

[illegible]



[illegible]

【EZW】零树小波算法 (EZW) 全配, 即快速搜索算法。EZW 算法有两个中心部分组成: 零树数据结构和逐次逼近量化。

在给定码率下获得最佳的图像质量。

以嵌入式的方式完成这项工作。

逐次逼近量化 (SAQ) 生成嵌入码。

一种将所有低码率编码含在位流开头的部分, 每个位在位流中按重要性排列。

允许编码器或解码器在任意点结束编码或解码。

进一步利用零树特性。

最细致的搜索方法, 在零树数据上逐点搜索, 每搜索一点计算 MAD, 当 MAD 达到最小值时求得最佳匹配参数。

需要计算 MAD 的次数是  $(M+2d_{max}) \times (N+2d_{min})$ 。

当图像空间分辨率需要快速搜索时, 这时要采用快速的搜索算法是相称的。

快速搜索算法: 二维对数法、三步法、正交搜索法、共轭方向法等。

矢量

水平和平直方向的最大运动矢量范围为  $+16$ 。

自适应环路滤波器, 水平和垂直分别采用一维滤波, 其脉冲响应为  $[1/4, 1/2, 1/4]$ 。

运动矢量采用差分编码。

H.263 运动补偿

图像大小为  $16 \times 16$  像素宏块。

每个宏块与帧 (参考帧) 中周围的匹配。

搜索区域 = 水平/垂直  $15\text{-pixel}$  范围内。

半像素精度 (双线性像素内插)。

最佳匹配用于宏块预测。

(音调、响度、音色)

音高: 单位为美取, 于时, 强度的基频频率不变高, 强度的变化对音调略有影响。

强度增大时, 低频率音调显得更低, 而高频率音调显得更高。

响度

一定强度的声波作用于人听觉器官所引起的辨别声音强弱的感觉。

单位为宋 (sone)。

频率  $1\text{kHz}$ , 强度为听阈, 以上  $40\text{dB}$  的纯音所产生的响度为  $1\text{sone}$ 。

响度级

响度的相对量, 它表示的是其响度与基准响度的相对量。

部分信息, 它对声音音色的确定或音色和音高的区分, 与不相关, 但信号不能传输利用掩蔽效应来掩盖量化失真。

人耳听觉特性

人耳分辨能力

人耳对不同频段声音的敏感程度不同, 通常对低频比高频敏感。

人耳对语音信号的相位不敏感。

人耳掩蔽效应

Masking Effect

语音编解码器的分类

波形编解码器 (waveform coder): 不利用生成语音的信号的任何知识, 将话

语音分为清音、浊音和抖动语音三种状态。

浊音: 周期脉冲 + 白噪声。

抖动语音: 非周期脉冲 + 白噪声。

24kbs 的 MELP 编码方法已经被确定为美国新的联邦语音编码标准。

编码速率为  $2.4\text{kbs}$ , 帧长为  $2.5\text{ms}$ , 每帧数据用  $54\text{bit}$  进行编码。

在低码率下的合成语音的自然度大大提高, 主观语音质量接近美国联邦  $4.8\text{kbs}$  的 CELP。

的一个子集, 按视频帧的个数和帧数确定。按照相互包容的关系, 5 个类依次为:

简单类 (Simple Profile) 主类 (Main Profile) 信噪比可标度类 (SNR Profile) 空间分辨率可标度类 (Spatial Profile) 高级类 (High Profile) 级是对于每个类, 按其参数数目的变化, 按其所划分的子级。

由低到高分:

低等级 (Low Level) 主等级 (Main Level)  $1440$  等级 (High-1440 Level)

<p>进行编码对更多的重要图</p> <p><b>预测编码</b></p> <p><b>DPCCM 差分脉冲编码调制</b></p> <p>DPCCM 的一个重要方面是预测值于输出不是基于帧的未量化量，而是基于帧的未量化量，所以在此误差是在帧后处理的失真重建误差。</p> <p>在帧的 DC 漂移和误差积累。</p>	<p>■ 它们共同之处在于把判决准则 (例如 MAD) 趋于最小的方向作为最小失真方向。</p> <p>假定判决准则函数在最小失真方向上为单峰函数，则在其 (1) 平滑区域，(2) 快速搜索从任一猜测点开始沿最小失真方向进行。</p> <p>■ 这些快速搜索算法在实际上都是不同的，在帧的 DC 漂移和误差积累和步长不同，因此</p> <p>测</p> <p>■ 相关的位移或运动矢量被编码，并且直接解码。</p> <p>■ 所有帧的预测差值构成差值图像。</p> <p><b>H.263 DCT 变换</b></p> <p>■ 差值图像分成 8×8 象素块。</p> <p>■ 每个 8×8 块分别采用 2D DCT 变换。</p> <p>■ DCT 系数描述了块的空间频率。</p> <p>■ 高频对应于图像细节，例如轮廓和纹理。</p>	<p>度比值的对数值</p> <p>单位为方 (phon)</p> <p>当人耳感到某声音与 1kHz 的语音同样响时，1kHz 语音的声压级即为其响度级。</p> <p>■ 音色：</p> <p>■ 泛音与基音的比例，泛音的多少和出现的时间的长短变化决定乐音的不同发音源 (乐器) 的材质、形状不同，其泛音的排列组合也不同，也就构成了这一物体特殊的音色。</p> <p>■ 混合编解码器 (Hybrid coder)</p>	<p>音视为一种普遍的声音，直接对波形信号进行采样和量化。例如 PCM、DPCM 和 ADPCM 等。</p> <p>■ 音频编解码器 (Source coder)</p> <p>■ 也称为参量编解码器 (parametric coder)，它从语音波形信号中提取出语音生成模型的参数生成模型参数通过语音重构出语音。</p> <p>■ 混合编解码器 (Hybrid coder)</p>	<p>MPEG1/Audio 的通用编码概念</p> <p>■ 层的概念：根据不同的复杂性和性能的系统编解码器分为三层：Layer I、II、III。</p> <p>一个标准的 ISO 解码器应能对任何一层编解码器流解码。</p> <p>一个标准第 x 层编解码器能对 x-n 层编解码器流解码。</p> <p>■ Layer I</p>	<p>高等级 (High Level) 由内容上相互关联的文本、图形、图像、音频、视频等媒体数据构成的复合信息实体。</p> <p>■ 多媒体数据包含的多种媒体对象不是相互独立的</p> <p>媒体对象之间存在着多种相互约束的关系。</p> <p>内容约束 Content relations</p> <p>空间约束 Spatial relations</p> <p>时间约束 Temporal</p>
---	---	---	---	---	---

<p>由图 4.10</p> <p><b>DCPM 的量化噪声</b></p> <p>量化噪声 (Granular Noise)</p> <p>图像平部分处处的随机噪声</p> <p>边缘混乱 (Edge Busyness):</p> <p>视频边界处的抖动现象。量化步长大于信号精度，大小变化，故不能准确反映这种小的幅度变化；</p> <p>斜度过载 (Slope Overload):</p> <p>边界产生模糊，周期性的精确的级数图，幅度变化的斜度步长小，故不能准确反映这种大斜率的幅度变化。或者当输入信号幅度为 <math>-(\infty, +\infty)</math> 或</p>	<p><b>光流</b> (Optical Flow) 是运动物体在观察帧空平面上的像素运动的瞬时速度。</p> <p>通常将二维图像平面时变坐标点上的灰度函数变化率定义为光流矢量。</p> <p>平面上的每一点在时刻 <math>t</math> 与 <math>t+\Delta t</math> 的位移矢量，率就是该点的光流矢量。平面上各点的光流向量组成了该平面的光流场。</p> <p>光流法是利用图像上序列帧中像素在时间轴上的变化以及相邻帧之间像素的相似性来找到它们之间的运动关系，从而计算出相邻帧之间物体的运动信息的一种方法。</p>	<p>低频对应于大的特征，例如大面积图像。</p> <p>最低频率系数称为直流 DC 系数，对应于块的平均亮度。</p> <p><b>H.263 量化和编码码</b></p> <p>对变换系数进行量化；</p> <p>小的系数被丢弃；</p> <p>剩余系数编码；</p> <p>帧间编码，所有的系数获得相同的比特数，除了 DC 更多外；</p> <p>帧内编码，低频率系数获得更多比特，以保持良好的大的特征；</p> <p>实际比特数取决于量化参数 <math>QP</math>，它的值取决于比特分配策略。</p> <p>最后，采用传统的 Huffman 码对量化后的</p>	<p><b>掩蔽效应</b></p> <p>一个声音的存在掩蔽了另一个声音，掩蔽一个信号的掩蔽和生成原理与两个声音的相关性有密切关系。</p> <p>分为频域效应和时域掩蔽效应。在频域掩蔽效应，稳态短时掩蔽信号特性，利用频域掩蔽效应。</p> <p>对于声音的时域结构特性，利用时域掩蔽效应。</p> <p><b>时域掩蔽效应</b></p> <p>在时间上相邻的声音之间存在的掩蔽现象</p> <p>后向掩蔽：一个声音影响了在时间上先于它的声音的听觉能力</p> <p>同时掩蔽：在一定时间同时产生两个声音，一个声音对另一个声音的掩蔽作用</p>	<p>综合使用上述两种技术，使用的激励信号波形尽可能接近于原始话音信号的波形，例如 CELP。</p> <p><b>LPC-10 声码器</b></p> <p>M=10: LPC-10</p> <p>美国国家标准</p> <p>输入语音：8000 样本/秒</p> <p>180 样本 / 帧 (22.5ms)</p> <p>语音信号在 10 ~ 30ms 内具有短时平稳的特点</p> <p>输出带宽：2.4 kbps，54 比特/帧 44.44 帧/秒 (22.5ms)</p> <p>LPC-10 问题和改进存在的问题</p>	<p>它是 MUSICAM 编码方案的高化版本</p> <p>编码技术</p> <p>数字音频输入变换为 32 个固定子带</p> <p>心理模型 I</p> <p>自适应比特分配；均匀量化</p> <p>每帧包含 384 个样值</p> <p>应用：那些不要求很低码率的场合，但时是要求高的编码效率。</p> <p>适合民用，例如家庭数字磁带录音</p> <p>温彻斯特盘、磁光盘等</p> <p>Layer II</p> <p>它和 MUSICAM 系统</p>	<p><b>多媒体同步机制</b></p> <p>■ 流内同步</p> <p>基于播放时限 (dead line) 的方法</p> <p>基于缓存器数据量的控制方法</p> <p>■ 流间同步</p> <p>基于全局时钟的时间戳方法</p> <p>基于反馈的流间同步方法</p> <p>■ 时钟同步</p> <p>基于接收缓存器的方法</p> <p>基于时间戳的锁相方法</p> <p>基于网络时间协议的方法</p>
---	---	--	--	--	---	--

<p>在大时，在两帧会产生过裁。</p> <p>视频信号的<b>帧间编码</b></p> <p>帧间编码利用时间连续图像的相关性和人眼的时域特性。</p> <p>重要的帧间编码方法： ■ 自适应帧内-帧间编码</p>	<p><b>基本假设条件：</b></p> <ol style="list-style-type: none"><li>亮度恒定不变：即同一像素在不同帧中产生的误差是基本无流变的假定。</li><li>时间连续或运动量小：即当时间的自适应性不会引起目标位置的周期性变化。相邻帧之间帧差较小。</li></ol> <p>H264 I和P slice (not B slice)</p> <p>■ 帧编码CAVLC（基于上下文的自适应变长码） long Context based adaptive VLC</p> <p>■ 去块效应环路滤波</p>	<p>音谱掩蔽效应：已经结音的听觉能力还起着影响作用。</p> <p>利用时域掩蔽效应：将时间上相邻的一些取样值合并成块，降低码率。</p> <p>三种时域掩蔽效应的时间常数不同</p> <p>频率掩蔽</p>	<p>语音的自然度较差：采用过于简单的二元模型组成的语音信号在自然的语言冗余信息中，相当一部分是非周期脉冲冲并非线性成分。</p> <p>非线性的周期成分没有得观察到，造成听音有明显的失真感。</p> <p>语言的简单程度：与相同(除了例外外)，提供了比例因子。</p> <p>特分配和取样的附加编码</p> <p>在民用和音频专业领域都有许多应用：数字声广播、DAB和电视；通信和多媒体。</p> <p>Layer III</p>	<p>DPCM 编/解码原理图</p>
--	---	---	--	---------------------

**帧内预测**

- 有条件的补充;
  - 运动补偿预测和内插。
- 框图见后。
- 适应帧内-帧间 DPCM 原理

预测器在两种状态之间切换:

- 活动或变化区域采用帧内预测
- 静止区域采用帧间预测

**帧内预测**

- 无须了解场景信息，即可对当前帧进行预测。
- 就运动补偿而言，帧内预测仍然适用。
- 光流不仅携带了运动信息，也携带了物体本身的信息。因此，在运动检测的情况下，检测出运动的物体，并知道其运动方向，是有帮助的。

**帧内预测**

- 支持 ASO (Arbitrary slice ordering, ASO)
- 支持 FMO (Flexible macroblock ordering, FMO)
- 支持冗余片

**帧间预测**

- 同时发出的两个声道信号一个相近频率的较强音另一个相近频率的较弱音
- 声音所掩蔽的两个声道最明显
- 在高频段，随着两者频率增高，掩蔽效果缓慢减小
- 在低频段，随着两者频率增高，掩蔽效果急剧减小

**帧间预测**

- 如果语音用噪声值激励，听起来会觉得沙哑。
- 鲁棒性差：在低信噪比的情况下不能取得与期望值和不能正确区分清音和静音。
- 非均匀量化器；自适应分段的心理声学模型 II。

**帧间预测**

- 是 ASPEC 中最有效部分和 MUSICAM 系统结合。
- 采用的技术多相滤波组子带合成和 MDCT (带延迟) 结合的混合滤波器。
- 非均匀量化器；自适应分段的心理声学模型 II。

**帧间预测**

$$e = s - \hat{s}, \quad d = e + \hat{d}, \quad s' - s = e' - e = g$$

假设是线性预测，DPCM 编码器等效于下列结构：

预测器的传输函数： $\hat{E}(z) = [1 - P(z)]X(z)$

码元量化的代数表示：

**运动补偿估计**

- 对于静止或运动缓慢的图象时，帧间差值信号比帧内相邻像素间的差值信号的直方图有更为尖锐的、以零为中心的分布。
- 缺点：即两个基本假设在现实情况下均不容易满足。
  - 光流法的适用条件：适用于 IP 视频、视频会议应用。

**H.265 新技术**

- 四叉树编码结构
- 多角度帧内预测
- 广义 B 帧预测技术
- 自适应运动矢量预测

**线性预测编码**

语音信号具有自相关性，引入了“临界频率”的概念。

一个声音可被另一个声音掩蔽的频率范围称为“临界频带”。

**MUSICAM 心理声学模型**

主要包含以下方面：

- 激励源基音提取方法
- 声道滤波器参数量化方法
- 线谱对（Line Spectrum Pair, LPS）
- 多脉冲线性预测编码（MP-LPC）
- 掩蔽扩散效应

每帧 1152 样值，广泛适用于通信领域，尤其是窄带的 ISDN、Internet 和要求很低的码率的专业领域。

公式： $E(n) = Q(n) \cdot E(n-1) + (1 - Q(n)) \cdot S(n)$   
 $Q(n) = [E(n-1) - Q(n-1)V(n)] / E(n) + Q(n-1)V(n)/E(n)$   
 $V(n) = (1 - P(n))G(n)$

<p>性较差。</p> <p>■ 如果能够对运动物体进行位置差值, 再求差值, 则相关性增加, 从而提高了压缩比。</p>	<p>■ Laplace 分布, 表现出更强的相关性。</p> <p>■ 对于运动剧烈的物体, 其间差值信号的幅度剧烈变化。</p> <p>■ 如果能够对运动物体进行位置差值, 再求差值, 则相关性增加, 从而提高了压缩比。</p>	<p>AV1/H.266 新算法</p> <p>■ 块结构</p> <p>元划分和更多的划分模式</p> <p>■ 帧内预测模式</p> <p>50-65 帧内预测方向</p> <p>亮度分量预测度分量</p> <p>■ 帧内预测模式</p>	<p>频带内的频率</p> <p>■ 听觉特性: 声音的方位</p> <p>■ 人对信噪比和频率的方向辨别能力有所降低, 大多数声音是多频道的, 人的听觉心理会人根据声音中的中频成分的方向</p> <p>(判断方向灵敏度较高的频段), 来判断那些会</p>	<p>■ 多脉冲线性预测编码 (multiple linear predictive coding, MP-LPC):</p> <p>对清音/语音, 有限个脉冲经过最优估值后作为激励信号, 不同脉冲的间隔根据</p> <p>■ 分析过程: 混合编码</p>	<p>■ 频率单音对临界频率</p> <p>带内频率相近临界频带的掩蔽效应</p> <p>■ 频率单音对临界频率</p> <p>带外声音的掩蔽效应</p> <p>信噪比 (SNR)=信号峰值-噪声有效值</p> <p>信号掩蔽比 (SMR)=信号峰值-最小掩蔽阈值</p>	<p>线性预测声码器 (LPC)</p> <p>■ 激励: 二元激励源</p> <p>■ 激励: 基音周期脉冲</p> <p>■ 激励: 随机噪声</p> <p>■ 激励: 线性滤波器 <math>y_i = \sum_{j=1}^N a_{ij}x_{ij} + G_i</math></p>
---	--	---	--	--	--	---

ISO/IEC 11173-1 标准定义了两种基本压缩算法：

- 基于空间预测编码 DPCM 的无失真压缩算法。
- 基于运动估计和运动补偿的有失真压缩算法。

前者用于静态图像，后者用于动态图像。前者使用两种类型的参数来表示：当前要编码的图像宏块和参考图像的宏块之间的差值、宏块的运动矢量。

后者使用两种类型的参数来表示：当前要编码的图像宏块和参考图像的宏块之间的差值、宏块的运动矢量。

后者使用两种类型的参数来表示：当前要编码的图像宏块和参考图像的宏块之间的差值、宏块的运动矢量。

**运动矢量**

编码图像中的当前宏块 MPI 相对于参考图像中的宏块 MRJ 所移动的距离和方向, 这就是运动矢量 (Motion Vector)。

宏块 MRJ 是宏块 MPI 的最佳匹配块, 所谓最佳匹配是指这两个宏块之间的差值按某种判决准则最小。

运动估值 (Motion Estimation) 寻找最佳匹配宏块的过程。

基于 DCT 变换有失真的压缩算法, DCT 第一次成功的应用。

**运动补偿**

x-x 差值进行无失真的运动补偿或运动矢量压缩比可达 2:1。

**量化**

为了达到压缩的目的, 对 DCT 系数  $F(u,v)$  进行量化处理, MPEG 采用量化均匀化处理, 量化用的映射产生误差。

**应用**

应用于 Internet, 彩色

局部光照补偿(LIC) 仿射变换模拟物体在实际运动中的卷曲和形变

模式匹配运动矢量推导

- 变换
- 自适应多核变换支持 DCT/ADST/ADST/IDT 及类型组合
- 熵编码
- CABAC 调整
- 非二进制算术编码

将合成语音与原语音混合, 得到合成语音, 得到加权误差信号。

根据最小加权均方误差, 分析出一组位置、幅度最佳的激励脉冲, 然后最佳激励脉冲与 LP 系数一起编码送入信道。

**码激励线性预测编码**

- CELP: 矢量量化 + 线性预测
- 码书中每个码字代表一个基信号, 有可能

帧内编码帧, 压缩没有参考其它图像, 只利用空间冗余度。

P 预测编码帧, 利用运动补偿进行帧内预测, 帧的前向时间冗余度。

B - 双向预测 (Bidirectionally-predicted) 编码帧, 利用了 P 帧的前向和后向时间冗余度。

MPEG-2 和 MPEG1 的帧结构

- 保持尽可能大的兼容性

Layer 1, Layer 1 编码器框图