Module 4

CHAPTER 4

Digital Video

Introduction

Video is the technology of electronically capturing, recording, processing, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion. Human eye is able to distinguish approximately 20 images per second. Thus, when more than 20 images are displayed per second, it is possible to mislead the eye and create the illusion of an animated image. The fluidity of a video is characterized by the number of images per second (*frame rate*), expressed in *FPS* (*Frames per second*).

Video is a medium of communication that delivers more information per second than any other element of multimedia we have considered. We are used to seeing analog video on TV and are well aware of its impact on our life. Nowadays, a lot of video we see on TV and in movies has a digital component. For example, many of the special effects that you see in movies are digitally generated using the computer.

Syllabus Topic : Digitization of Video

4.1 Digitization of Video

Video digitizing is the process of capturing and converting and storing video images for use by a computer.

Analog video recordings have been produced in a multitude of physical formats, including 8mm, 16mm, and 35mm film, 4 sizes of reel-to-reel tape, and videocassettes in 8mm (VHS and S-VHS) and 1/2" (Betacam and M-II), each requiring a specific player. Digital video camcorders have recently become popular, providing quality video at an affordable price, but even digital video camcorders initially store the data on magnetic tapes. All of these media are vulnerable to physical decay. There are clear advantages in transferring video to a computer, including easier presentation on the Internet, distribution on DVD or CD, and consolidation from multiple physical formats.

To capture (transfer to computer) video from a digital video camcorder, the main requirements are the proper cable and a video editing application. Digitization from an analog format, such as VHS or 16mm film, requires analog-to-digital conversion, preferably using a video capture card.

Video is an especially complex multimedia format, requiring the synchronization of a series of still images with a soundtrack. Digitization of video requires a codec, an algorithm that determines how to encode and (usually) compress the data electronically. The file can then be stored in a number of formats.

☞ **The main advantages of the digital medium are :**

- Random Access allowed by the digital format enables us to quickly jump to any point in a movie. In analog format, we have to wind the tape backward and forward to reach a point in the movie.
- Digital format also allows us to quickly cut, paste, or edit video in other ways. It also allows easy addition of special effects.
- It is easy to duplicate digital video without loss of quality. In the analog tapes, video producers lose some quality after the first time they edit the video. This loss in quality is known as generation loss. Now video producers can convert real-life video that they have shot into digital format and edit it without losing the original quality of the film.
- Finally, digital video allows for interactivity. Currently this is not widely available, but once we get digital TVs it should be possible for us to have interactive TV.

Syllabus Topic : Types of Video Signals

Types of Video Signals

Video signals can be organized in three different ways : Component video, Composite video and S-video.

4.2.1 Component Video

Component video is a video signal that has been split into two or more components. In popular use, it refers to a type of analog video information that is transmitted or stored as three separate signals. component analog video signals do not use R, G, and B components but rather a colorless component, termed luma, which provides brightness information (as in black-and-white video). This combines with one or more color-carrying components, termed chroma, that give only color information.

In component video, the luminance (Y) and two color difference signals (U and V or I and Q) are separated into three separate analog signals that can be transmitted over three separate wires or stored in three separate tracks on an analog tape, or digitized separately. Component video is used in professional video production and provides the best quality and the most accurate reproduction of colors. The professional Betacam SP video cameras use component video.

4.2.2 Composite Video

Composite video signals are analog signals that combine luminance and chrominance (color) information in a single analog signal that can be transmitted over a single wire or stored in a single track on an analog magnetic tape.

The NTSC video signals used by commercial television sets in the United States and Japan are an example of composite signals. Composite video is particularly prone to errors in reproducing exact colors due to the overlap of the color and luminance signals.

Most of the analog home video equipment records a signal in the composite format and a composite video interface is used to connect a VHS tape player, game consoles or a DVD player to the television.

In composite video, three source signals are combined with sync pulses to form a composite video signal. The three source signals are referred as YUV in which Y represents the brightness of the picture and it also includes the synchronizing pulses. The color information is carried between U and V. However, two orthogonal phases of a color carrier signal are mixed with them in the first place to form a signal called as chrominance. The Y signal and the UV signal are then combined together and this is equivalent to the frequency-division multiplexing. The signals are compressed and then channeled through a single wire. Comb filter present in the television set is used to separate the signals. This results in degradation of signal quality.

4.2.3 S-video

S-video is one of a number of methods of separating a video signal into different components for transmission from a video cassette recorder or playback machine to a television set or video monitor. The technology was introduced to the market by JVC in 1987 as "separate video," which was quickly shortened to "s-video." S-video cables are the cables that connect two devices that are equipped with s-video capability to transfer the signal from one to another.

S-video was one of a number of enhancements in bringing the signal from the video cassette player to the television, and separates the video signal into *luma,* or luminescence, and *chroma,* or color.

S-video cables carry four or more wires wrapped together in an insulated sleeve, with S-video connectors at either end. The most common S-video connector has four pins – one for the chroma signal, one for the luma, and two ground wires, one for each signal.

S-Video is commonly used throughout the world with relative popularity. It is found on consumer TVs, DVD players, high-end video cassette recorders, digital TV receivers, DVRs, game consoles, and graphics cards. S-Video cables are used for computer-to-TV output for business or home use.

4.3 Digital Video

Digital video consists in showing a succession of digital images. Since these digital images are displayed at a certain rate, it is possible to know the necessary video display rate, i.e. the number of bytes displayed (or transferred) per unit of time. Thus, the necessary rate to display a video (in bytes per second) is equal to the size of the image multiplied by the number of images per second.

There are many advantages of digital representation for video, some of them are :

- Video can be stored on digital devices or in memory, ready to be processed and integrated into various multimedia applications.
- Direct access is possible, which makes non liner video editing.
- Repeated recording does not degrade image quality
- Ease of encryption and better tolerance to channel noise

4.3.1 Chroma Subsampling

Chroma subsampling is the process whereby the color information in the image is sampled at a lower resolution than the original. A reduced color resolution in digital component video signals. To accommodate storage and bandwidth limitations, the two color components (Cb, Cr) in digital video signals are compressed by sampling them at a lower rate than the brightness (Y). Color information is actually discarded.

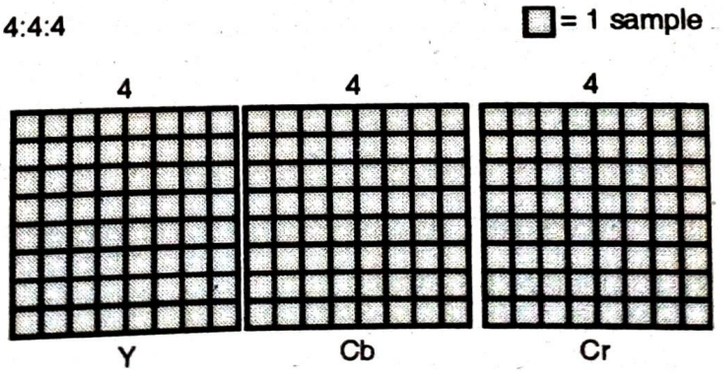
YCbCr is the YUV color space recorded digitally. Y is brightness (luma), and Cb and Cr are the U and V color difference signals. In chroma subsampling, only the colors are compressed, not the luma because the eye is more sensitive to brightness than to the color components.

☞ Various levels of YCbCr subsampling

YCbCr is designated as 4:n:n. The 4 represents a sampling rate of 13.5 MHz, which is the standard frequency for digitizing analog NTSC, PAL and SECAM. The next two digits represent the Cb and Cr rate. Review the illustrations below for details. Each 8x8 matrix represents a "macroblock" of 64 pixels in a video frame. The pink squares are the pixel locations where the sample is taken. Sony's HDCAM uses a different notation because it compresses both the luma and the colors

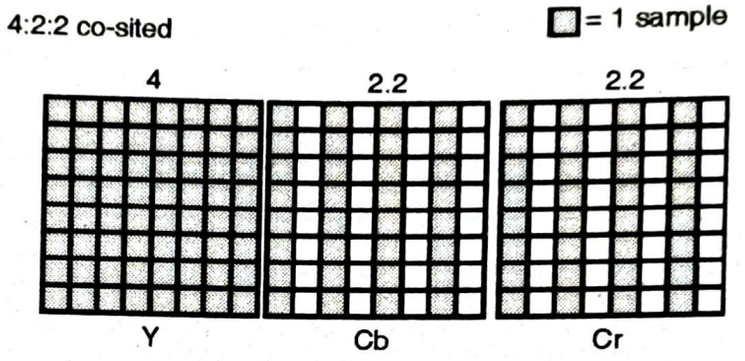
☞ 4:4:4 (Cb/Cr Same as Luma)

The resolution of chrominance information (Cb & Cr) is preserved at the same rate as the luminance (Y) information also known as 1x1 (or subsampling disabled). Cb and Cr are sampled at the same full rate as the luma. MPEG-2 supports 4:4:4 coding, but having the same number of color difference samples as the luma is considered overkill and not worth the additional bandwidth to transmit it. When video is converted from one color space to another, it is often resampled to 4:4:4 first.



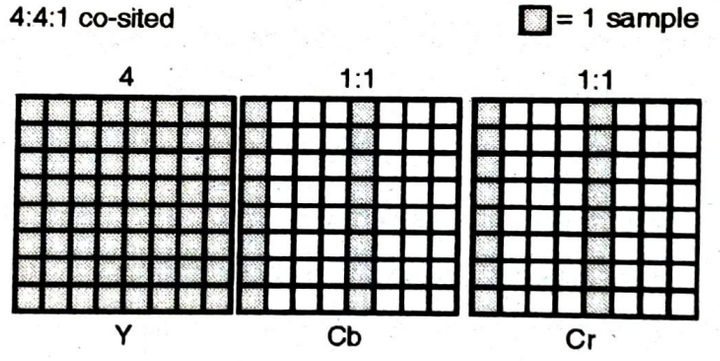
4:2:2 (1/2 the Luma Samples)

Half of the horizontal resolution in the chrominance is dropped (Cb & Cr), while the full resolution is retained in the vertical direction, with respect to the luminance. This is also known as 2 × 1 chroma subsampling, and is quite common for digital cameras. Cb and Cr are sampled at half the horizontal resolution of Y. Co-sited means that Cb/Cr samples are taken at the same time as Y. 4:2:2 color sampling is widely used and considered very high quality. It is used for prosumer and professional digital video recording, including DV (at 50 Mbps), Digital Betacam and DVCPRO 50 and is an option in MPEG-2.



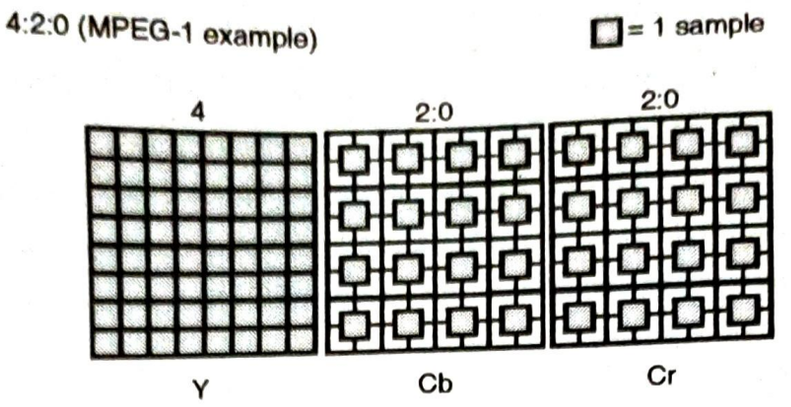
4:1:1 (1/4 the Luma Samples)

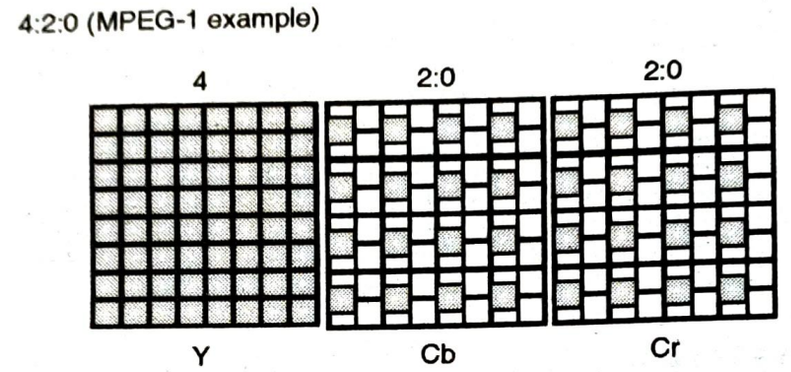
Only a quarter of the chrominance information is preserved in the horizontal direction with respect to the luminance information. I don't think this format is nearly as common as the other variations. Cb and Cr are sampled at one quarter the horizontal resolution. Co-sited means that Cb/Cr samples are taken at the same time as the Y. Co-sited 4:1:1 is used in DV, DVCAM and DVCPRO formats.



4:2:0 (1/4 the Luma Samples)

With respect to the information in the luminance channel (Y), the chrominance resolution in both the horizontal and vertical directions is cut in half. This form is also known as 2 × 2 chroma subsampling. The zero in 4:2:0 means that Cb and Cr are sampled at half the vertical resolution of Y. MPEG-1 and MPEG-2 use 4:2:0, but the samples are taken at different intervals. By the time MPEG-2 came along, it was known that 4:2:2 coding was often converted to 4:2:0, which is why MPEG-2 sampling more closely lines up with the 4:2:2 pattern. H.261/263 also uses 4:2:0.





4.3.2 CCIR Standards for Digital Video

CCIR is the Consultative Committee for International Radio, and one of the most important standards it has produced is CCIR-601, for component digital video. This standard has since become standard ITU-R-601, an international standard for professional video applications. This standard is adopted by certain digital video formats including the popular DV video. The CCIR 601 standard uses an interlaced scan, so each field has only half as much vertical resolution

CIF stands for Common Intermediate Format specified by the CCITT. The idea of CIF is to specify a format for lower bitrate. CIF is about the same as VHS quality. It uses a progressive (non-interlaced) scan. QCIF stands for "Quarter-CIF". All the CIF/QCIF resolutions are evenly divisible by 8, and all except 88 are divisible by 16; this provides convenience for block-based video coding in H.261 and H.263.

4.3.3 HDTV (High Definition TV)

High-definition television (or **HDTV**, or just **HD**) refers to video having resolution substantially higher than traditional television systems (standard-definition TV, or SDTV, or SD). HD has one or two million pixels per frame, roughly five times that of SD. Early HDTV broadcasting used analog techniques, but today HDTV is digitally broadcast using video compression.

The first generation of HDTV was based on an analog technology developed by Sony and NHK in Japan in the late 1970s. MUSE (MUltiple sub-Nyquist Sampling Encoding) was an improved NHK HDTV with hybrid analog/digital technologies that was put in use in the 1990s. It has 1,125 scan lines, interlaced (60 fields per second), and 16:9 aspect ratio. Since uncompressed HDTV will easily demand more than 20 MHz bandwidth, which will not fit in the current 6 MHz or 8 MHz channels, various compression techniques are being investigated. It is also anticipated that high quality HDTV signals will be transmitted using more than one channel even after compression.

HDTV broadcast systems are identified with three major parameters :

- **Frame size** in pixels is defined as *number of horizontal pixels* × *number of vertical pixels*, for example *1280* × *720* or *1920* × *1080*. Often the number of horizontal pixels is implied from context and is omitted, as in the case of *720p* and *1080p*.
- **Scanning system** is identified with the letter ***p*** for progressive scanning or ***i*** for interlaced scanning.
- **Frame rate** is identified as number of video frames per second. For interlaced systems an alternative form of specifying number of fields per second is often used.

If all three parameters are used, they are specified in the following form : *[frame size][scanning system][frame or field rate]* or *[frame size]/[frame or field rate][scanning system]*.

Often, frame size or frame rate can be dropped if its value is implied from context. In this case the remaining numeric parameter is specified first, followed by the scanning system.

For example, *1920×1080p25* identifies progressive scanning format with 25 frames per second, each frame being 1,920 pixels wide and 1,080 pixels high. The *1080i25* or *1080i50* notation identifies interlaced scanning format with 25 frames (50 fields) per second, each frame being 1,920 pixels wide and 1,080 pixels high. The *1080i30* or *1080i60* notation identifies interlaced scanning format with 30 frames (60 fields) per second, each frame being 1,920 pixels wide and 1,080 pixels high. The *720p60* notation identifies progressive scanning format with 60 frames per second, each frame being 720 pixels high; 1,280 pixels horizontally are implied.

There is no standard for HDTV color support. Until recently the color of each pixel was regulated by three 8-bit color values, each representing the level of red, blue, and green which defined a pixel color. Together the 24 total bits defining color yielded just under 17 million possible pixel colors. Recently some manufacturers have produced systems that can employ 10 bits for each color (30 bits total) which provides for a palette of 1 billion colors, saying that this provides a much richer picture, but there is no agreed way to specify that a piece of equipment supports this feature.

Syllabus Topic : File Formats - MPEG Video

4.4 Computer Video Format

Video can be stored in many different formats. Some popular video formats are as follow :

1. The AVI Format

- The AVI (Audio Video Interleave) format was developed by Microsoft.
- The AVI format is supported by all computers running Windows, and by all the most popular web browsers. It is a very common format on the Internet, but not always possible to play on non-Windows computers.
- Videos stored in the AVI format have the extension .avi.

2. The Windows Media Format

- The Windows Media format is developed by Microsoft.
- Windows Media is a common format on the Internet, but Windows Media movies cannot be played on non-Windows computer without an extra (free) component installed. Some later Windows Media movies cannot play at all on non-Windows computers because no player is available.
- Videos stored in the Windows Media format have the extension .wmv.

The MPEG Format

- The MPEG (Moving Pictures Expert Group) format is the most popular format on the Internet. It is cross-platform and supported by all the most popular web browsers.
- Videos stored in the MPEG format have the extension .mpg or .mpeg.

4. The QuickTime Format

- The QuickTime format is developed by Apple.
- QuickTime is a common format on the Internet, but QuickTime movies cannot be played on a Windows computer without an extra (free) component installed.
- Videos stored in the QuickTime format have the extension .mov.

5. The RealVideo Format

- The RealVideo format was developed for the Internet by Real Media.
- The format allows streaming of video (on-line video, Internet TV) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.
- Videos stored in the RealVideo format have the extension .rm or .ram.

6. The Shockwave (Flash) Format

- The Shockwave format was developed by Macromedia.
- The Shockwave format requires an extra component to play. This component comes preinstalled with the latest versions of Netscape and Internet Explorer.
- Videos stored in the Shockwave format have the extension .swf.

Syllabus Topic : Compression - MPEG

4.5 Video Compression

Video compression is essential before transmitting or storing video data. A video encoder utilizes the redundancies in the raw data to produce a compressed output. Two types of redundancies are used in this respect. One is spatial redundancy and the other one is temporal redundancy. An encoder exploits both redundancies during compression. In a video frame, neighboring pixels usually have similar color information. This is called spatial redundancy.

Consecutive frames (or very closely located frames) of a video sequence, which have been captured during a very short time interval, have similar scenes with or without slight displacement. For this reason, they are often highly correlated. This type of redundancy is called temporal redundancy.

Discrete Cosine Transformation (DCT) is applied on video data to remove spatial redundancy and Motion Compensation is applied for removing temporal redundancy.

The basic concept of motion compensation arises from the fact that though the consecutive or very closely located frames have similar scenes due to temporal redundancy, a little displacement of the objects occurs from frame to frame. Finding this displacement is called motion vector estimation and the amount of displacement is called motion vector. Basically, the huge gain in data compression comes from the fact that, instead of encoding an entire image frame, it is far more efficient to encode only the displacement vectors.

Motion estimation techniques form the core of video compression and video processing applications. Motion estimation extracts motion information from the video sequence. The motion is typically represented using a motion vector (x, y). The motion vector indicates the displacement of a pixel or a pixel block from the current location due to motion.

Motion information is used in video compression to find best matching block in reference frame to calculate low energy residue, used in scan rate conversion to generate temporally interpolated frames. It is also used in applications such motion compensated de-interlacing, video stabilisation, motion tracking etc.

Motion vector estimation is the most computationally expensive and resource hungry operation in the entire compression process. That is why it is important to devise improved and efficient motion vector estimation algorithm. There are different approaches for finding motion vectors like pel-recursive technique, block-based

technique, optical flow etc. Among these, block-based technique is the most widely accepted technique due to its effectiveness and simplicity.

Block-based technique divides a frame into blocks and searches them in the previous frame, which is generally called the Reference Frame. Here, the displacement of a block in the current frame with respect to the reference frame is the motion vector. A very good and thorough searching in the reference frame is necessary to get an exact or a very close match. At the same time it is imperative that the searching remains computationally inexpensive.

4.5.1 Block Matching Algorithm

Block Matching Algorithm (BMA) is the most popular motion estimation algorithm. BMA calculates motion vector for an entire block of pixels instead of individual pixels. The same motion vector is applicable to all the pixels in the block. This reduces computational requirement and also results in a more accurate motion vector since the objects are typically a cluster of pixels.

BMA algorithm is illustrated in Fig. 4.5.1. The current frame is divided into pixel blocks and motion estimation is performed independently for each pixel block. Motion estimation is done by identifying a pixel block from the reference frame that best matches the current block, whose motion is being estimated. The reference pixel block is generated by displacement from the current block's location in the reference frame. The displacement is provided by the Motion Vector (MV). MV consists of is a pair (x, y) of horizontal and vertical displacement values.

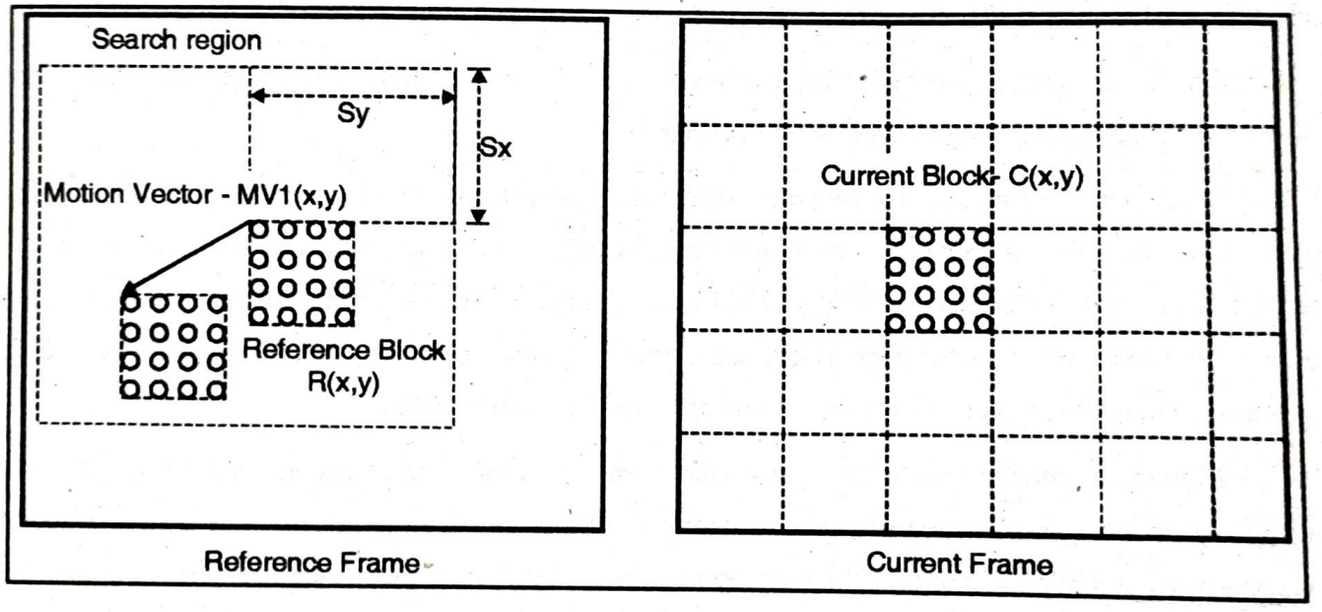


Fig. 4.5.1 : Block matching algorithm

There are various criteria available for calculating block matching.

Two popular criteria are listed below :

$$\text{Sum of Square Error (SSE)} = \sum_{x=1}^{x=N} \sum_{y=1}^{y=N} (C(x, y) - R(x, y))^2$$

$$\text{Sum of Absolute Difference (SAD)} = \sum_{x=1}^{x=N} \sum_{y=1}^{y=N} |C(x, y) - R(x, y)|$$

SSE provides a more accurate block matching, however requires more computations. SAD provides fairly good match at lower computational requirement. Hence it is widely used for block matching. There are various other criteria also available such as cross correlation, maximum matching pixel count etc. The reference pixel blocks are generated only from a region known as the search area. Search region defines the boundary for the motion vectors and limits the number of blocks to evaluate. The height and width of the search region is dependent on the motion in video sequence. The available computing power also determines the search range. Bigger search region requires more computation due to increase in number of evaluated candidates.

Typically the search region is kept wider (i.e. width is more than height) since many video sequences often exhibit panning motion. The search region can also be changed adaptively depending upon the detected motion. The horizontal and vertical search range, Sx & Sy, define the search area (+/-Sx and +/- Sy) as illustrated in Fig. 4.5.1.

4.5.2 Full Search Block Matching

Full search block matching algorithm evaluates every possible pixel block in the search region. Hence, it can generate the best block matching motion vector. This type of BMA can give least possible residue for video compression.

But, the required computations are prohibitively high due to the large amount of candidates to evaluate. The number of candidates to evaluate are (2Sx + 1)*(2Sy + 1). Hence, full search is typically not used. Also, it does not guarantee consistent motion vectors required for video processing applications.

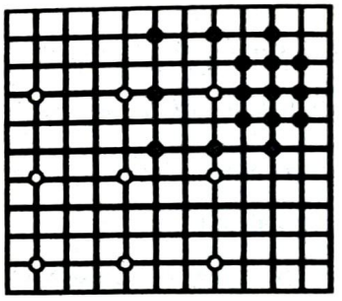


Fig. 4.5.2

There are several other fast block-matching algorithms, which reduce the number of evaluated candidates yet try to keep a good block matching accuracy. Note that since these algorithms test only limited candidates, they might result in selecting a candidate corresponding to local minima, unlike full search, which always results in global minima.

4.5.3 Three Step Search

In a three-step search (TSS) algorithm, the first iteration evaluates nine candidates as shown in Fig. 4.5.3.

The candidates are centered around the current block's position. The step size for the first iteration is typically set to half the search range.

During the next iteration, the search centre is shifted to the best matching candidate from the first iteration. Also, the step size is reduced by half. The same process continues till the step size becomes equal to one pixel. This is the last iteration of the three-step search algorithm.

The best matching candidate from this iteration is selected as the final candidate. The motion vector corresponding to this candidate is selected for the current block.

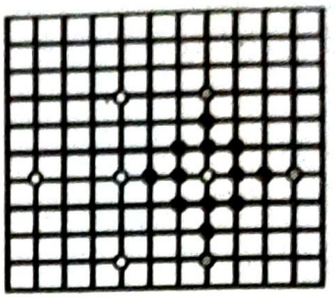


Fig. 4.5.3 : 2D logarithmic search

The number of candidates evaluated during three-step search is very less compared to the full search algorithm. The number of evaluated candidate is fixed depending upon the step size set during the first iteration.

4.5.4 2D Logarithmic Search

2D Logarithmic search is another algorithm, which tests limited candidates. It is similar to the three-step search.

During the first iteration, a total of five candidates are tested. The candidates are centered around the current block location in a diamond shape. The step size for first iteration is set equal to half the search range.

For the second iteration, the centre of the diamond is shifted to the best matching candidate. The step size is reduced by half only if the best candidate happens to be the centre of the diamond. If the best candidate is not the diamond centre, same step size is used even for second iteration.

In this case, some of the diamond candidates are already evaluated during first iteration. Hence, there is no need for block matching calculation for these candidates during the second iteration. The results from the first iteration can be used for these candidates. The process continue still the step size becomes equal to one pixel. For this iteration all eight surrounding candidates are evaluated. The best matching candidate from this iteration is selected for the current block. The number of evaluated candidate is variable for the 2D logarithmic search. However, the worst case and best case candidates can be calculated.

4.5.5 Hierarchical Block Matching

Hierarchical Block Matching algorithm is a more sophisticated motion estimation technique. Hierarchical Block Matching motion estimation provides consistent motion vectors by successively refining the motion vector at different resolutions.

In Hierarchical motion estimation, a pyramid of reduced resolution video frame is formed. The original video frame forms the highest resolution image and the other images in the pyramid are formed by down sampling the original image.

A simple bi-linear down sampling can be used. This is illustrated in Fig. 4.5.4. A three-level hierarchical search in which the original image is at Level 0, images at Levels 1 and 2 are obtained by down-sampling from the previous levels by a factor of 2, and the initial search is conducted at Level 2. Since the size of the macroblock is smaller and S(search range) can also be proportionally reduced, the number of operations required is greatly reduced .

The block size of N × N at the highest resolution is reduced to (N/2) × (N/2) in the next resolution level. Similarly, the search range is also reduced. The motion estimation process starts at the lowest resolution. Typically, full search motion estimation is performed for each block at the lowest resolution. Since the block size and the search range are reduced, it does not require large computations.

The motion vectors from lowest resolution are scaled and passed on as candidate motion vectors for each block to next level. At the next level, the motion vectors are refined with a smaller search area. A simpler motion estimation algorithm and a small search range is enough at close to highest resolution since the motion vectors are already close to accurate motion vectors.

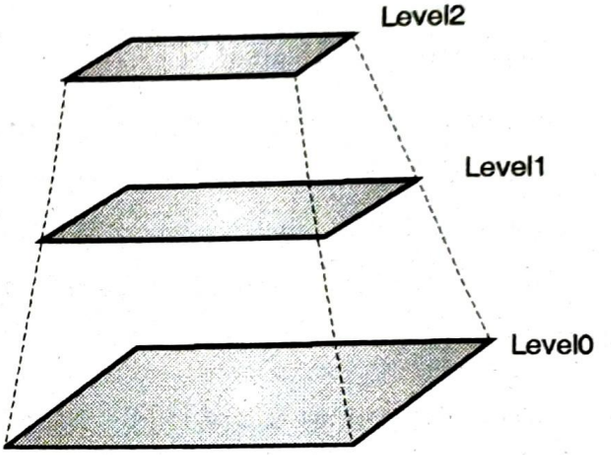


Fig. 4.5.4

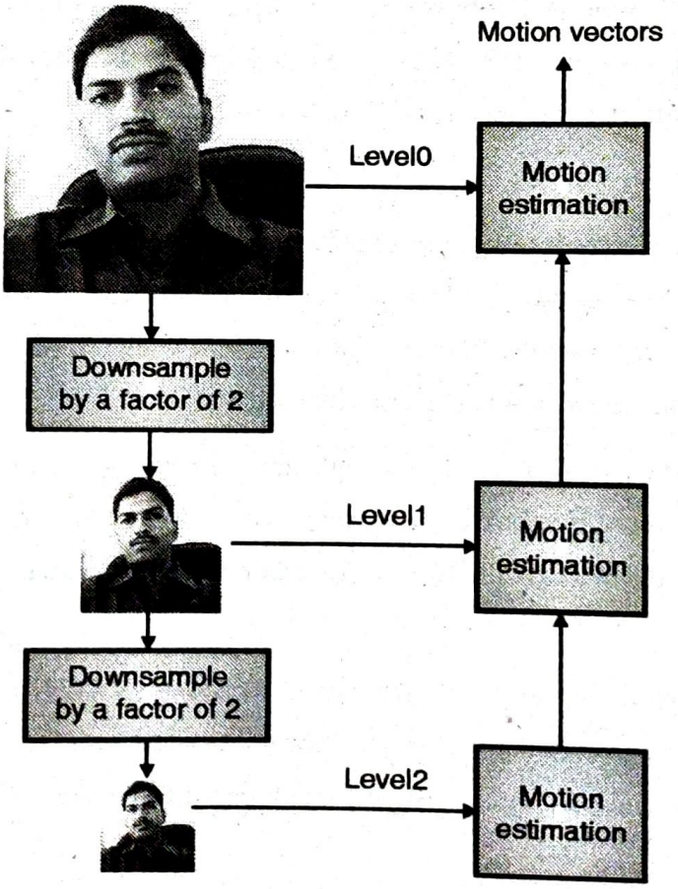


Fig. 4.5.5 : Three level hierarchical search for motion vectors

4.6 Video Compression Standards

MPEG stands for the Moving Picture Experts Group. MPEG is an ISO/IEC working group, established in 1988 to develop standards for digital audio and video formats. There are five MPEG standards being used or in development. Each compression standard was designed with a specific application and bit rate in mind, although MPEG compression scales well with increased bit rates.

They include :

MPEG-1

Designed for up to 1.5 Mbit/sec.

Standard for the compression of moving pictures and audio. This was based on CD-ROM video applications, and is a popular standard for video on the Internet, transmitted as .mpg files. In addition, level 3 of MPEG-1 is the most popular standard for digital compression of audio–known as MP3. MPEG-1 is the standard of compression for Video CD, the most popular video distribution format throughout much of Asia.

MPEG-2

Designed for between 1.5 and 15 Mbit/sec.

Standard on which Digital Television set top boxes and DVD compression is based. It is based on MPEG-1, but designed for the compression and transmission of digital broadcast television. The most significant enhancement from MPEG-1 is its ability to efficiently compress interlaced video. MPEG-2 scales well to HDTV resolution and bit rates, obviating the need for an MPEG-3.

MPEG-4

Standard for multimedia and Web compression. MPEG-4 is based on object-based compression. Individual objects within a scene are tracked separately and compressed together to create an MPEG4 file. This results in very efficient compression that is very scalable, from low bit rates to very high. It also allows developers to control objects independently in a scene, and therefore introduce interactivity.

- **DV** is a high-resolution digital video format used with video cameras and camcorders. The standard uses DCT to compress the pixel data and is a form of lossy compression. The resulting video stream is transferred from the recording device via FireWire (IEEE 1394), a high-speed serial bus capable of transferring data up to 50 MB/sec.
- **H.261** is an ITU standard designed for two-way communication over ISDN lines (video conferencing) and supports data rates which are multiples of 64Kbit/s. The algorithm is based on DCT and can be implemented in hardware or software and uses intraframe and interframe compression. H.261 supports CIF and QCIF resolutions.

- **H.263** is based on H.261 with enhancements that improve video quality over modems. It supports CIF, QCIF, SQCIF, 4CIF and 16CIF resolutions.

Syllabus Topic : H.261

4.7 H.261

H.261 is video coding standard by the ITU. It was designed for data rates which are multiples of 64 Kbit/s, and is sometimes called p × 64 K bit/s (p is in the range 1-30). These data rates suit ISDN lines, for which this video codec was originally designed for. H.261 transport video stream using the real-time transport protocol, RTP, with any of the underlying protocols that carry RTP.

The coding algorithm is a hybrid of inter-picture prediction, transform coding, and motion compensation. The data rate of the coding algorithm was designed to be able to be set to between 40 Kbits/s and 2 Mbits/s. INTRA coding where blocks of 8 × 8 pixels each are encoded only with reference to themselves and are sent directly to the block transformation process. On the other hand INTER coding frames are encoded with respect to another reference frame. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codec compensate for motion. To remove any further redundancy in the transmitted bitstream, variable length coding is used.

H261 supports motion compensation in the encoder as an option. In motion compensation a search area is constructed in the previous (recovered) frame to determine the best reference macroblock.

H261 supports two image resolutions, QCIF (Quarter Common Interchange format) which is (144 × 176 pixels) and CIF (Common Interchange format) which is (288 × 352)

The video multiplexer structures the compressed data into a hierarchical bitstream that can be universally interpreted. The hierarchy has four layers :

1. Picture layer : Corresponds to one video picture (frame).
2. Group of blocks : Corresponds to 1/12 of CIF pictures or 1/3 of QCIF.
3. Macroblocks : Corresponds to 16x16 pixels of luminance and the two spatially corresponding 8 × 8 chrominance components.
4. Blocks : Corresponds to 8 × 8 pixels.

4.7.1 H.261 Encoder

The source coder operates on only non-interlaced pictures. Pictures are coded as luminance and two color difference components(Y, Cb, Cr). The Cb and Cr matrices are half the size of the Y matrix.

The three main elements in an H.261 encoder as illustrated in Fig. 4.7.1 are :

Prediction

H261 defines two types of coding. INTRA coding where blocks of 8x8 pixels each are encoded only with reference to themselves and are sent directly to the block transformation process. On the other hand INTER coding frames are encoded with respect to another reference frame

A prediction error is calculated between a 16 × 16 pixel region (macroblock) and the (recovered) correspondent macroblock in the previous frame. Prediction error of transmitted blocks (criteria of transmission is not standardized) are then sent to the block transformation process.

- Blocks are inter or intra coded.
- Intra-coded blocks stand alone.
- Inter-coded blocks are based on predicted error between the previous frame and this one.
- Intra-coded frames must be sent with a minimum frequency to avoid loss of synchronisation of sender and receiver.

Block Transformation

H261 supports motion compensation in the encoder as an option. In motion compensation a search area is constructed in the previous (recovered) frame to determine the best reference macroblock. Both the prediction error as well as the motion vectors specifying the value and direction of displacement between the encoded macroblock and the chosen reference are sent.

INTRA coding where blocks of 8 × 8 pixels each are encoded only with reference to themselves and are sent directly to the block transformation process. On the other hand INTER coding frames are encoded with respect to another reference frame. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codec compensate for motion. To remove any further redundancy in the transmitted bitstream, variable length coding is used.

Quantization and Entropy Coding

The purpose of this step is to achieve further compression by representing the DCT coefficients with no greater precision than is necessary to achieve the required quality. The number of quantizers are 1 for the INTRA dc coefficients and 31 for all others.

Entropy coding involves extra compression (non-lossy) is done by assigning shorter code-words to frequent events and longer code-words to less frequent events. Huffman coding is usually used to implement this step.

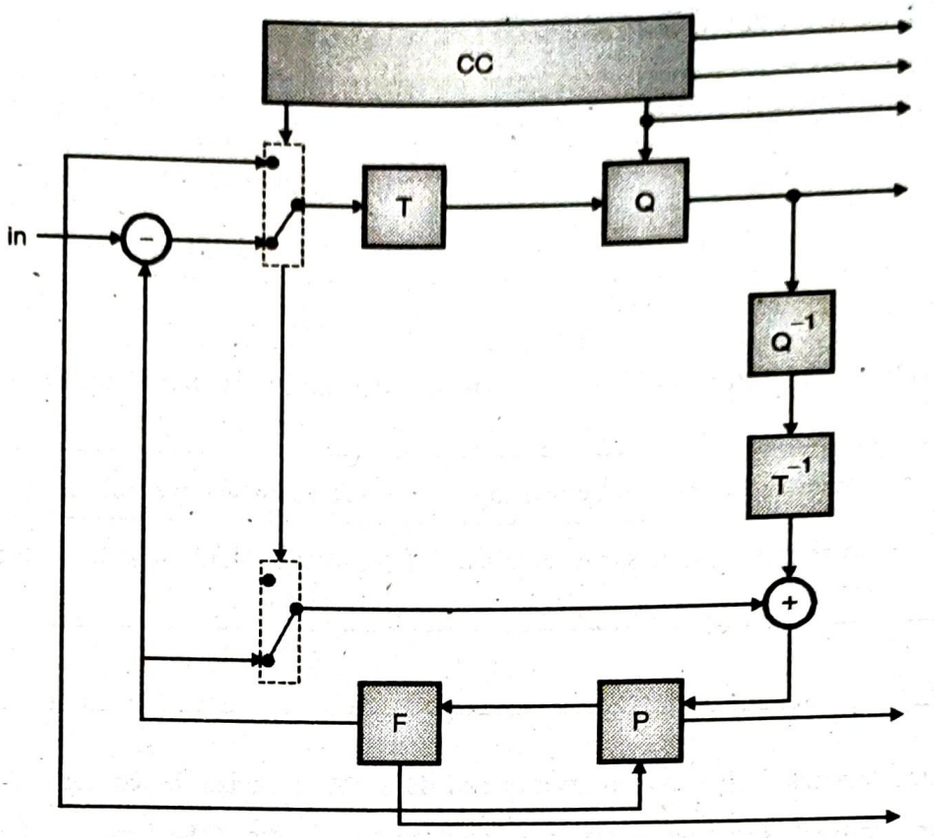


Fig. 4.7.1

In other words, for a given quality, we can lose coefficients of the transform by using less bits than would be needed for all the values This leads to a "coarser" picture. We can then entropy code the final set of values by using shorter words for the most common values and longer ones for rarer ones.

T	Transform	P	Flag for INTRA/INTER
Q	Quantiser	t	Flag for transmitted or not
P	Picture memory with motion compensated variable delay	qz	Quantiser indication
F	Loop filter	q	Quantizing index for transform co-efficients
CC	Coding control	v	Motion vector
		f	Switching on/off the loop filter

H261 Encoder

4.7.2 Comparison of MPEG and H.261

H.261	MPEG
It is based on JPEG.	It is based on H.261 and JPEG.
It encode video only.	It encodes audio and video.
It uses I and P frames.	It uses I, P and B frames.
It is best for video with little motion (e.g. video conferencing).	It is designed to handle moving picture components.
It is optimized for bandwidth efficiency and low delay.	It s less bandwidth efficient.
It is lossy algorithm with compression in space and time.	It is lossy algorithm with compression in space and time.

4.8 MPEG-1

MPEG-1 codes progressively-scanned images and does not recognize the concept of interlace, interlaced source video must be converted to a non-interlace format prior to encoding. The format of the coded video allows forward play and pause, typical coding and decoding methods allow random access, fast forward and reverse play also, the requirements for these functions are very much application dependent and different encoding techniques will include varying levels of flexibility to account for these functions.

Compression of the digitized video comes from the use of several techniques : Sub sampling of the chroma information to match the human visual system, differential coding to exploit spatial redundancy, motion compensation to exploit temporal redundancy, Discrete Cosine Transform (DCT) to match typical image statistics, quantization, variable length coding, entropy coding and use of interpolated pictures.

4.8.1 Algorithm Structure and Terminology

The MPEG hierarchy is arranged into layers. This layered structure is designed for flexibility and management efficiency, each layer is intended to support a specific function i.e. the sequence layer specifies sequence parameters such as picture size, aspect ratio, picture rate, bit rate etcetera , whereas the picture layer defines parameters such as the temporal reference and picture type. This layered structure improves robustness and reduces susceptibility to data corruption.

For convenience of coding, macroblocks are divided into six blocks of component Pixels - four luma and two chroma (Cr and Cb). Blocks are the basic coding unit and the DCT is applied at this block level. Each block contains 64 component Pixels arranged in an 8×8 array.

There are four picture types : I pictures or INTRA pictures, which are coded without reference to another pictures; P pictures or PREDICTED pictures which are coded using motion compensation from a previous picture;

B pictures or BIDIRECTION-ALLY predicted pictures which are coded using interpolation from a previous and a future picture and D pictures or DC pictures in which only the low frequency component is coded and which are only intended for fast forward search mode. B and P pictures are often called Inter pictures. Some other terminology that is often used are the terms M and N, M + 1 represents the number of frames between successive I and P pictures whereas N + 1 represents the number of frames between successive I pictures. M and N can be varied according to different applications and requirements such as fast random access.

A typical coding scheme will contain a mix of I, P and B pictures. A typical scheme will have and I picture every 10 to 15 pictures and two B pictures between successive I and P pictures.

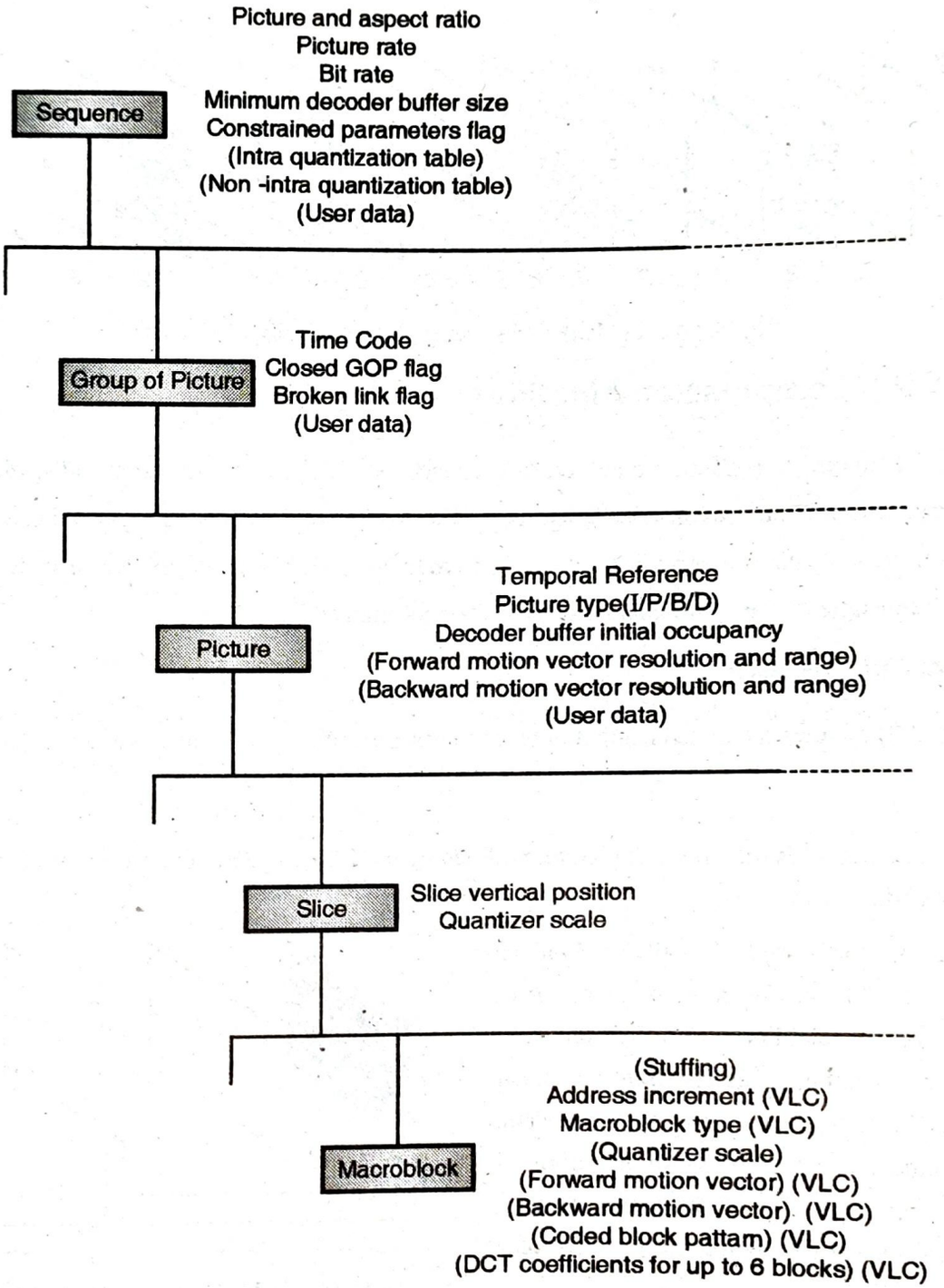


Fig. 4.8.1 : MPEG Bistream Hierarchy

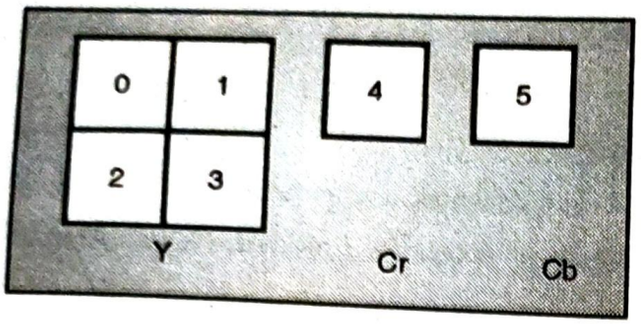


Fig. 4.8.2 : Macroblock Structure

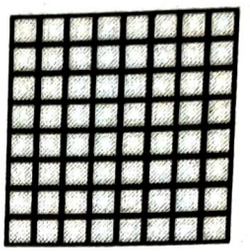


Fig. 4.8.3 : Block Structure

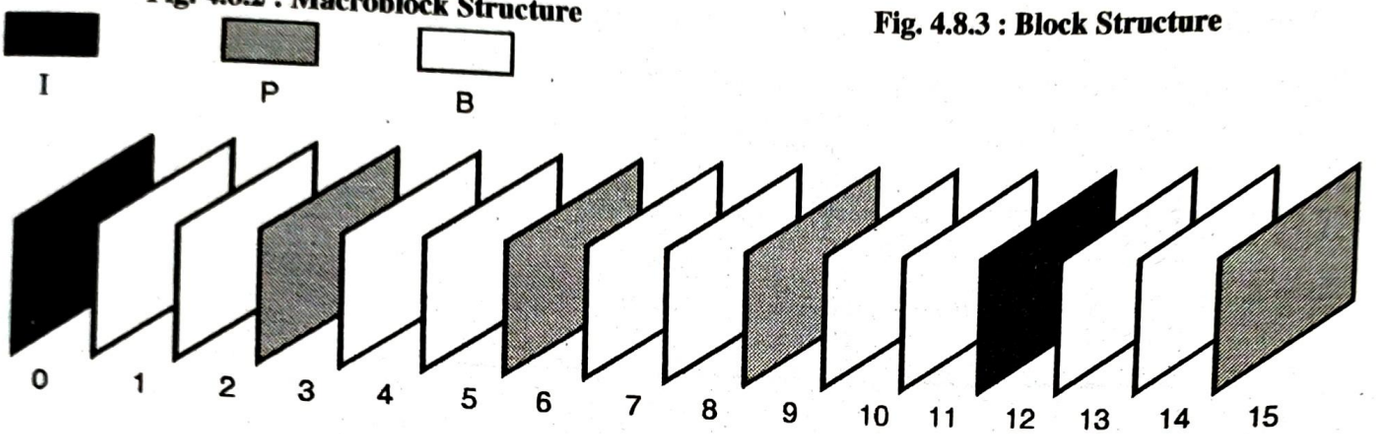


Fig. 4.8.4 : Typical sequence of pictures in display order

4.8.2 MPEG-1 Compression Algorithm

The MPEG-1 algorithm is based around two key techniques : Temporal compression and spatial compression. Temporal compression relies upon similarity between successive pictures using prediction and motion compensation whereas spatial compression relies upon redundancy within small areas of a picture and is based around the DCT transform, quantization and entropy coding techniques.

1) Temporal Compression

Inter (B and P) pictures are coded using motion compensation, primarily prediction and interpolation.

Prediction

The predicted picture is the previous picture modified by motion compensation. Motion vectors are calculated for each macroblock.

The motion vector is applied to all four luminance blocks in the macro block. The motion vector for both chrominance blocks is calculated from the luma vector. This technique relies upon the assumption that within a macroblock the difference between successive pictures can be represented simply as a vector transform (i.e. there is very little difference between successive pictures, the key difference being in position of the Pixels)

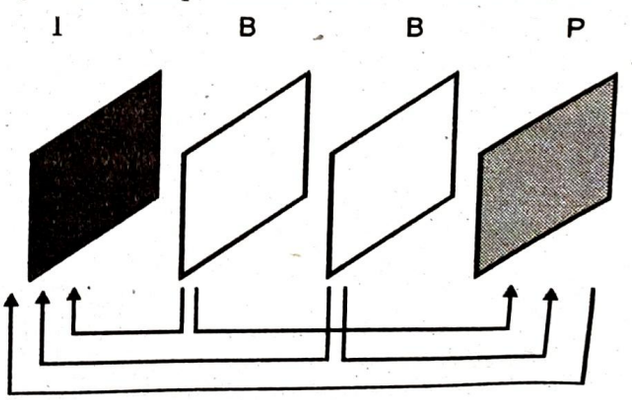


Fig. 4.8.5 : Make up of I, B and P pictures

Interpolation

Interpolation (or bidirectional prediction) generates high compression in that the picture is represented simply as an interpolation between the past and future I or P pictures (again this is performed on a macroblock level). Pictures are not transmitted in display order but in the order in which the decoder requires them to decode the bitstream (the decoder must of course have the reference picture(s) before any interpolated or predicted pictures can be decoded).

2) Spatial Compression

The spatial compression techniques are similar to those of JPEG, DCT, Quantization and entropy coding. The compression algorithm takes advantage of the redundancy within each block (8×8 Pixels).

The resulting compressed data stream is made up of a combination of spatial and temporal compression techniques which best suit the type of picture being compressed. Decoding is controlled through the use of MPEG system codes which are put into the data stream explaining how to reconstruct specific areas of picture.

Through a combination of techniques, MPEG-1 compression is designed to give good quality (typically similar or better quality to VHS) images from such storage media as CD-ROM. The quality is however, dependent upon the type of picture compressed and the level of redundancy within the sequence coded. Picture quality will also depend upon how well the sequence has been coded and which features are required - For Example : For fast random access, N will tend towards zero hence the quality of compression will deteriorate, if random access is not required then the number of P and B frames can increase, hence increasing the potential quality. The standard does not specify a method of compression but a syntax for the compressed data, this allows for differing compression techniques depending upon differing requirements. The decoding techniques are defined due to the nature of the compressed data stream.

This method allows for true flexibility in coding whilst retaining the format and hierarchy ensuring compatibility in the data stream and hence uniform readability.

4.9 MPEG-2

MPEG-2 is an extension of the MPEG-1 international standard for digital compression of audio and video signals. MPEG-1 was designed to code progressively scanned video at bit rates up to about 1.5 Mbit/s for applications such as CD-i (compact disc interactive). MPEG-2 is directed at broadcast formats at higher data rates; it provides extra algorithmic is 'tools' for efficiently coding interlaced video, supports a wide range of bit rates and provides for multichannel surround sound coding. This tutorial paper introduces the principles used for compressing video according to the MPEG-2 standard, outlines the general structure of a video coder and decoder, and describes the subsets ('profiles') of the toolkit and the sets of constraints on parameter values ('levels') defined to date.

1. Introduction

- Recent progress in digital technology has made the widespread use of compressed digital video signals practical.
- Standardisation has been very important in the development of common compression methods to be used in the new services and products that are now possible.
- This allows the new services to interoperate with each other and encourages the investment needed in integrated circuits to make the technology cheap.

2. Bit rate reduction principles

A bit rate reduction system operates by removing redundant information from the signal at the coder prior to transmission and re-inserting it at the decoder. A coder and decoder pair are referred to as a 'codec'. In video signals, two distinct kinds of redundancy can be identified.

- **Spatial and temporal redundancy :** Pixel values are not independent, but are correlated with their neighbours both within the same frame and across frames. So, to some extent, the value of a pixel is predictable given the values of neighbouring pixels.
- **Psychovisual redundancy :** The human eye has a limited response to fine spatial detail and is less sensitive to detail near object edges or around shot-changes. Consequently, controlled impairments introduced into the decoded picture by the bit rate reduction process should not be visible to a human observer.

Two key techniques employed in an MPEG code are intra-frame Discrete Cosine Transform (DCT) coding and motion-compensated inter-frame prediction. These techniques have been successfully applied to video bit rate reduction prior to MPEG, notably for 625-line video contribution standards at 34 Mbit/s and video conference systems at bit rates below 2 Mbit/s.

Step 1 : Intra-frame DCT coding

- A two-dimensional DCT is performed on small blocks (8 pixels by 8 lines) of each component of the picture to produce blocks of DCT coefficients (Fig. 4.9.1(a)).
- The magnitude of each DCT coefficient indicates the contribution of particular combination of horizontal and vertical spatial frequencies to the original picture block. The coefficient corresponding to zero horizontal and vertical frequency is called the DC coefficient.

The N × N two-dimensional DCT is defined as :

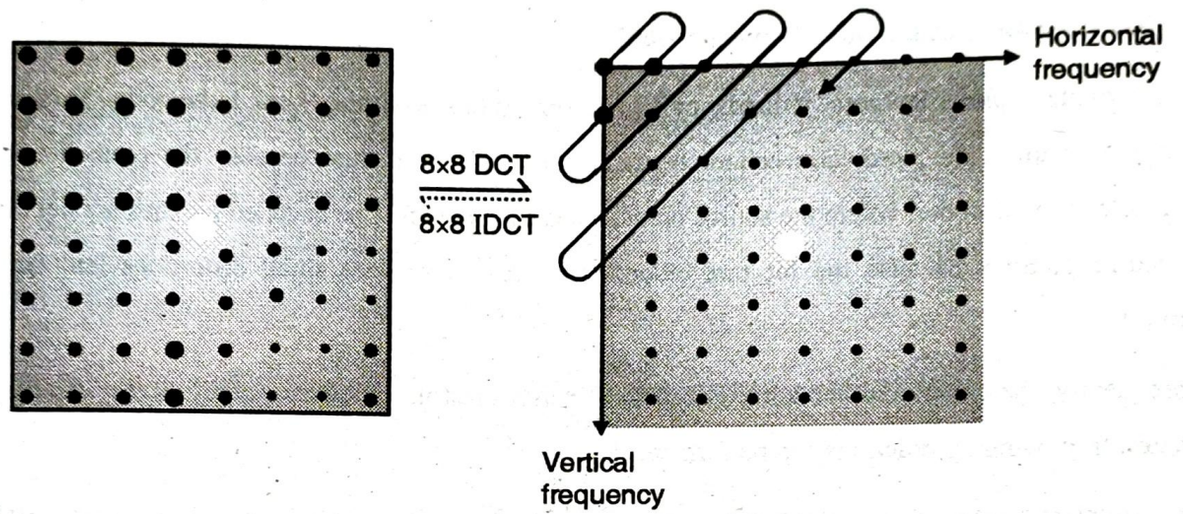
$$F(u,v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}$$

$$C(u), C(v) = \frac{1}{\sqrt{2}} \begin{cases} \text{for } u,v = 0 \\ 1 \quad \text{otherwise} \end{cases} \qquad \ldots(4.9.1)$$

The inverse DCT (IDCT) is defined as :

$$f(x,y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)\, C(v)\, F(u,v) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N} \qquad \ldots(4.9.2)$$

Where x, y are spatial co-ordinates in the image block u, v are co-ordinates in the DCT coefficient block.



(a) 8 × 8 image block **(b) 8 × 8 DCT coefficient block**

Fig. 4.9.1 : The distance cosine transform (DCT)

- Pixel value and DCT coefficient magnitude are represented by dot size.
- The DCT does not directly reduce the number of bits required to represent the block. In fact for an 8 × 8 block of 8 bit pixels, the DCT produces an 8 × 8 block of 11 bit coefficient (the range of coefficient values is larger than the range of pixel values.)
- The reduction in the number of bits follows from the observation that, for typical blocks from natural images, the distribution of coefficients is non-uniform. The transform tends to concentrate the energy into the low-frequency coefficients and many of the other coefficients are near-zero.
- The bit rate reduction is achieved by not transmitting the near-zero coefficients and quantising and coding the remaining coefficients as described below. The non-uniform coefficient distribution is a result of the spatial redundancy present in the original image block.

Step 2 : Quantisation

- The function of the coder is to transmit the DCT block to the decoder, in a bit rate efficient manner, so that it can perform the inverse transform to reconstruct the image.

- It has been observed that the numerical precision of the DCT coefficients may be reduced while still maintaining good image quality at the decoder. Quantisation is used to reduce the number of possible values to be transmitted, reducing the required number of bits.
- The degree of quantisation applied to each coefficient is weighted according to the visibility of the resulting quantisation noise to a human observer. In practice, this results in the high-frequency coefficients being more coarsely quantised than the low-frequency coefficients. Note that the quantisation noise introduced by the coder is not reversible in the decoder, making the coding and decoding process 'lossy'.

Step 3 : Motion-compensated inter-frame prediction

- The technique exploits temporal redundancy by attempting to predict the frame to be coded from a previous 'reference' frame. The prediction cannot be based on a source picture because the prediction has to be repeatable in the decoder, where the source pictures are not available (the decoded pictures are not identical to the source pictures because the bit rate reduction process introduces small distortions into the decoded picture.)
- Consequently, the coder contains a local decoder which reconstructs pictures exactly as they would be in the decoder, from which predictions can be formed.
- The simplest inter-frame prediction of the block being coded is that which takes the co-sited (i.e. the same spatial position) block from the reference picture. Naturally this makes a good prediction for stationary regions of the image, but is poor in moving areas. A more sophisticated method, known as motion-compensated inter-frame predication, is to offset any translational motion which has occurred between the block being coded and the reference frame and to use a shifted block from the reference frame as the prediction.
- One method of determining the motion that has occurred between the block coded and the reference frame is a 'block-matching' search in which a large number of trial offsets are tested by the coder using the luminance component of the picture. The 'best' offset is selected on the basis of minimum error between the block being coded and the prediction.

3. MPEG–2 Details

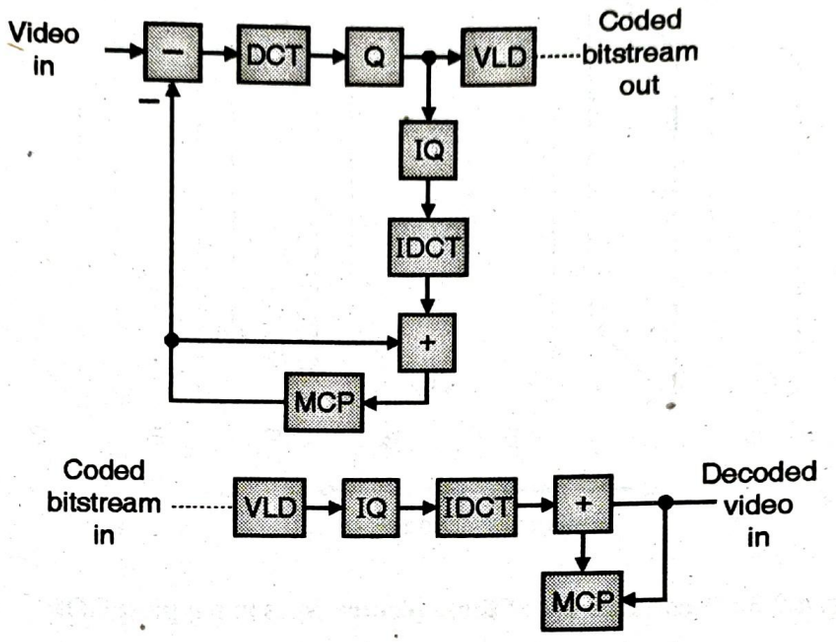
☞ Codec structure

In an MPEG-2 system, the DCT and motion-compensated interframe prediction are combined, as shown in Fig. 4.9.2. The coder subtracts the motion-compensated prediction from the source picture to form a 'prediction error' picture.

The prediction error is transformed with the DCT, the coefficients are quantised and these quantised values coded using a VLC. The coded luminance and chrominance prediction error is combined with 'side information'

required by the decoder, such as motion vectors and synchronising information, and formed into a bitstream for transmission.

In the decoder, the quantised DCT coefficients are reconstructed and inverse transformed to produce the prediction error. This is added to the motion-compensated prediction generated from previously decoded pictures to produce the decoded output.



(I)DCT = (inverse) discrete cosine transform
(I)Q = (inverse) quantization
MCP = motion-compensated prediction
VLC = variable-lenght coder
VLC = variable-lenght decoder

Fig. 4.9.2 : (a) Motion-compensated DCT coder; (b) Motion compensated DCT decoder

Picture types

In MPEG-2, three 'picture types' are defined. The picture type defines which prediction modes may be used to code each block.

- **'Intra' pictures (I-pictures)** are coded without reference to other pictures. Moderate compression is achieved by reducing spatial redundancy, but not temporal redundancy. They can be used periodically to provide access points in the bitstream where decoding can begin.
- **'Predictive' pictures (P-pictures)** can use previous I-or P-pictures for motion compensation and may be used as a reference for further prediction. Each block in a P-picture can either be predicted or intra-coded. By reducing spatial and temporal redundancy, P-pictures offer increased compression compared to I-pictures.
- **'Bidirectionally-predictive' pictures (B-pictures)** can use the previous and next I-or P-pictures for motion-compensation, and offer the highest degree of compression. Each block in a B-picture can be forward, backward or bidirectionally predicted or intra-coded.

- To enable backward prediction from a future frame, the coder recorders the pictures from natural 'display' order to 'bitstream' order so that the B-picture is transmitted after the previous and next pictures it references. This introduces a recording delay dependant on the number of consecutive B-pictures.

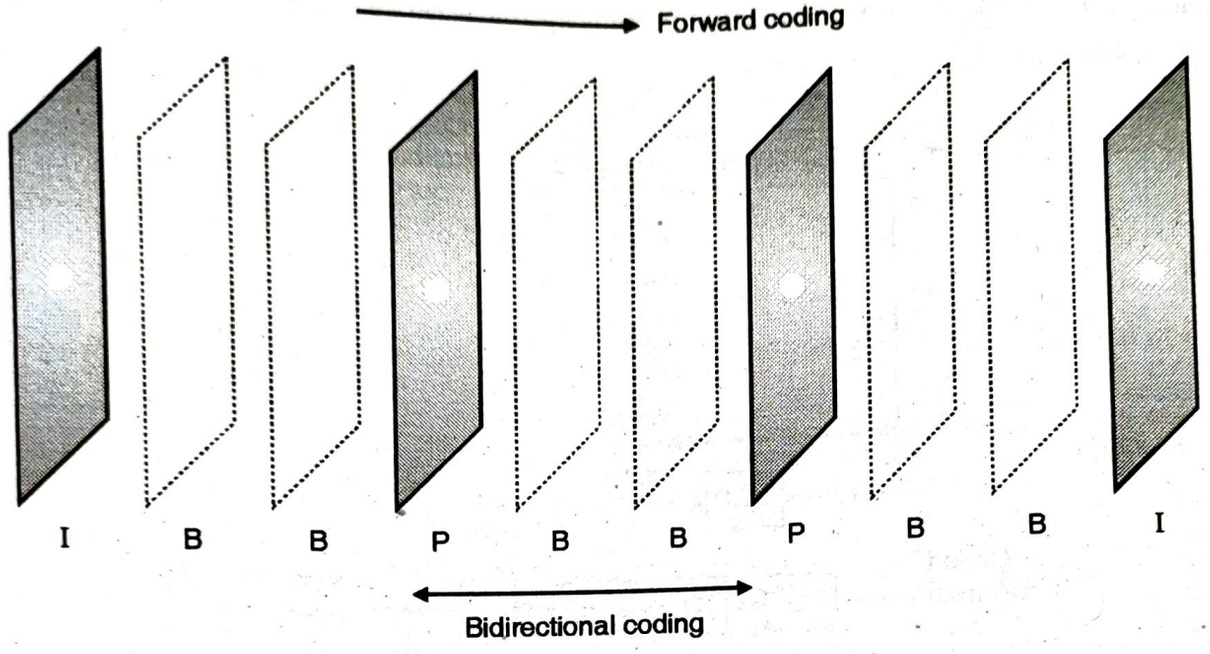


Fig. 4.9.3 : Illustrated use of three picture types in a typical GOP

The different picture types typically occur in a repeating sequence, termed a 'Group of Pictures' or GOP. A typical GOP in display order is :

B_1 B_2 I_3 B_4 B_5 P_6 B_7 B_8 P_9 B_{10} B_{11} P_{12}

The corresponding bitstream order is:

I_3 B_1 B_2 P_6 B_4 B_5 P_9 B_7 B_8 P_{12} B_{10} B_{11}

A regular GOP structure can be described with two parameters : N, which is the number of pictures in the GOP, and M, which is the spacing of P-pictures. The GOP given here is described as N = 12 and M = 3. MPEG-2 does not insist on a regular GOP structure. For example, a P-picture following a shot-change may be badly predicted since the reference for prediction is completely different from the picture being predicated. Thus, it may be beneficial to code it as an I-picture instead.

For a given decoded picture quality, coding using each picture type produces a different number of bits. In a typical example sequence, a coded I-picture was three times larger than a coded P-picture, which was itself 50% larger than a coded B-picture.

4.10 MPEG-4

- MPEG-4 is a collection of methods defining compression of audio and visual (AV) digital data.
- MPEG-4 is a standard used to compress audio and visual data. The MPEG-4 standard is generally used for streaming media and CD distribution, video conversation, and broadcast television. MPEG-4 incorporates many features of MPEG-1, MPEG-2 and other related standards.
- MPEG-4 is still a developing standard and is divided into several parts. The standard includes the concept of "profiles" and "levels," allowing a specific set of capabilities to be defined in a manner appropriate for a subset of applications.
- MPEG-4 is able to crunch massive video files into pieces small enough to send over mobile networks. While these blurry pictures are unlikely to persuade millions of people to upgrade immediately their mobile phones but holds enough promise for future.
- Perhaps more important are the interactive features that MPEG-4 offers. The video functions almost like a Web page, but allowing people to interact with the picture on the screen or to manipulate individual elements in real time.
- MPEG-4 also allows other types of content to be bundled into a file, such as video or images. These files require special software to play.
- MPEG-4 would allow the interactivity of the video which may open potential to do far more than just point and click at links on the screen. Individual elements of the video like a character, a ball in a sporting event, a rocket ship in a science-fiction epic can exist in a separate layer from the rest of the video. This could allow viewers to interact with these elements somehow, even changing the direction of the story.
- The MPEG-4 Video VM supports the representation of video objects (VOs) of natural or synthetic origin, coding them as separate entities in the bitstream which the user can access and manipulate (cut, paste, scale, etc.).
- In the MPEG-4 context, a VO can still be the traditional case of a sequence of rectangular frames formed by pixels. But a VO can also correspond to a sequence of arbitrarily shaped sets of pixels possibly with a semantic meaning, given that this higher level information is somehow made available (e.g. by providing shape or transparency information).
- The way VOs are identified is not within the scope of the MPEG-4 standard - it is considered as a pre-processing step. MPEG-4 wants to provide the means to represent any composition of objects, whatever the methods used to achieve the composition information. The arbitrarily shaped VOs can be obtained by a variety of means such as : automatic, or assisted segmentation of natural data, chroma key techniques, or synthetic computer generated data. The video test material currently used in MPEG-4 contains both rectangular and pre-segmented VOs. Since MPEG-4 usefulness will strongly depend on the availability of robust tools for video analysis and content production, it is expected that significant developments will happen

in the near future in this area, notably by taking into account techniques already in use in the area of Computer Vision

- For natural VOs, the shape or boundary of the object needs to be identified first. The shape information and texture of the object are then encoded separately. The texture information is encoded in the similar way to MPEG-1 video. It should be noted that MPEG-4 does not specify how to determine object boundaries. It is up to the MPEG-4 product developers to develop techniques for object segmentation.
- With object-based coding, efficient compression and content-based scalability are possible. MPEG-4 is an important development in audiovisual coding - from pixel based coding to object-based coding. The object based coding makes content-based indexing and retrieval of multimedia data achievable.

4.11 MPEG-7

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2, and the MPEG-4 standard. MPEG-1 and MPEG-2 standards made interactive video on CD-ROM and Digital Television possible. MPEG-4 is the multimedia standard for the fixed and mobile web enabling integration of multiple paradigms.

MPEG-7, formally named "Multimedia Content Description Interface", is a standard for describing the multimedia content data that supports some degree of interpretation of the information meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as broad a range of applications as possible.

MPEG-7 offers a comprehensive set of audiovisual Description Tools (the metadata elements and their structure and relationships, that are defined by the standard in the form of Descriptors and Description Schemes) to create descriptions (i.e., a set of instantiated Description Schemes and their corresponding Descriptors at the users will), which will form the basis for applications enabling the needed effective and efficient access (search, filtering and browsing) to multimedia content. This is a challenging task given the broad spectrum of requirements and targeted multimedia applications, and the broad number of audiovisual features of importance in such context.

MPEG-7 will also standardize ways to define other descriptors as well as structures (description schemes) for the descriptors and their relationships. This description (i.e., the combination of descriptors and description schemes) will be associated with the content itself, to allow fast and efficient searching for material of a user's interest. AV material that has MPEG-7 data associated with it can be indexed and searched. This 'material' may include still pictures, graphics, 3D models, audio, speech, video, and information about how these elements are combined in a multimedia presentation ("scenarios" composition information). Special cases of these general data types may include facial expressions and personal characteristics.

Because the descriptive features must be meaningful in the context of the application, they will be different for erent user domains and different applications. This implies that the same material can be described using

different types of features, tuned to the area of application. All these descriptions will be coded in an efficient way - efficient for search, that is.

MPEG-7 data may be physically located with the associated AV material, in the data stream or on the same storage system, but the descriptions could also live somewhere else on the globe. When the content and its descriptions are not colocated, mechanisms that link AV material and their MPEG-7 descriptions are useful; these links should work in both directions.

4.11.1 Main elements of the MPEG-7

- **Description Tools :** Descriptors (D), that define the syntax and the semantics of each feature (metadata element); and Description Schemes (DS), that specify the structure and semantics of the relationships between their components, that may be both Descriptors and Description Schemes;
- **A Description Definition Language (DDL)** to define the syntax of the MPEG-7 Description Tools and to allow the creation of new Description Schemes and, possibly, Descriptors and to allow the extension and modification of existing Description Schemes;
- **System tools,** to support binary coded representation for efficient storage and transmission, transmission mechanisms (both for textual and binary formats), multiplexing of descriptions, synchronization of descriptions with content, management and protection of intellectual property in MPEG-7 descriptions, etc.

Therefore, MPEG-7 Description Tools allows to create descriptions (i.e., a set of instantiated Description Schemes and their corresponding Descriptors at the users will), to incorporate application specific extensions using the DDL and to deploy the descriptions using System tools.

4.11.2 MPEG-7 parts

☞ **The MPEG-7 Standard consists of the following parts**

- MPEG-7 Systems : The tools needed to prepare MPEG-7 descriptions for efficient transport and storage and the terminal architecture.
- MPEG-7 Description Definition Language : The language for defining the syntax of the MPEG-7 Description Tools and for defining new Description Schemes.
- MPEG-7 Visual : The Description Tools dealing with (only) Visual descriptions.
- MPEG-7 Audio : The Description Tools dealing with (only) Audio descriptions.
- MPEG-7 Multimedia Description Schemes : The Description Tools dealing with generic features and multimedia descriptions.
- MPEG-7 Reference Software : A software implementation of relevant parts of the MPEG-7 Standard with normative status.

- MPEG-7 Conformance Testing : Guidelines and procedures for testing conformance of MPEG-7 implementations.
- MPEG-7 Extraction and use of descriptions : Informative material about the extraction and use of some of the Description Tools.
- MPEG-7 Profiles and levels : Provides guidelines and standard profiles.
- MPEG-7 Schema Definition : Specifies the schema using the Description Definition Language.

4.11.3 MPEG-7 Application Areas

The elements that MPEG-7 standardizes provide support to a broad range of applications (for example, multimedia digital libraries, broadcast media selection, multimedia editing, home entertainment devices, etc.). MPEG-7 will also make the web as searchable for multimedia content as it is searchable for text today. This would apply especially to large content archives, which are being made accessible to the public, as well as to multimedia catalogues enabling people to identify content for purchase. The information used for content retrieval may also be used by agents, for the selection and filtering of broadcasted "push" material or for personalized advertising. Additionally, MPEG-7 descriptions will allow fast and cost-effective usage of the underlying data, by enabling semi-automatic multimedia presentation and editing.

All application domains making use of multimedia will benefit from MPEG-7. Considering that at present day it is hard to find one not using multimedia, please extend the list of the examples below using your imagination :

- Architecture, real estate, and interior design (e.g., searching for ideas).
- Broadcast media selection (e.g., radio channel, TV channel).
- Cultural services (history museums, art galleries, etc.).
- Digital libraries (e.g., image catalogue, musical dictionary, bio-medical imaging catalogues, film, video and radio archives).
- E-Commerce (e.g., personalized advertising, on-line catalogues, directories of e-shops).
- Education (e.g., repositories of multimedia courses, multimedia search for support material).
- Home Entertainment (e.g., systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke).
- Investigation services (e.g., human characteristics recognition, forensics).
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face).
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems).
- Multimedia editing (e.g., personalized electronic news service, media authoring).

- Remote sensing (e.g., cartography, ecology, natural resources management).
- Shopping (e.g., searching for clothes that you like).
- Social (e.g. dating services).
- Surveillance (e.g., traffic control, surface transportation, non-destructive testing in hostile environments).

4.12 DVI Technology

Intel is the current owner of DVI, which was one of the first systems that provided practical full-motion video incorporating real-time decompression technology.

DVI technology has defined a file format for storing audio/video objects. Applications should use this and other industry standard file formats, to increase interoperability with other applications such as media editing and manipulation tools. The DVI multimedia file format is particularly appropriate for motion video objects that use the compression algorithms, and media objects that use ActionMedia II board pixel formats.

The DVI multimedia file format was designed to grow into a general purpose repository for complex multimedia objects, including information that might be added by media object editors.

DVI is actually both the name of the Digital Video Interactive hardware system sold by Intel and the file format associated with that system. DVI technology is essentially a PC-based interactive audio/video system used for multimedia applications. The DVI system consists of a board for use in an Intel-based PC, drivers, and associated software. The four components of DVI technology are :

1. DVI hardware chipset
2. Run-time software interface
3. Data compression and decompression schemes
4. Data file formats

The heart of the DVI system is the hardware architecture based on the video display processor (VDP) chipset. DVI technology was originally designed for implementation on the IBM PC AT platform. A collection of three 16-bit, ISA-bus DVI interface boards (audio, video, and CD-ROM) were plugged into the AT, and all of the hardware capabilities were accessed through the run-time software interface. The functions in the interface were called by writing a software program using a programming language such as assembly or C.

Today, Intel distributes licenses to third-party developers who want to incorporate DVI technology into their platforms and multimedia products. All of IBM's multimedia hardware platforms (such as the Action Media II boards) and software applications are based upon DVI technology.

DVI is a major competitor of QuickTime, AVI, and MPEG for market share in digital audio/video applications.

DVI allows the storage and playback of audio and video information. All DVI images have a 5:4 pixel aspect ratio and are 256 × 240 pixels in size. DVI is also capable of storing still images and supports both a lossy and a lossless native compression method for such images. DVI works across MS-DOS, Microsoft Windows, and OS/2 platforms and supports the capability of using its own proprietary compression scheme, or using user-definable algorithms, such as JPEG, as well. Audio compression is achieved using either the ADPCM or PCM8 algorithms.

☞ File Organization

The DVI file format is extremely flexible in its design and is used to store a wide variety of data. This format is capable of storing both still-image and motion-video/audio data. As you can see, a common practice of DVI is to store each color plane of an image in a separate disk file. This allows the easy reading and writing of bitmap information, without the need to buffer data to read or write a single file.

A still image is saved using three color-channel files and possibly a colormap and alpha-channel file as well. Motion-video/audio data is stored using the Audio/Video Support System (AVSS) file format. AVSS (pronounced "avis") allows audio and video data to be stored in the same file and played back in a synchronized manner. All AVSS files have the extension .AVS or the file type AVSS.

The data in AVSS files is primarily stream-based, and there is at least one data stream per AVSS file. Each file contains a standard header, an AVL file header, one stream header per data stream, one substream header per substream, frame data, and a frame directory.

Review Questions

Q. 1 Explain different types of video.

Q. 2 Explain Chroma subsampling.

Q. 3 Explain analog video.

Q. 4 Explain different video format.

Q. 5 Explain video format.

Q. 6 Explain Block Matching Algorithm.

Q. 7 Explain h.261 encoder.

Q. 8 How does an H.264 codec work ?

Q. 9 Explain MPEG-1 Compression Algorithm.

Q. 10 Explain MPEG 7.

Q. 11 Explain DVI file format.

Chapter Ends...

❑❑❑