CMPSCI 754 Multimedia Systems | Spring 2001

Lecture 3: 8th February

Lecturer: Prashant Shenoy | *Scribe: Akash Jain*

3.1 MPEG

MPEG is the international standard for moving picture video compression. The name MPEG is an acronym for Moving Picture Experts Group. MPEG exploits temporal redundancy over successive frames in a video sequence to achieve further compression over JPEG.

Different MPEG standards:

- *MPEG-1:* Intended for intermediate data rates on the order of 1.5 Mbit/s.
- *MPEG-2:* Intended for higher data rates on the order of 10 Mbit/s or more.
- *MPEG-4:* Intended for very low bit rates of 64 KBit/s or less and compatible with MPEG-1 and MPEG-2. It can also efficiently code or represent multiviewpoint scenes, graphics and synthetic models, and allow functions such as interactivity and manipulation of scene contents.
- *MPEG-7:* Describes streaming media content to enable efficient search on audio and video.

MPEG was designed to cater to the following requirements:

- *Random Access:* The user should be able to access the media at any point of time.
- *Fast searches:* Forward and reverse search should be quick.
- *Reverse playback of the stream*
- *Audio-video synchronization*
- *Robustness:* Specially needed when transmitting over a network. Even if some packets are lost, decoder should be able to decode other packets.
- *Low encoding/decoding delay:* To enable clients to view live events.
- *Editability:* For example, to insert an advertisement.

3.1.1 MPEG Fundamentals

Encoding categories:

- *Intra-frame encoding:* Encoding is independent to other frames.
- *Inter-frame encoding:* Encoding is dependent on other frames.

To achieve a high compression ratio, temporal redundancies of subsequent pictures must be exploited (inter-frame), whereas the demand for random access requires intraframe coding.

MPEG Algorithm:

- Spatial redundancy is reduced using DCT-based compression.
- Temporal redundancy is exploited using block-based motion compensation where both causal (predictive coding) and non-causal (interpolative coding) predictors are used.

3.1.2 MPEG frame types

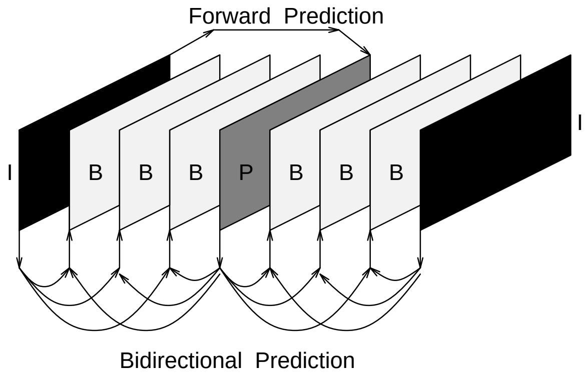

Figure 3.1: Typical ordering of frames

- *I frames:* Intra-coded images, coded without any reference to other images. The compression rate of I frames is lowest within MPEG. I frames are points for random access in MPEG streams.
- *P frames:* Predictively coded frames. Require information of the previous I and/or P frames for encoding and decoding. Can achieve higher compression ratio than I frames.
- *B frames:* Bidirectionally predictively coded frames. Require information from the previous and the following I and/or P frames for encoding and decoding. Acheives highest compression ratio. B frames can never be directly accessed in a random fashion.
- *D frames:* DC coded frames. These frames are intra-frame encoded. They can be used for fast forward or rewind mode. The DC parameters are DCT coded and the AC coefficients are neglected. It is a very coarse grain representation.

A *Group of Pictures (GoP)* in MPEG is a self contained group of images. They serve as boundary for random access.

3.1.3 Motion Compensation in MPEG

As discussed above, MPEG exploits temporal redundancy in video frames. This can be achieved in the following ways:

- *Prediction:* Current information is modeled as a translation of a previous frame. One also needs to keep in mind that displacement need not be the same everywhere in the frame so encoding of motion information must be at a finer granularity.

- *Bidirectional Interpolation:* Uses successive frames also and achieves high degree of compression. Successive frames are useful for area that are uncovered in a frame. This new part of the frame is not available in past frames. Effect of noise and errors can be reduced by averaging between past and future references. There is a trade-off in that if the frequency of B-frames is high, the correlation between them and their references decrease.

3.1.4 Motion Representation in MPEG

In most cases, predictive coding only makes sense for parts of an image and not for the whole image. Each image is therefore divided into areas called macroblocks which is partitioned into 16 * 16 pixels for the luminance components and 8 * 8 pixels for the chrominance components. A macro block can be of the following types: Intra coded, forward-predicted, backward-predicted, average (which has both forward and backward predicted blocks). A macro block contains the following two types of information:

- *Motion vector:* The motion vector stores the displacement between the spatial locations of the macro blocks. One motion vector is required for forward or backward predicted blocks and two are required for average blocks. Adjacent motion vectors are encoded using differential encoding techniques as they differ only slightly.
- *Difference between the encoded macro block and its predictor blocks:* DCT based coding techniques are used to encode it.

3.1.5 Motion Estimation in MPEG

Motion vector is obtained by searching for a block where there is minimum mismatch between the block being encoded and its predictor. Exhaustive search is costly. So, there is a trade-off between the quality of motion vector and the complexity of motion estimation process.

3.2 Need for Hierarchical/Scalable profiling

The need for hierarchical/scalable profiling was brought about with the aim to support applications such as compatible terrestrial TV/HDTV, packet-network video systems, backward compatibility with existing standards, and other applications like hand-held devises for which multilevel coding is required. Encoding is done in a way such that there is base layer and multiple enhancement layers. In progressive transmission, base layer is sent in the first stage and depending on the need/capacity of the client, other layers can be sent.

3.2.1 Fixed-resolution hierarchies

- All images have the same spatial resolution.
- Value at a pixel is refined with each level of hierarchy. For example, the base layer might have a resolution of 2 bits/pixel and enhancement layer has 4 bits/pixel and so on.

3.2.2 Variable-resolution hierarchies

- Different versions of the same image differing in spatial resolutions are kept.
- Upsampling technique is used where the base layer is the full image and each higher level results in reduction in spatial resolution.

3.3 MPEG-2

MPEG-2 achieves data rates of up to 100 Mbits/s. It strives for a higher resolution, similar to the digital video studio standard leading to HDTV.

In MPEG-2, scaling may act on the following parameters:

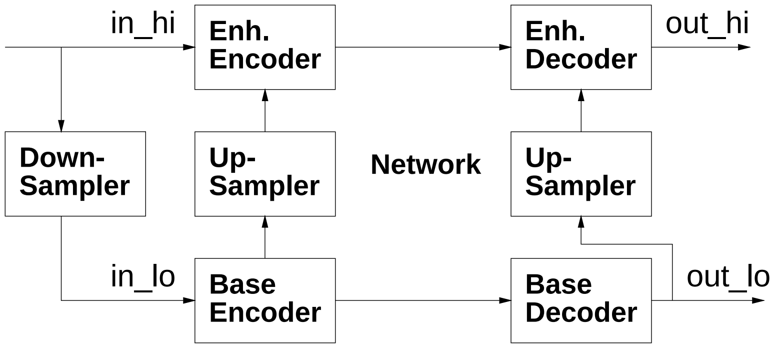

Figure 3.2: Spatial scaling

- Spatial scaling: As shown in fig 3.2, spatial scalability generally starts with a base layer at a lower resolution and adds an enhancement layer at higher resolution. The input source video is preprocesses to create the lower resolution image. This is independently coded, perhaps with MPEG-1 coding. In the enhancement layer, the differences between the interpolated version of the base layer and source image are coded.
- Temporal scaling: It provides an extension to higher temporal picture rates while maintaining backward compatibility with lower rates.
- Chroma scaling:
  - SNR scalability: It uses different quantizers for the base and enhancement layers.
  - Data partitioning: Some applications sometime have two channels available for transmission. The video bitstream can be split so that one channel contains all the key headers, motion vectors and low frequency DCT coefficients and the other channel has less critical information.