

对一个宏块的编码。

P帧的用途在图7-10所示的例子中可以看出。在图7-10中我们看到三个连续的帧具有相同的背景，但是在一个人所在的位置上存在差异。对于摄像机固定在三脚架上，而演员在摄像机面前活动的情形中，这种场景是常见的。包含背景的宏块是严格匹配的，但是包含人的宏块在位置上存在某一未知数量的偏移，编码时必须追踪到前一帧中相应的宏块。



图7-10 三个连续的视频帧

MPEG标准没有规定如何搜索、搜索多远以及如何计算一个匹配的好坏，这些都留给每一具体的实现。例如，一种实现可能在前一帧中的当前位置以及所有在 $x$ 方向偏移 $\pm \Delta x$ 、在 $y$ 方向偏移 $\pm \Delta y$ 的位置搜索一个宏块。对于每个位置，可以计算出亮度矩阵中匹配的数目。具有最高得分的位置将成为获胜者，只要其得分高于某一预设的阈值。否则，宏块就被称为失配。当然，更复杂的算法也是可能的。

如果一个宏块被找到，则通过以其值与前一帧中的值求差对其进行编码（针对亮度和两个色度），然后，对这些差值矩阵进行JPEG编码。输出流中宏块的值是运动矢量（宏块在每一方向从其前一位置移动多远的距离），随后是以JPEG进行编码的与前一帧的差值。如果宏块在前一帧中查找不到，则当前值以JPEG进行编码，如同在I帧中一样。

B帧与P帧相类似，不同的是它允许参考宏块既可以在前一帧中，也可以在后续的帧中，既可以在I帧中，也可以在P帧中。这一额外的自由可以改进运动补偿，并且在物体从前面（或后面）经过其他物体时非常有用。例如，在一场垒球比赛中，当三垒手将球掷向一垒时，可能存在某些帧其中垒球遮蔽了在背景中移动的二垒手的头部。在下一帧中，二垒手的头部可能在垒球的左面有一部分可见，头部的下一个近似可以从垒球已经通过了头部的后续的帧中导出。B帧允许一个帧基于未来的帧。

要进行B帧编码，编码器需要在内存中同时保存三个解码的帧：过去的一帧、当前的一帧和未来的一帧。为了简化解码，各帧必须以依赖的顺序而不是以显示的顺序出现在MPEG流中。因而，当一段视频通过网络被观看时，即使有完美的定时，在用户的机器上也需要进行缓冲，对帧进行记录以便正常地显示。由于这一依赖顺序和显示顺序间的差异，试图反向播放一部电影而没有相当可观的缓冲和复杂的算法是无法工作的。

有许多动作以及快速剪切（比如战争电影）的电影需要许多I型帧。而那种在导演对准了摄像机之后便出去喝咖啡，只留下演员背台词（比如爱情故事）的电影，就可以使用长段的P帧与B帧，而这两种帧结构与I帧相比使用很少的存储空间。从磁盘效率的角度来看，一个运营多媒体服务的公司应该尝试得到尽可能多的女性消费群体。

## 7.4 音频压缩

就像我们刚刚看到的，CD品质的音频需要一个1.411 Mbps带宽的传送。很清楚，在Internet的实际传送中，需要有效的压缩。正是因为这一点，已经发展起来许多不同的音频压缩算法。或许最流行的算法是拥有三个层（变体）的MPEG音频，其中，MP3(MPEG 音频层3)是功能最强大也是最出名的。在Internet上随处可见大量MP3格式的音乐，它们并非都合法，因此引发了许多来自艺术家与版权拥有者的案件。MP3属于MPEG视频压缩标准里的音频部分。

音频压缩可以用两种方法完成。在波形编码技术中，信号通过傅立叶变换（Fourier transform）转换成频率分量。图7-11给出一个时间与它最初的15个傅立叶振幅的实例函数。然后每一个分量的振幅用最简短的方法编码。目标是在另一端用尽可能少的二进制位精确地重建其波形。

另一种方法是感知编码，这种技术是在人类听觉系统中寻找某种细纹，用来对信号编码，这种信号听起来与人的正常收听相同，尽管在示波器上看起来却大相径庭。感知编码是基于心理声学的——人们如何感知声音的科学。MP3正是基于感知编码。