原 视频压缩编码和音频压缩编码的基本原理

2014年06月03日 00:01:20　阅读数：52307

本文介绍一下视频压缩编码和音频压缩编码的基本原理。其实有关视频和音频编码的原理的资料非常的多，但是自己一直也没有去归纳和总结一下，在这里简单总结一下，以作备忘。

1.视频编码基本原理

（1） 视频信号的冗余信息

以记录数字视频的YUV分量格式为例，YUV分别代表亮度与两个色差信号。例如对于现有的PAL制电视系统，其亮度信号采样频率为13.5MHz；色度信号的频带通常为亮度信号的一半或更少，为6.75MHz或3.375MHz。以4：2：2的采样频率为例，Y信号采用13.5MHz，色度信号U和V采用6.75MHz采样，采样信号以8bit量化，则可以计算出数字视频的码率为：

13.5*8 + 6.75*8 + 6.75*8= 216Mbit/s

如此大的数据量如果直接进行存储或传输将会遇到很大困难，因此必须采用压缩技术以减少码率。数字化后的视频信号能进行压缩主要依据两个基本条件：

l 数据冗余。例如如空间冗余、时间冗余、结构冗余、信息熵冗余等，即图像的各像素之间存在着很强的相关性。消除这些冗余并不会导致信息损失，属于无损压缩。

l 视觉冗余。人眼的一些特性比如亮度辨别阈值，视觉阈值，对亮度和色度的敏感度不同，使得在编码的时候引入适量的误差，也不会被察觉出来。可以利用人眼的视觉特性，以一定的客观失真换取数据压缩。这种压缩属于有损压缩。

数字视频信号的压缩正是基于上述两种条件，使得视频数据量得以极大的压缩，有利于传输和存储。一般的数字视频压缩编码方法都是混合编码，即将变换编码，运动估计和运动补偿，以及熵编码三种方式相结合来进行压缩编码。通常使用变换编码来消去除图像的帧内冗余，用运动估计和运动补偿来去除图像的帧间冗余，用熵编码来进一步提高压缩的效率。下文简单介绍这三种压缩编码方法。

（2） 压缩编码的方法

（a） 变换编码

变换编码的作用是将空间域描述的图像信号变换到频率域，然后对变换后的系数进行编码处理。一般来说，图像在空间上具有较强的相关性，变换到频率域可以实现去相关和能量集中。常用的正交变换有离散傅里叶变换，离散余弦变换等等。数字视频压缩过程中应用广泛的是离散余弦变换。

离散余弦变换简称为DCT变换。它可以将L*L的图像块从空间域变换为频率域。所以，在基于DCT的图像压缩编码过程中，首先需要将图像分成互不重叠的图像块。假设一帧图像的大小为1280*720，首先将其以网格状的形式分成160*90个尺寸为8*8的彼此没有重叠的图像块，接下来才能对每个图像块进行DCT变换。

经过分块以后，每个8*8点的图像块被送入DCT编码器，将8*8的图像块从空间域变换为频率域。下图给出一个实际8*8的图像块例子，图中的数字代表了每个像素的亮度值。从图上可以看出，在这个图像块中各个像素亮度值比较均匀，特别是相邻像素亮度值变化不是很大，说明图像信号具有很强的相关性。

一个实际8*8图像块

下图是上图中图像块经过DCT变换后的结果。从图中可以看出经过DCT变换后，左上角的低频系数集中了大量能量，而右下角的高频系数上的能量很小。

图像块经过DCT变换后的系数

信号经过DCT变换后需要进行量化。由于人的眼睛对图像的低频特性比如物体的总体亮度之类的信息很敏感，而对图像中的高频细节信息不敏感，因此在传送过程中可以少传或不传送高频信息，只传送低频部分。量化过程通过对低频区的系数进行细量化，高频区的系数进行粗量化，去除了人眼不敏感的高频信息，从而降低信息传送量。因此，量化是一个有损压缩的过程，而且是视频压缩编码中质量损伤的主要原因。

量化的过程可以用下面的公式表示：

其中 F_Q （u,v）表示经过量化后的DCT系数；F（u,v）表示量化前的DCT系数；Q（u,v）表示量化加权矩阵；q表示量化步长；round表示归整，即将输出的值取为与之最接近的整数值。

合理选择量化系数，对变换后的图像块进行量化后的结果如图所示。

量化后的DCT系数

DCT系数经过量化之后大部分经变为0，而只有很少一部分系数为非零值，此时只需将这些非0值进行压缩编码即可。

（b） 熵编码

熵编码是因编码后的平均码长接近信源熵值而得名。熵编码多用可变字长编码（VLC，Variable Length Coding）实现。其基本原理是对信源中出现概率大的符号赋予短码，对于出现概率小的符号赋予长码，从而在统计上获得较短的平均码长。可变字长编码通常有霍夫曼编码、算术编码、游程编码等。其中游程编码是一种十分简单的压缩方法，它的压缩效率不高，但编码、解码速度快，仍被得到广泛的应用，特别在变换编码之后使用游程编码，有很好的效果。

首先要在量化器输出直流系数后对紧跟其后的交流系数进行Z型扫描（如图箭头线所示）。Z型扫描将二维的量化系数转换为一维的序列，并在此基础上进行游程编码。最后再对游程编码后的数据进行另一种变长编码，例如霍夫曼编码。通过这种变长编码，进一步提高编码的效率。

（c） 运动估计和运动补偿

运动估计（Motion Estimation）和运动补偿（Motion Compensation）是消除图像序列时间方向相关性的有效手段。上文介绍的DCT变换、量化、熵编码的方法是在一帧图像的基础上进行，通过这些方法可以消除图像内部各像素间在空间上的相关性。实际上图像信号除了空间上的相关性之外，还有时间上的相关性。例如对于像新闻联播这种背景静止，画面主体运动较小的数字视频，每一幅画面之间的区别很小，画面之间的相关性很大。对于这种情况我们没有必要对每一帧图像单独进行编码，而是可以只对相邻视频帧中变化的部分进行编码，从而进一步减小数据量，这方面的工作是由运动估计和运动补偿来实现的。

运动估计技术一般将当前的输入图像分割成若干彼此不相重叠的小图像子块，例如一帧图像的大小为1280*720，首先将其以网格状的形式分成40*45个尺寸为16*16的彼此没有重叠的图像块，然后在前一图像或者后一个图像某个搜索窗口的范围内为每一个图像块寻找一个与之最为相似的图像块。这个搜寻的过程叫做运动估计。通过计算最相似的图像块与该图像块之间的位置信息，可以得到一个运动矢量。这样在编码过程中就可以将当前图像中的块与参考图像运动矢量所指向的最相似的图像块相减，得到一个残差图像块，由于残差图像块中的每个像素值很小，所以在压缩编码中可以获得更高的压缩比。这个相减过程叫运动补偿。

由于编码过程中需要使用参考图像来进行运动估计和运动补偿，因此参考图像的选择显得很重要。一般情况下编码器的将输入的每一帧图像根据其参考图像的不同分成3种不同的类型：I（Intra）帧、B（Bidirection prediction）帧、P（Prediction）帧。如图所示。

典型的I，B，P帧结构顺序

如图所示，I帧只使用本帧内的数据进行编码，在编码过程中它不需要进行运动估计和运动补偿。显然，由于I帧没有消除时间方向的相关性，所以压缩比相对不高。P帧在编码过程中使用一个前面的I帧或P帧作为参考图像进行运动补偿，实际上是对当前图像与参考图像的差值进行编码。B帧的编码方式与P帧相似，惟一不同的地方是在编码过程中它要使用一个前面的I帧或P帧和一个后面的I帧或P帧进行预测。由此可见，每一个P帧的编码需要利用一帧图像作为参考图像，而B帧则需要两帧图像作为参考。相比之下，B帧比P帧拥有更高的压缩比。

（d） 混合编码

上面介绍了视频压缩编码过程中的几个重要的方法。在实际应用中这几个方法不是分离的，通常将它们结合起来使用以达到最好的压缩效果。下图给出了混合编码（即变换编码+ 运动估计和运动补偿+ 熵编码）的模型。该模型普遍应用于MPEG1，MPEG2，H.264等标准中。

混合编码模型

从图中我们可以看到，当前输入的图像首先要经过分块，分块得到的图像块要与经过运动补偿的预测图像相减得到差值图像X，然后对该差值图像块进行DCT变换和量化，量化输出的数据有两个不同的去处：一个是送给熵编码器进行编码，编码后的码流输出到一个缓存器中保存，等待传送出去。另一个应用是进行反量化和反变化后的到信号X'，该信号将与运动补偿输出的图像块相加得到新的预测图像信号，并将新的预测图像块送至帧存储器。

2.音频编码基本原理

（1） 音频信号的冗余信息

数字音频信号如果不加压缩地直接进行传送，将会占用极大的带宽。例如，一套双声道数字音频若取样频率为44.1KHz，每样值按16bit量化，则其码率为：

2*44.1kHz*16bit=1.411Mbit/s

如此大的带宽将给信号的传输和处理都带来许多困难，因此必须采取音频压缩技术对音频数据进行处理，才能有效地传输音频数据。

数字音频压缩编码在保证信号在听觉方面不产生失真的前提下，对音频数据信号进行尽可能大的压缩。数字音频压缩编码采取去除声音信号中冗余成分的方法来实现。所谓冗余成分指的是音频中不能被人耳感知到的信号，它们对确定声音的音色，音调等信息没有任何的帮助。

冗余信号包含人耳听觉范围外的音频信号以及被掩蔽掉的音频信号等。例如，人耳所能察觉的声音信号的频率范围为20Hz～20KHz，除此之外的

其它频率人耳无法察觉，都可视为冗余信号。此外，根据人耳听觉的生理和心理声学现象，当一个强音信号与一个弱音信号同时存在时，弱音信号将被强音信号所掩蔽而听不见，这样弱音信号就可以视为冗余信号而不用传送。这就是人耳听觉的掩蔽效应，主要表现在频谱掩蔽效应和时域掩蔽效应，现分别介绍如下：

（a） 频谱掩蔽效应

一个频率的声音能量小于某个阈值之后，人耳就会听不到，这个阈值称为最小可闻阈。当有另外能量较大的声音出现的时候，该声音频率附近的阈值会提高很多，即所谓的掩蔽效应。如图所示：

频率掩蔽效应

由图中我们可以看出人耳对2KHz～5KHz的声音最敏感，而对频率太低或太高的声音信号都很迟钝，当有一个频率为0.2KHz、强度为60dB的声音出现时，其附近的阈值提高了很多。由图中我们可以看出在0.1KHz以下、1KHz以上的部分,由于离0.2KHz强信号较远，不受0.2KHz强信号影响,阈值不受影响；而在0.1KHz～1KHz范围，由于0.2KHz强音的出现,阈值有较大的提升，人耳在此范围所能感觉到的最小声音强度大幅提升。如果0.1KHz～1KHz范围内的声音信号的强度在被提升的阈值曲线之下，由于它被0.2KHz强音信号所掩蔽，那么此时我们人耳只能听到0.2KHz的强音信号而根本听不见其它弱信号，这些与0.2KHz强音信号同时存在的弱音信号就可视为冗余信号而不必传送。

（b） 时域掩蔽效应

当强音信号和弱音信号同时出现时，还存在时域掩蔽效应。即两者发生时间很接近的时候，也会发生掩蔽效应。时域掩蔽过程曲线如图所示，分为前掩蔽、同时掩蔽和后掩蔽三部分。

时域掩蔽效应

由图我们可以看出，时域掩蔽效应可以分成三种：前掩蔽，同时掩蔽，后掩蔽。前掩蔽是指人耳在听到强信号之前的短暂时间内，已经存在的弱信号会被掩蔽而听不到。同时掩蔽是指当强信号与弱信号同时存在时，弱信号会被强信号所掩蔽而听不到。后掩蔽是指当强信号消失后，需经过较长的一段时间才能重新听见弱信号，称为后掩蔽。这些被掩蔽的弱信号即可视为冗余信号。

（2） 压缩编码方法

当前数字音频编码领域存在着不同的编码方案和实现方式, 但基本的编码思路大同小异, 如图所示。

数字音频编码系统模型

对每一个音频声道中的音频采样信号,首先都要将它们映射到频域中,这种时域到频域的映射可通过子带滤波器实现。每个声道中的音频采样块首先要根据心理声学模型来计算掩蔽门限值， 然后由计算出的掩蔽门限值决定从公共比特池中分配给该声道的不同频率域中多少比特数，接着进行量化以及编码工作，最后将控制参数及辅助数据加入数据之中，产生编码后的数据流。

文章标签： 视频编码 音频编码 原理 冗余信息 编码方法

个人分类： 音频编码 视频编码