

26

1.2.3 纹理编码

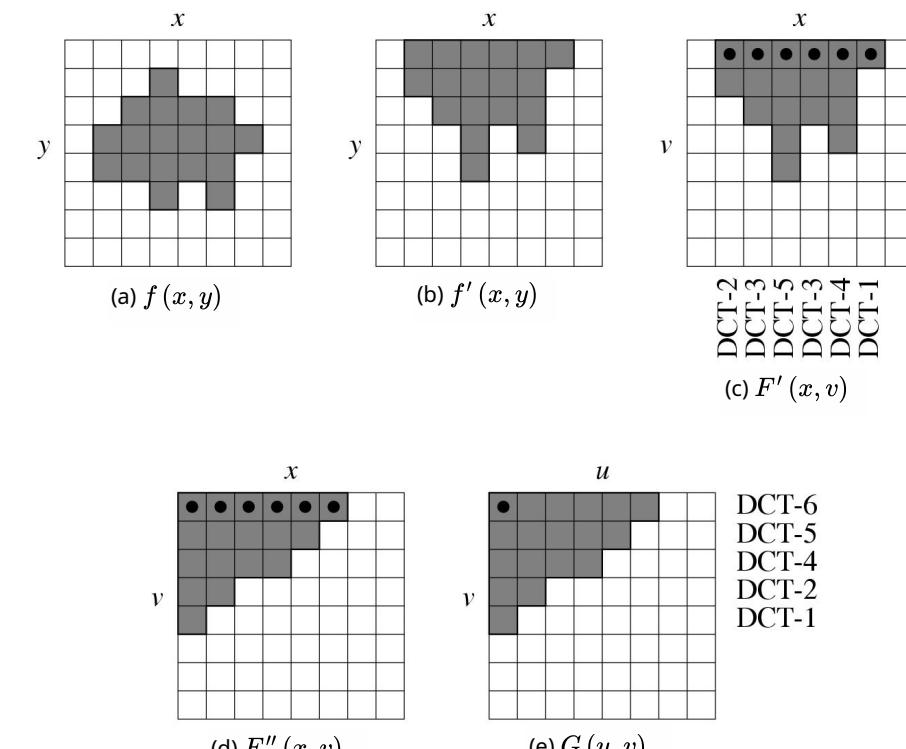


图12.8：使用形状自适应离散余弦变换 (SA - DCT) 对边界宏块进行纹理编码
形状自适应离散余弦变换 (SA - DCT)。

多媒体基础——MPEG视频编码II (2025年春季)

27

多媒体基础 - MPEG视频编码II (2025年春季)

28

1.2.4 形状编码

•MPEG - 4 支持两种类型的形状信息, 即二值和灰度。

•二值形状信息可以采用二值图 (也称为二值阿尔法图) 的形式, 其大小与视频对象平面 (VOP) 的矩形边界框相同。

•位图中的值“1”(不透明) 或“0”(透明) 表示该像素是在VOP 内部还是外部。

•或者, 灰度形状信息实际上指的是形状的透明度, 灰度值范围从 0 (完全透明) 到 255 (不透明)。

1.2.4 形状编码

• 二进制形状编码



不在视频对象平面 (VOP) 中的像素
表示为透明

(基于上下文的二进制算术编码)

多媒体基础 - MPEG视频编码II (2025年春季)

29

多媒体基础 - MPEG视频编码II (2025年春季)

30

1.2.4 形状编码

• 二进制形状编码

1. 计算当前上下文值 X 每个像素为二进制, 上下文 X 根据已编码的 10 个像素计算得出。 X 有 10 位: C9C8C7C6C5C4C3C2C1C0X 的范围是 0~1024
通过查找表 (MPEG - 4 标准) 获取相应的值。

2. 算术编码查找表中的值表示

c9	c8	c7
c6	c5	c4
c3	c2	
c1	c0	x

1024个上下文中每
个上下文的出现概
率。

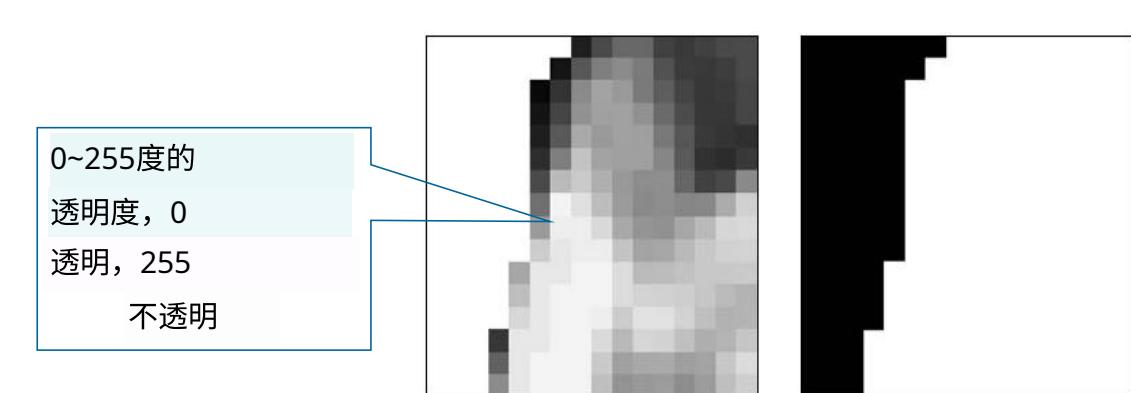
上下文 (二进制)	上下文 (十进制)	描述	P(0)
0000000000	0	所有上下文像素均为 0	65267/65535 = 0.9959
0000000001	1	c_0 为 1, 其他均为 0	16468/65535 = 0.2513
1111111111	1023	所有上下文像素均为 1	235/65535 = 0.0036

多媒体基础——MPEG视频编码II (2025年春季)

31

1.2.4 形状编码——灰度形 状编码

•灰度用于描述形状的透明度, 而非纹理。



•光栅图形使用额外的位平面来创建Alpha通道图。- 灰度形状编
码是有损的, 而二进制形状编码是无损的

多媒体基础 - - MPEG视频编码II (2025年春季)

32

1.2.5 静态纹理编码

静态物体纹理的小波编码

- 最低频率的子带使用差分脉冲编码调制（DPCM）进行编码
 - 每个系数的预测基于三个相邻系数
- 其他子带基于多尺度零树小波编码。
 - 多尺度零树具有父子关系树——所有系数的位置信息得到了更好的利用。
 - 使用量化器，并用算术编码对最重要的系数进行编码。
 - 在下一次迭代中对差值进行编码，此次迭代采用较小的量化器。

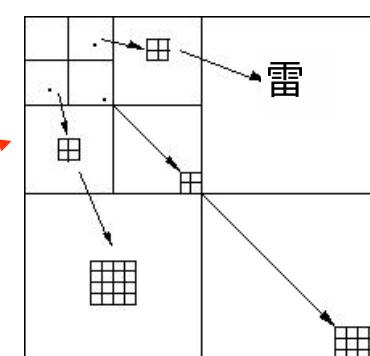
1.2.5 静态纹理编码

零树

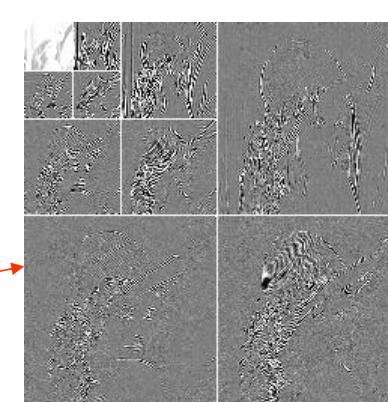


原始图像

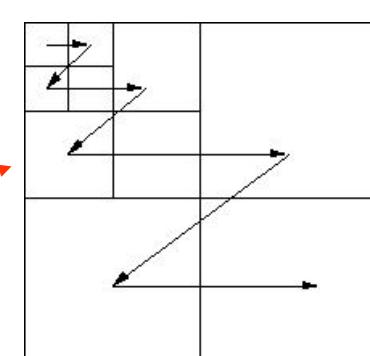
结构



三级
分解



扫描
顺序



多媒体基础——MPEG视频编码II (2025年春季)

33

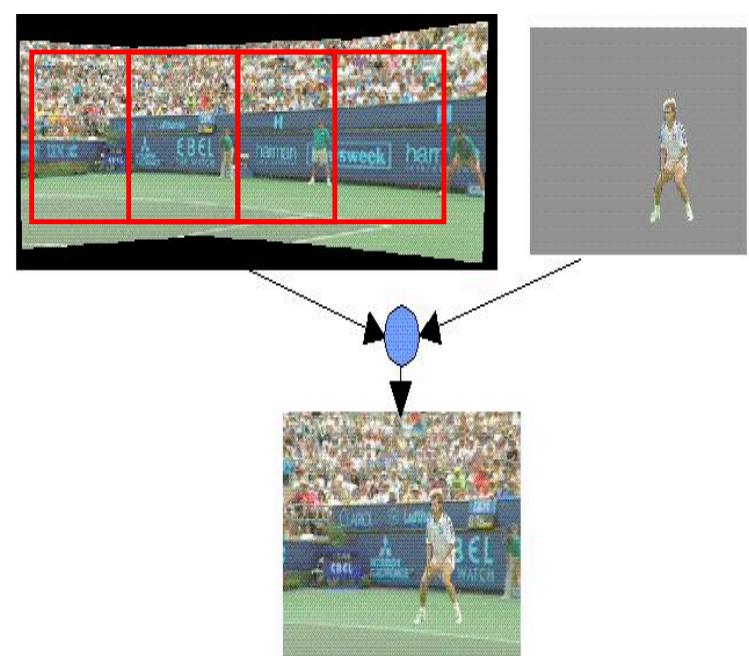
34

1.2.6 精灵编码

一些背景可以视为静态

图像

背景受摄像机移动的影响



背景可以单独编码，前景对象可用于创建基于对象的灵活MPEG - 4视频合成

1.2.7 全局运动补偿

- 诸如平移、倾斜、旋转和缩放等相机运动通常会导致连续帧之间的内容快速变化，基于块的运动补偿在此情况下并非有效方法
- 全局运动补偿是更好的选择
- 全局运动补偿有四个主要组成部分

- 全局运动估计
- 变形与融合
- 运动轨迹编码
- 局部运动补偿 (LMC) 或全局运动补偿 (GMC) 的选择

多媒体基础——MPEG视频编码II (2025年春季)

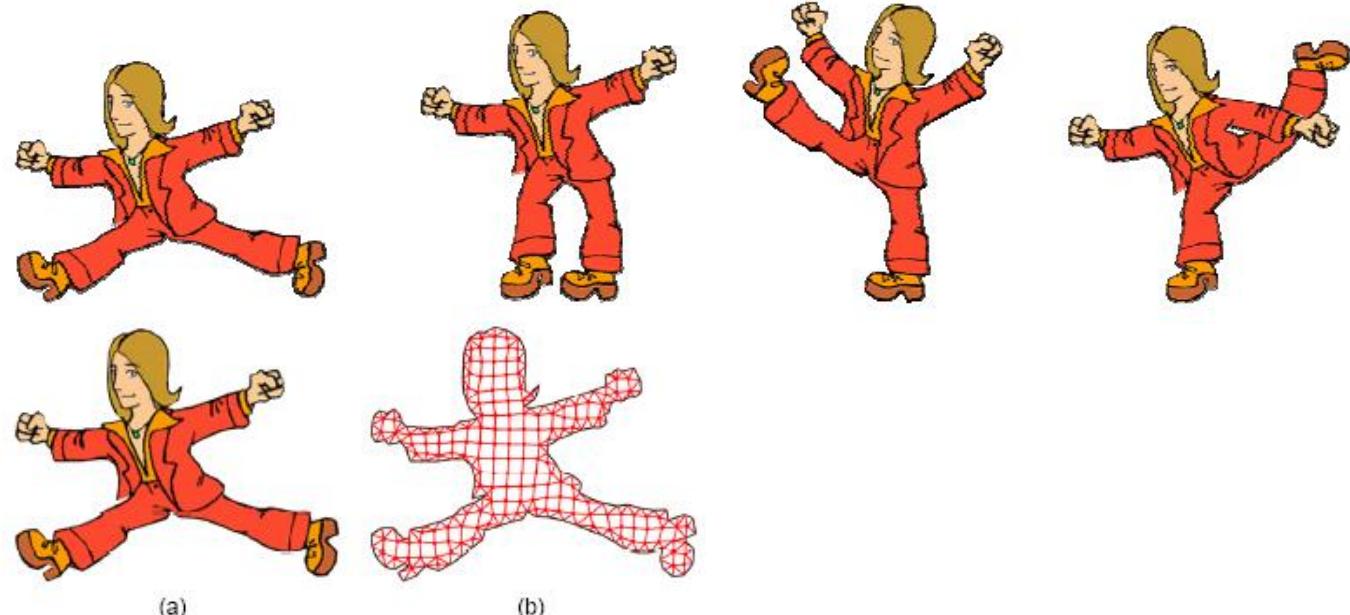
35

36

1.3 合成对象编码

合成对象：使用计算机创建的对象

二维网格对象编码：使用多边形面片对二维平面区域进行细分（或划分）：



多媒体基础 - - MPEG视频编码II (2025年春季)

36

1.3 合成对象编码

多边形的顶点被称为网格的节点。

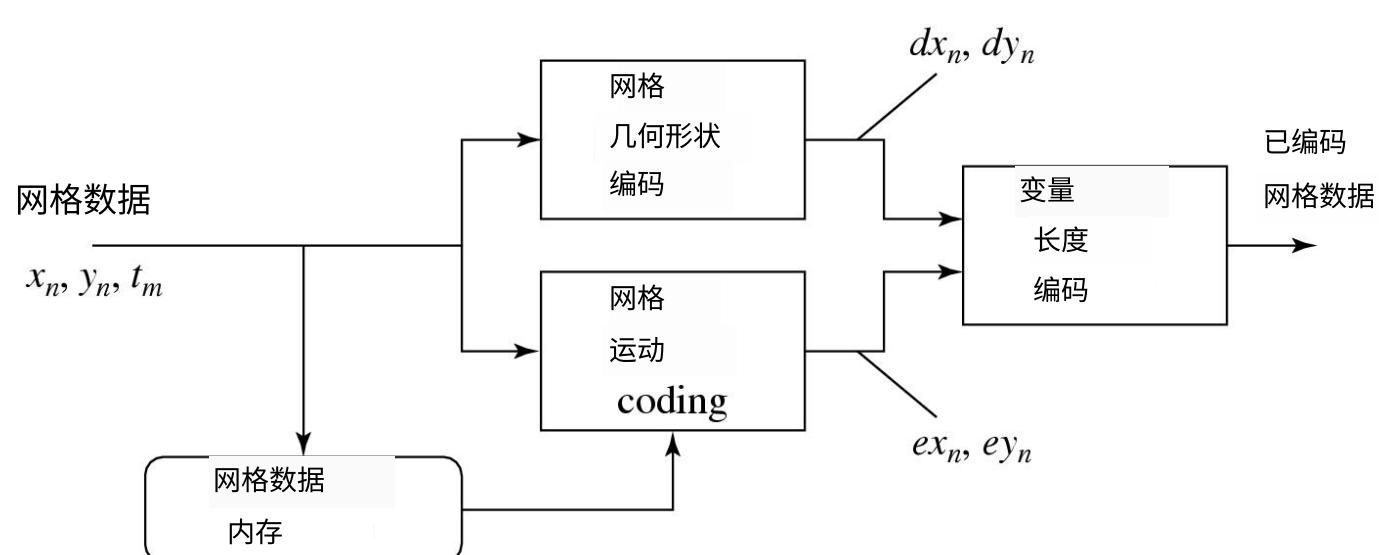
- 最常用的网格是三角形网格，其中所有多边形都是三角形。
 - MPEG - 4标准使用两种类型的二维网格：均匀网格和德劳内网格
 - 二维网格对象编码紧凑。网格的所有坐标值均以半像素精度编码。
 - 每个二维网格都被视为一个网格对象平面 (MOP)。

多媒体基础 - - MPEG视频编码II (2025年春季)

38

38

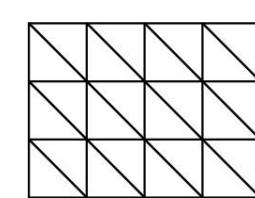
1.3 合成对象编码



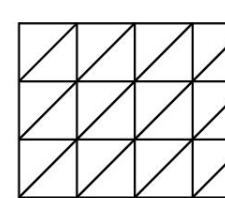
1.3 合成对象编码

二维网格对象编码

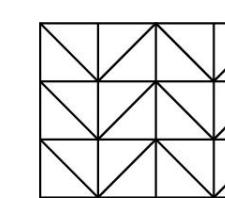
二维网格几何编码



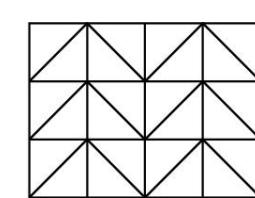
(a) 类型0



(b) 类型1



(c) 类型2



(d) 类型3

四种均匀网格

德劳内网格是一种更好的对象网格表示 - 选择网格的边界节点 - 选择内部节点 - 执行德劳内三角剖分

多媒体基础 -- MPEG视频编码II (2025年春季)

39

40

图12.11: 二维网格对象平面 (MOP) 编码过程

1.3 合成对象编码

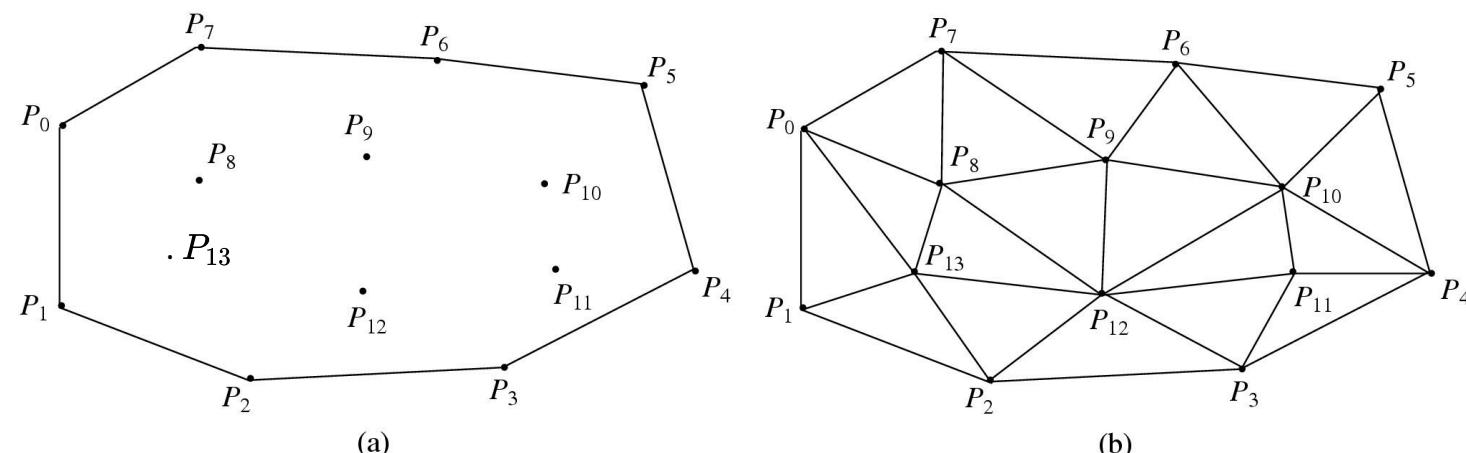


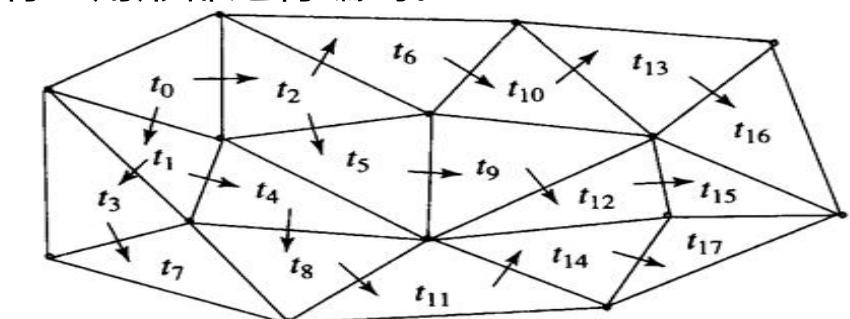
图 12.13：德劳内网格：(a) 边界节点 (P_0 至 P_7) 以及内部节点 (P_8 至 P_{13})。(b) 通过约束德劳内三角剖分得到的三角网格。

1.3 合成对象编码

•二维网格对象编码

•二维网格运动编码

- 对于任何MOP三角形i、j、k。如果已知i、j的运动矢量，则k的运动矢量可预测为
$$\text{Pred}_k = 0.5 (\text{Pred}_i + \text{Pred}_j)$$
- 当对一个三角形的运动矢量进行编码时，对与前一个三角形共享一条边的相邻MOP三角形的未编码顶点进行编码，依此类推，直到对所有三角形都进行编码。

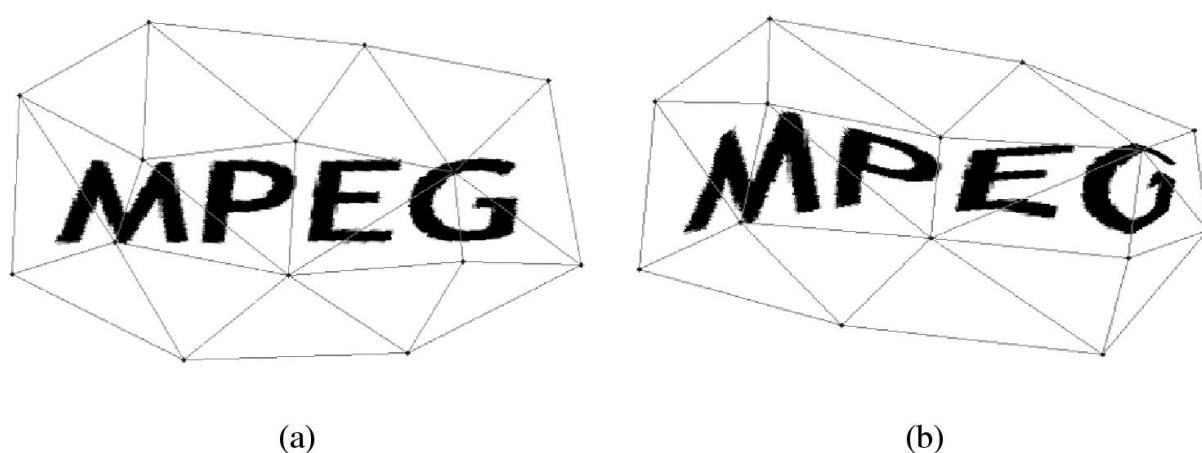


1.3 合成对象编码

•二维网格对象编码

•二维对象动画

- 上一步在参考MOP中的网格三角形与目标MOP之间建立了——映射——使用仿射变换来实现动画序列



(a)

(b)

- ### •由于人脸和人体在视频中频繁出现，MPEG - 4为面部对象和人体对象定义了特殊的3D模型。

- 这些新视频对象的一些潜在应用包括电话会议、人机界面、游戏和电子商务。

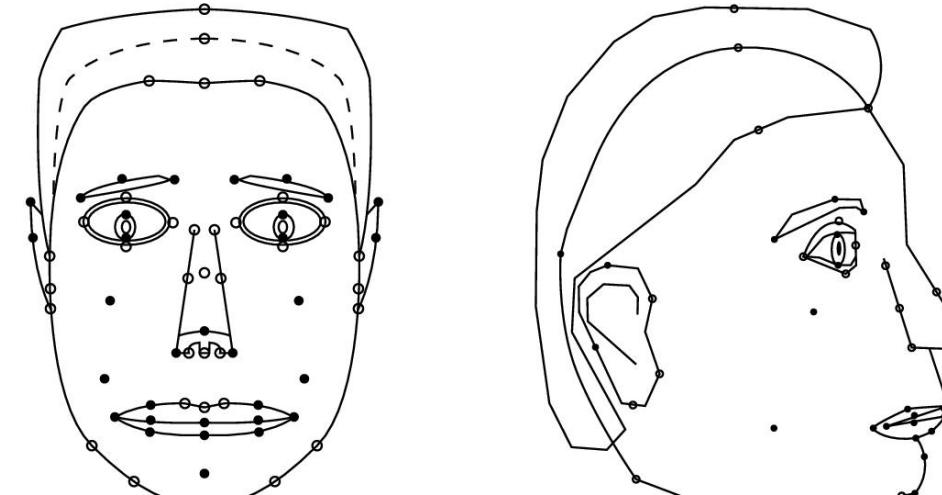
- MPEG - 4超越了线框模型，因此面部或身体对象的表面可以进行着色或纹理映射。

1.3 合成对象编码

•面部对象编码与动画

- MPEG - 4采用了由VRML联盟开发的通用默认面部模型。
- 可以指定面部动画参数 (FAPs) 来实现理想的动画效果——与原始“中性”面部的偏差。
- 此外，可以指定面部定义参数 (FDP) 以更好地描述个体面部特征。
- 图12.16展示了面部定义参数 (FDP) 的特征点。可能受动画影响的特征点 (FAP) 用实心圆表示，不受影响的特征点用空心圆表示。

1.3 合成对象编码



(a)

(b)

图12.16：面部定义参数 (FDP) 的特征点。（未显示牙齿和舌头的特征点。）

1.3 合成对象编码

人体对象编码与动画

- MPEG - 4 版本 2 引入了人体对象，这是对面部对象的自然扩展。
- 与虚拟现实建模语言 (VRML) 联盟中的类人动画 (H - Anim) 小组合作，采用了具有默认姿势的通用虚拟人体。
- 默认姿势是双脚朝前站立、双臂下垂且掌心向内的姿势。

1.3 合成对象编码

- 大量的身体动画参数 (BAPs) 用于描述连接不同身体部位的关节角度：脊柱、肩部、锁骨、肘部、腕部、手指、髋部、膝部、踝部和脚趾——使身体具有186个自由度，仅每只手就有25个自由度。
- 某些身体动作可以在多个细节级别上进行指定。

- 对于特定的身体，可以为身体尺寸、身体表面几何形状以及可选的纹理指定身体定义参数 (BDPs)。

1.4 MPEG4对象类型、配置文件和级别

- 与MPEG2一样，MPEG4定义了许多配置文件和级别：视觉配置文件、音频配置文件、图形配置文件、场景描述配置文件、对象描述符配置文件

MPEG4定义
创建视频对象所
需的工具以及它
们在场景中组合
的方式

工具	对象类型					
	简单	核心	主要	简单可扩展	N位	可扩展
基于基本运动补偿的工具	*	*	*	*	*	
B-VOP		*	*	*	*	
二值形状编码		*	*			*
灰度形状编码			*			
精灵图			*			
隔行扫描			*			
时间可伸缩性 (P-VOP)		*	*		*	
空间和时间可伸缩性 (矩形VOP)				*		
N位					*	
可伸缩静态纹理						*
容错能力	*	*	*	*	*	

1.4 MPEG4对象类型、配置文件和级别

- 不同配置文件中的对象类型和级别

配置文件	级别	典型数量 图片对象 大小	比特率		最大值 of
			(比特/秒)	of	
简单	1	176 × 144 (四分之一通用中间格式)	64K	4	
	2	352 × 288(通用中间格式)	128K	4	
	3	352 × 288(通用中间格式)	384K	4	
核心	1	176 × 144 (QCIF格式)	384K	4	
	2	352 × 288 (CIF格式)	2M	16	
主要	1	352 × 288(CIF)	2M	16	
	2	720 × 576 (CCIR601标准)	15M	32	
	3	1920 × 1080 (高清电视)	38.4M	32	

简单、核心和主要视觉配置文件中的级别

MPEG4自然视觉对象类型与配置文件

1.5 MPEG - 4第10部分/H.264

- 2001年，MPEG和国际电信联盟电信标准化部门 (ITU - T) 视频编码专家组 (VCEG) 联合成立了联合视频组 (JVT)
- JVT于2003年向国际标准化组织 (ISO) 提交了H.264草案

- 与MPEG - 2相比，H.264的压缩率提高了50%，与 H.263+和MPEG - 4高级简单配置文件相比，压缩率提高了 30%

1.5 MPEG - 4第10部分/H.264

•核心特性

- 熵解码
 - 统一可变长编码 (UVLC) 和上下文自适应可变长编码 (CAVLC)
- 运动补偿或帧内预测
 - 可变块大小和更精确的运动补偿。
- 变换、扫描、量化
 - 非线性量化与不同量化尺度
- 帧内预测
 - 所有帧内编码宏块均使用相邻重建像素
- 环路去块滤波器
 - 采用了先进的信号自适应去块滤波器。

1.5 MPEG - 4 第10部分/H.264

- H.264中的去块滤波器可获得理想效果



未去块

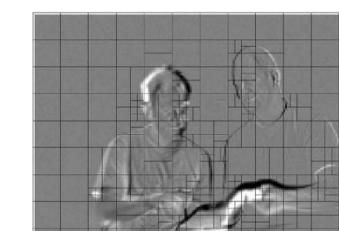
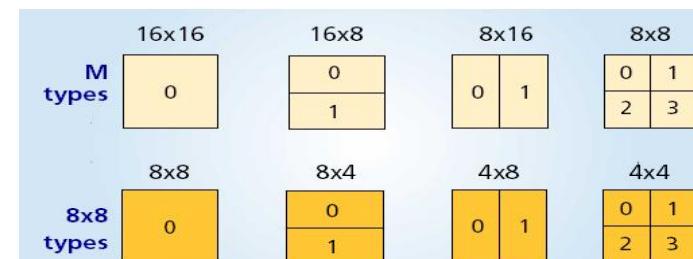


去块后

1.5 MPEG - 4 第10部分/H.264

•帧间预测

- 树状结构运动补偿
- H.264支持不同的块大小，块大小可低至4*4

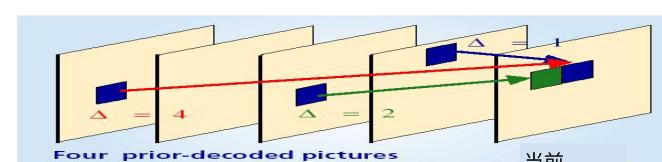


•选择最佳块大小，最小化

当前帧与参考帧之间的差异

•P帧可以使用多个前帧作为参考

帧。



MPEG - 7概述(1)

- 越来越多的多媒体内容成为各种应用程序不可或缺的一部分，有效且高效的检索成为首要关注的问题。

- MPEG7旨在满足基于视听内容的检索需求。

- MPEG7于1998年启动，2001年完成。
- MPEG7支持多种多媒体应用。
- MPEG7不描述任何特征提取方法。其正式名称是“多媒体内容描述接口”。

2. MPEG-7

- MPEG-7
 - 描述符 (D)、描述方案 (DS)
 - 描述方案 (DDL)
- 描述符 (D)
 - 颜色、纹理、形状、运动、定位
- 描述方案 (DS)
 - 基本元素、内容管理、内容
描述、导航与访问
- XML模式语言与MPEG7扩展

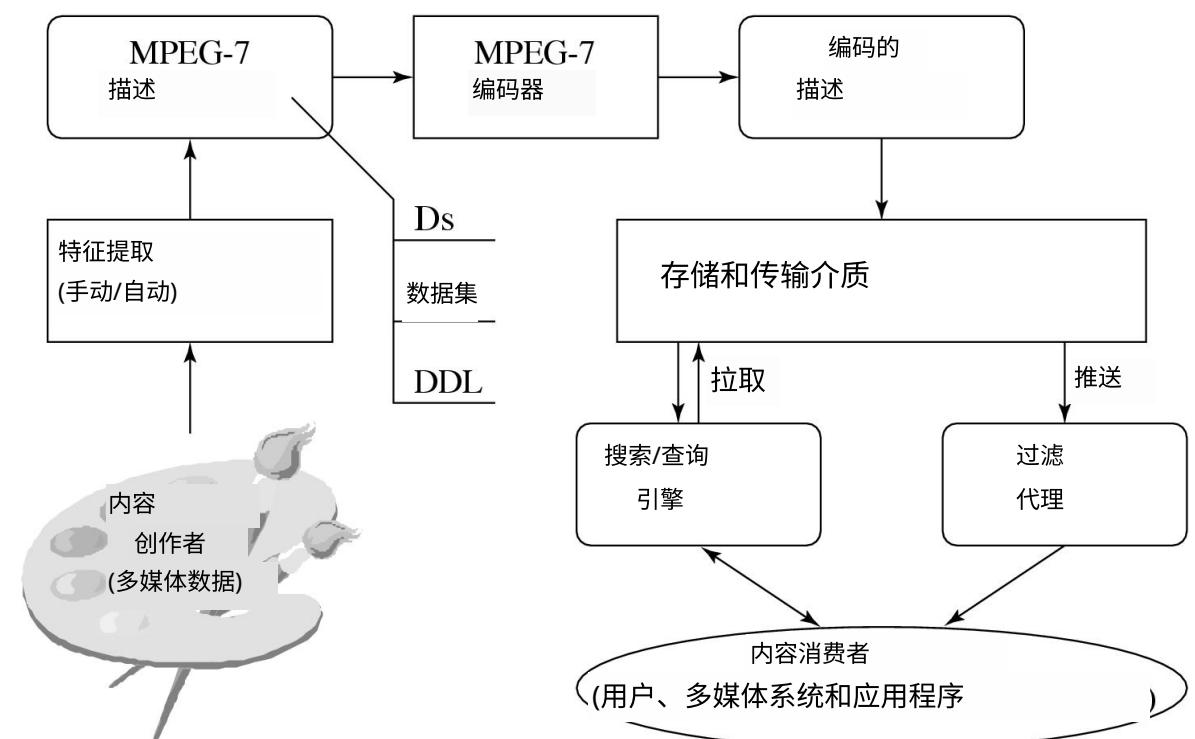


图12.17：使用MPEG - 7的可能应用。

3. MPEG - 21

MPEG - 21概述

- MPEG - 21旨在定义一种统一的方式来定义、识别、描述、管理和保护多媒体数据。
 - MPEG21有7个关键部分- 数字项声明- 数字项识别与描述- 内容管理与使用- 知识产权管理与保护- 终端与网络- 内容表示- 事件报告

结束

谢谢！

邮箱：junx@cs.zju.edu.cn