01076566 Multimedia Systems

Chapter 8: Media Compression: Video

Pakorn Watanachaturaporn
pakorn.wa@KMITL.ac.th

Bachelor Program in Computer Engineering (B.Eng.)
Faculty of Engineering

King Mongkut's Institute of Technology Ladkrabang

Outline

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- General Theory of Video Compression
- Types of Predictions
- Complexity of Motion Compensation
- Video-Coding Standards

General Theory of Video Compression

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

Multimedia video data	NTSC video	QCIF video	CIF video	HDTV (progressive) video	HDTV (interlaced) video
Frame size	720 × 486	176 × 144	352 × 288	1280 × 720	1920 × 1080
Bits/pixel	16 bpp	12 bpp	12 bpp	12 bpp	12 bpp
Frame rate	29.97	30	30	59.94	29.97
Uncompressed frame size (bytes)	700 KB	38 KB	152 KB	1.38 MB	3.11 MB
Uncompressed data produced per second (bits per second)	167.79 Mbps	9.12 Mbps	36.5 Mbps	662.89 Mbps	745.75 Mbps

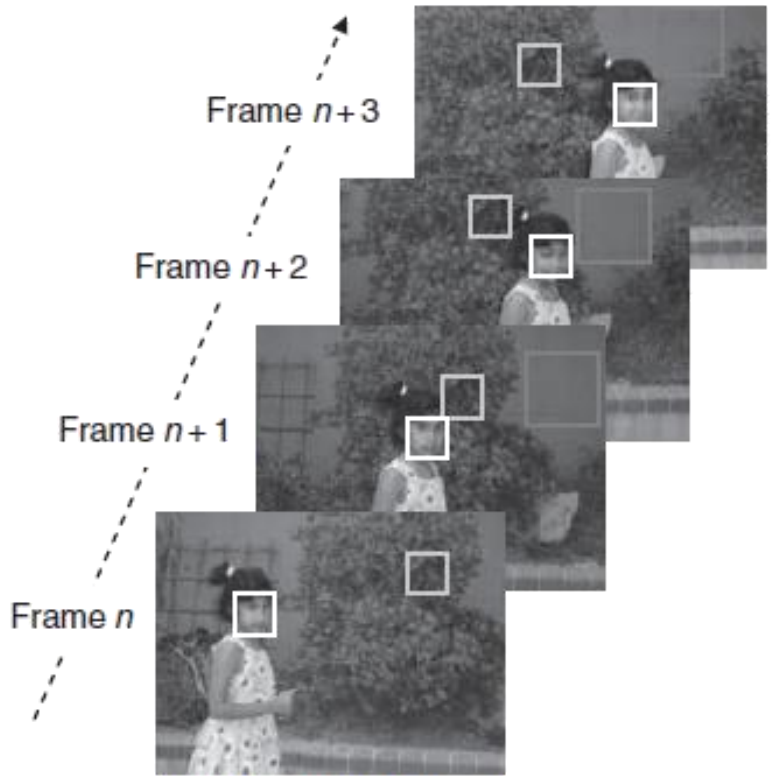


Figure 8-2 Temporal redundancy in video. Four frames of a video sequence are shown. The girl in the foreground is moving, whereas the background objects are static; however, the camera is also moving. Although each frame is different, there are areas of the background and the foreground that remain the same or are very similar.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

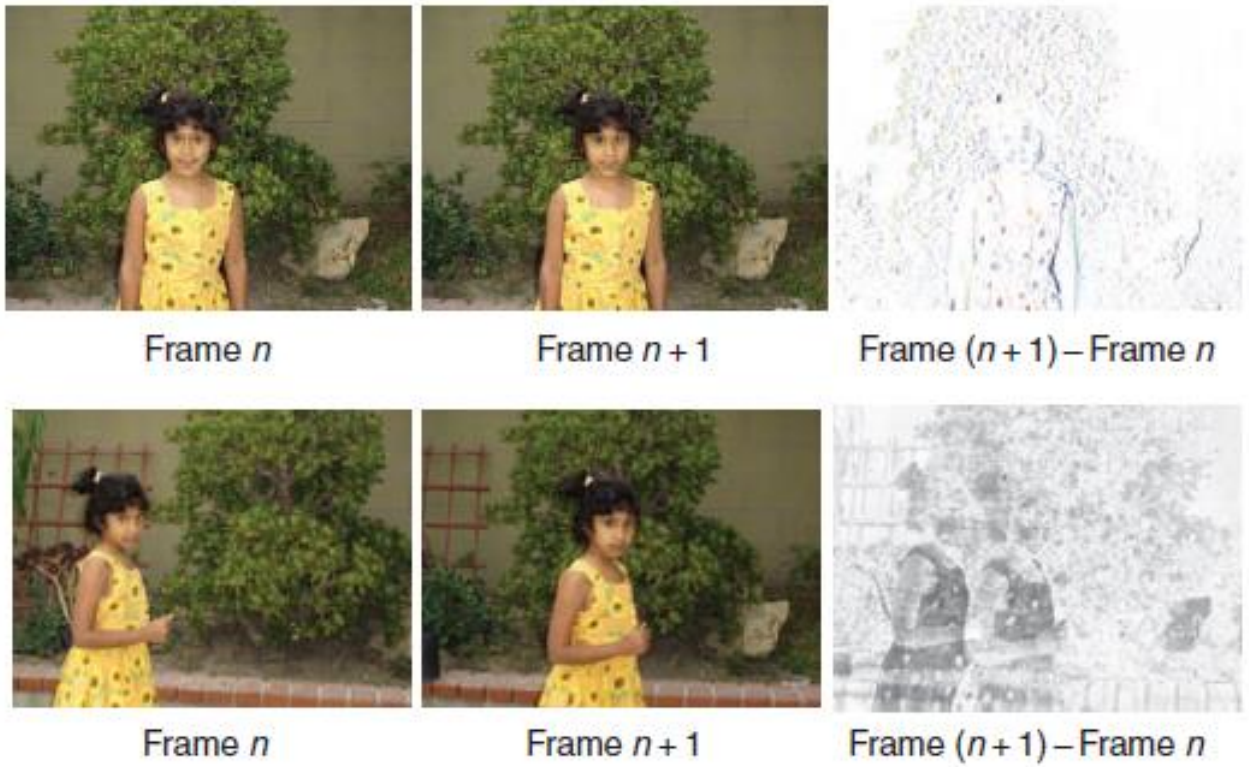


Figure 8-3 The top row shows two successive frames in a low-motion video. The frame difference of the Y channel is shown on the right. The bottom row shows two successive frames where the object/camera motion is higher. The difference image in this case contains a lot more information than the previous case. See the color insert in this textbook for a full-color version of this image.

Temporal Redundancy

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

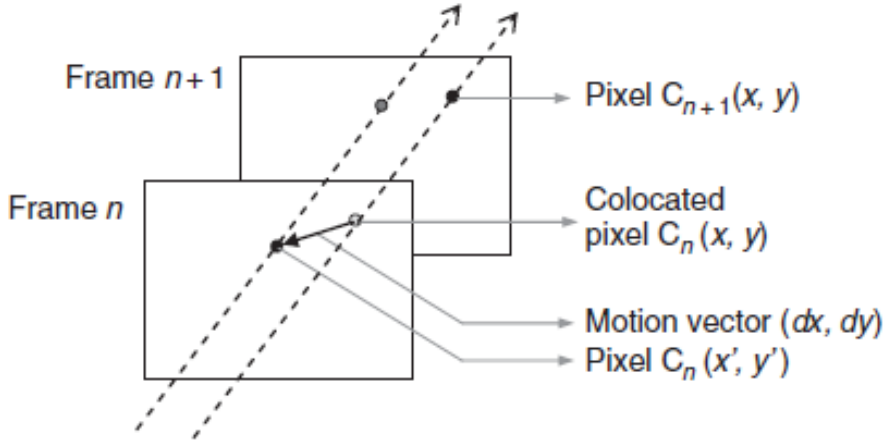


Figure 8-4 Pixel motion prediction. The pixel shown in the frame n has moved to a new location in frame n + 1. Consequently, $C_{n+1}(x, y)$ in frame n + 1 is not the same as $C_n(x, y)$ but is offset by the motion vector (dx,dy).

- $C_{n+1}(x, y) = C_n(x, y)$
- $C_{n+1}(x, y) \neq C_n(x, y)$
- $C_{n+1}(x, y) = C_n(x', y') = C_n(x - dx, y - dy)$
- $C_{n+1}(x, y) = C_n(x - dx, y - dy) + e(x, y)$

Block-Based Frame Prediction

KMITL
พระจอมเกล้าลาดกระบัง
korn.wa@kmitl.ac.th

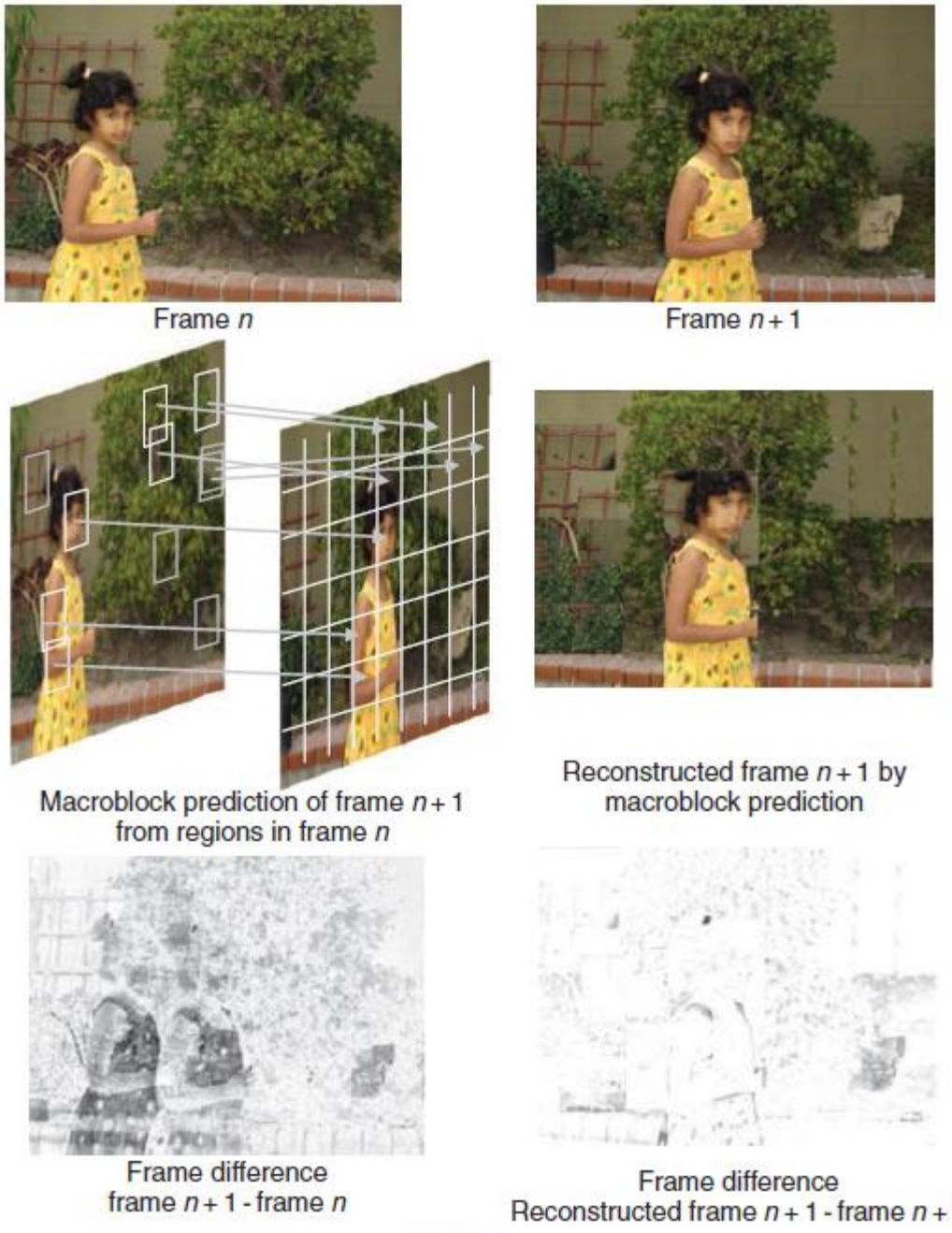
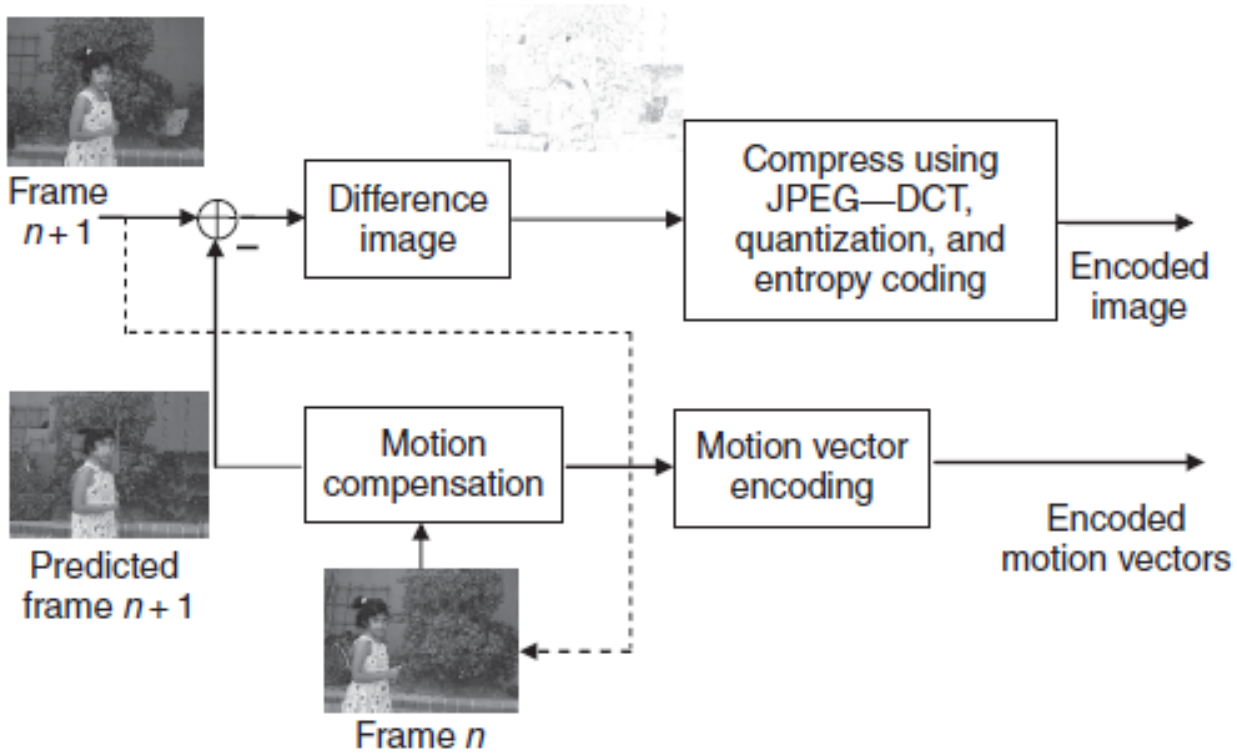


Figure 8-5

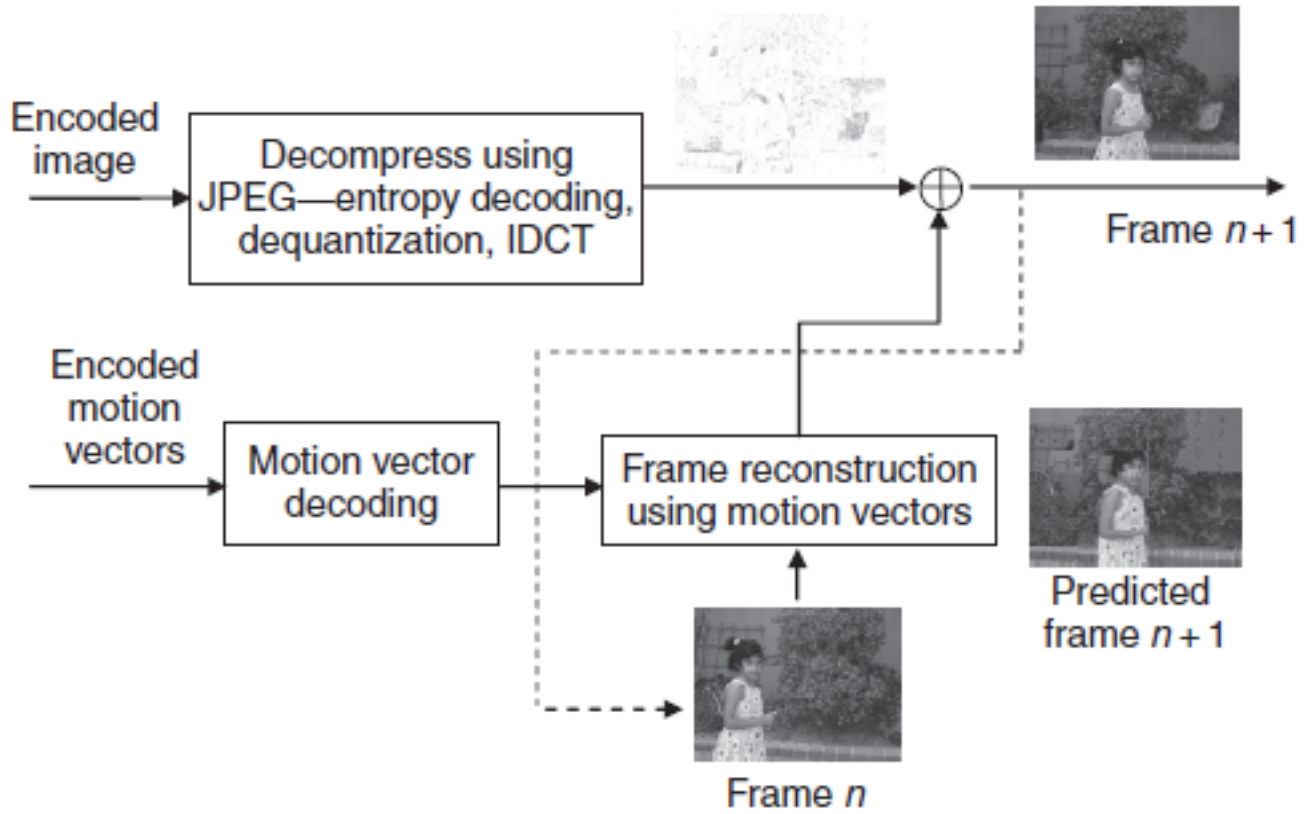
Figure 8 – 5 Macroblock-based frame prediction.

- Frame *n + 1* is divided into macroblocks and each macroblock is predicted from an area in frame *n*.
- The reconstructed frame based on this prediction is formed by copying the frame *n* areas into the appropriate frame *n+1* macroblocks

pakorn.wa@kmitl.ac.th



KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th



Computing Motion Vectors

KMITL
พระจอมเกล้าลาดกระบัง
korn.wa@kmitl.ac.th

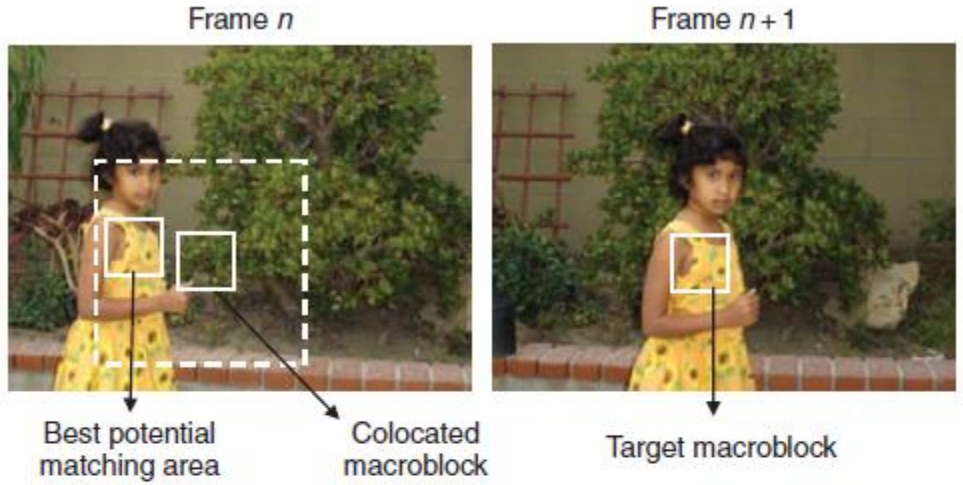
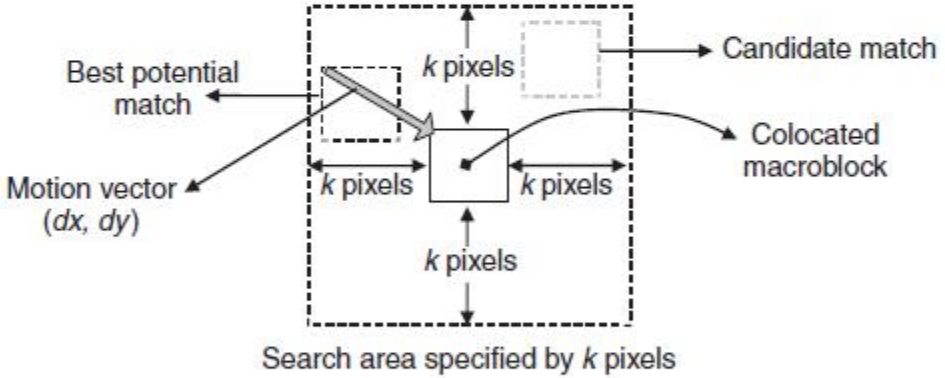
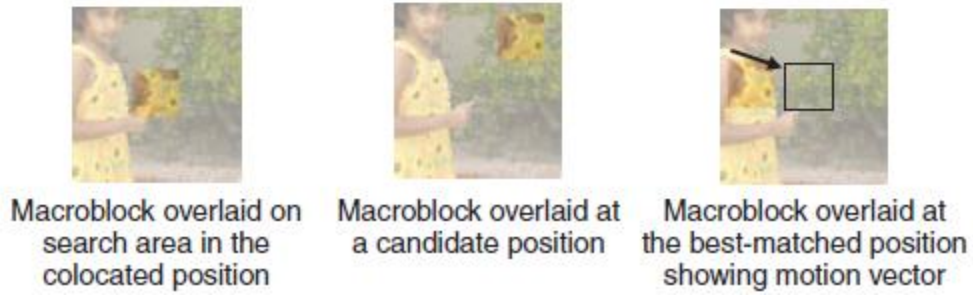


Figure 8 – 7 Motion vector search.

- The encoder needs to search the best match for macroblock in frame *n+1* from frame *n.*
- The search is performed within a search area specified by parameter *k* around the colocated position.
- The best-matching region obtained in the search area by a pixel-by-pixel difference gives the motion vector.





pakorn.wa@kmitl.ac.th

- Compute the difference between the candidate area and the target block by their *mean absolute difference* (MAD)

$$MAD(i,j) = \frac{\sum_{p=1}^{m}\sum_{q=1}^{n}|C_{n+1}[p,q] - C_n[p+i,q+j]|}{mn}$$

- The goal of the search task is to find a motion vector (i, j) such that MAD (i, j) is minimized.

pakorn.wa@kmitl.ac.th

- Other metrics can be used
  - Mean square difference (MSD)

$$\mathrm{MSD}(i,j) = \frac{\sum_{p=1}^{m}\sum_{q=1}^{n}(C_{n+1}[p,q]-C_n[p+i,q+j])^2}{mn}$$

  - Pel difference classification

$$\mathrm{PEL}(i,j) = \sum_{p=1}^{m}\sum_{q=1}^{n}[ord(|C_{n+1}[p,q] - C_n[p+i,q+j]| \le t)]$$

where *t* is some predefined threshold that decides a match
and *ord* (x) = 1 if *x* is true

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Other metrics can be used (Cont'd)
  - Projective ordering

$$\mathrm{PO}(i,j) = \sum_{p=1}^{m} \left| \sum_{q=1}^{n} C_{n+1}[p,q] - \sum_{q=1}^{n} C_n[p+i, q+j] \right| + \sum_{q=1}^{n} \left| \sum_{p=1}^{m} C_{n+1}[p,q] - \sum_{p=1}^{m} C_n[p+i, q+j] \right|$$

Size of Macroblocks

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

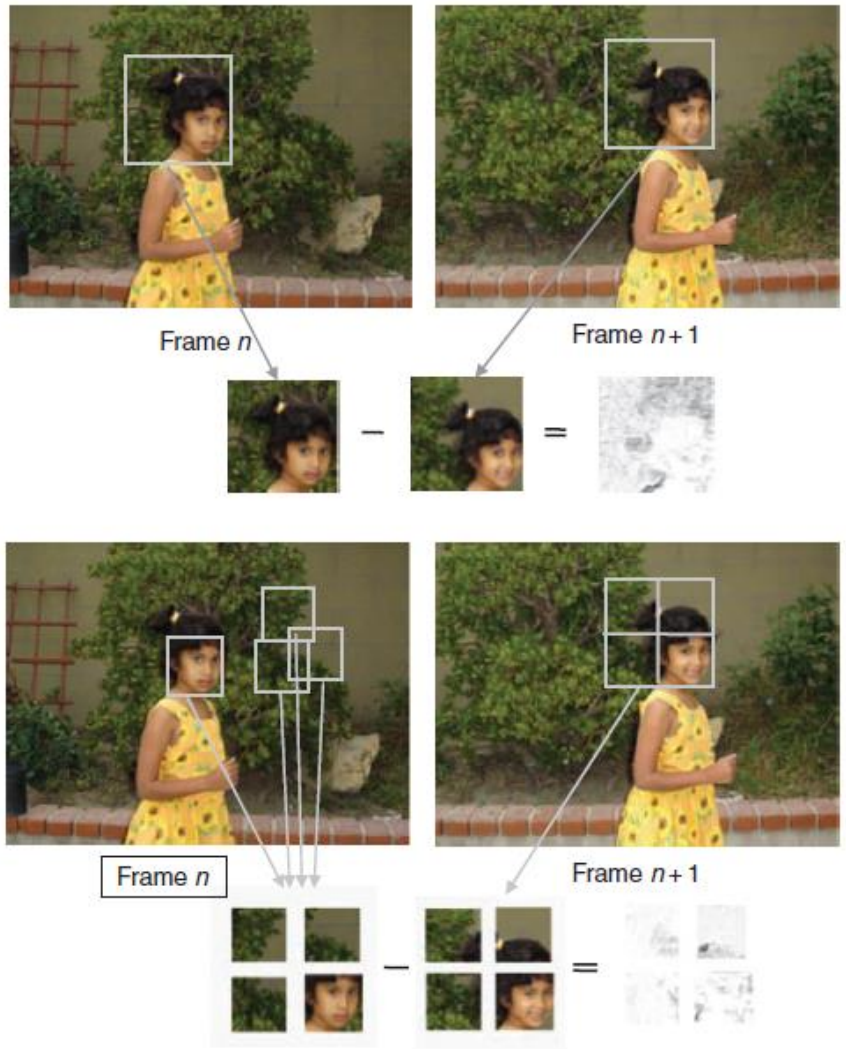


Figure 8-8

Closed Loop Motion Compensation

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

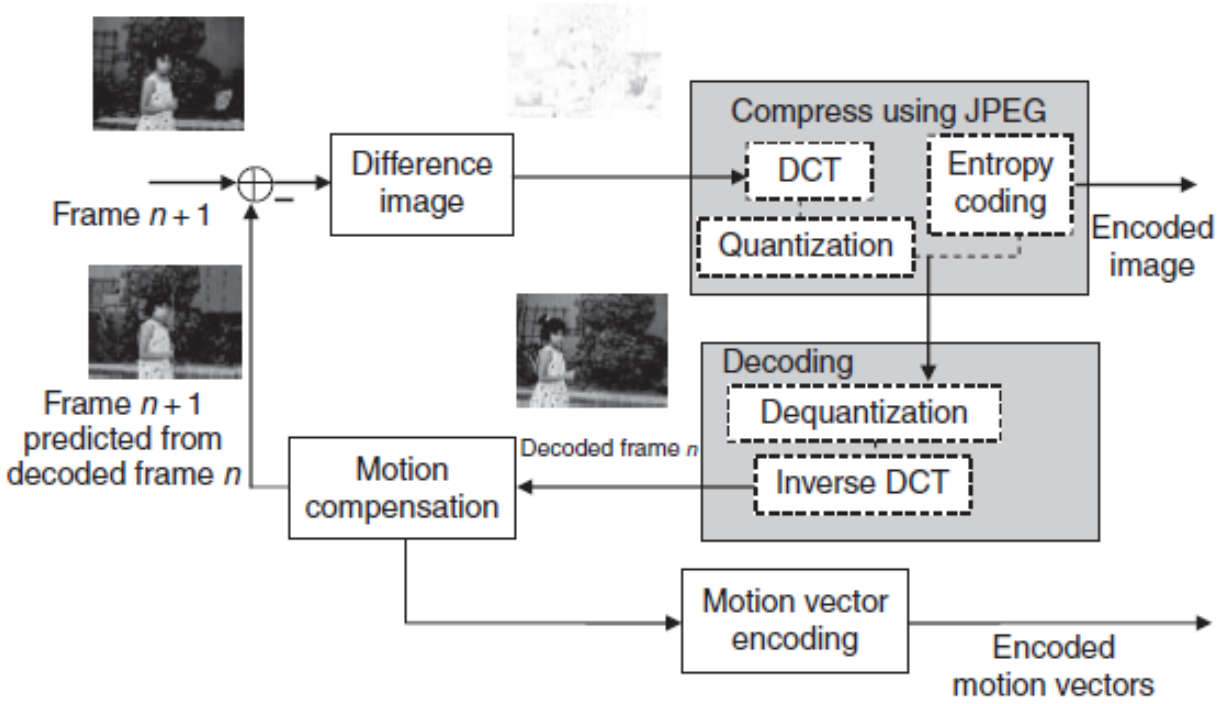


Figure 8-9 Closed loop motion compensation. The encoder simulates a decoder and decodes frame n. The decoded version of frame n is used to predict frame n + 1. The resulting reconstruction at the decoder does not accumulate errors.

Types of Predictions

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- I Frames
  - An I frame is encoded as a single image with no reference to any past of future frames
  - Using standard JPEG pipeline
- P Frames
  - P frames are predictive coded
  - Exploiting temporal redundancy by comparing them with the immediately preceding frame, which is known as a reference frame
  - The preceding reference frame might have been coded as an I frame or even a P frame.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

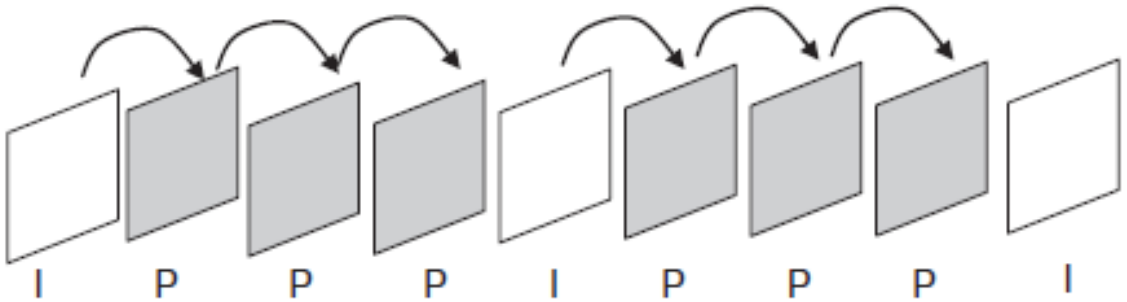


Figure 8-12 Interframe dependency in P frames. P frames are always forward predicted from a previous I frame or a P frame.

pakorn.wa@kmitl.ac.th

- B Frames
  - Bidirectionally coded frames
  - Differ from the P frames by using two reference frames instead of one
  - The previous and future frame are used to predict a B frame
  - Sometimes, area of the current frame can be better predicted by the next future frame
  - Using both frames increases the correctness in prediction during motion compensation.

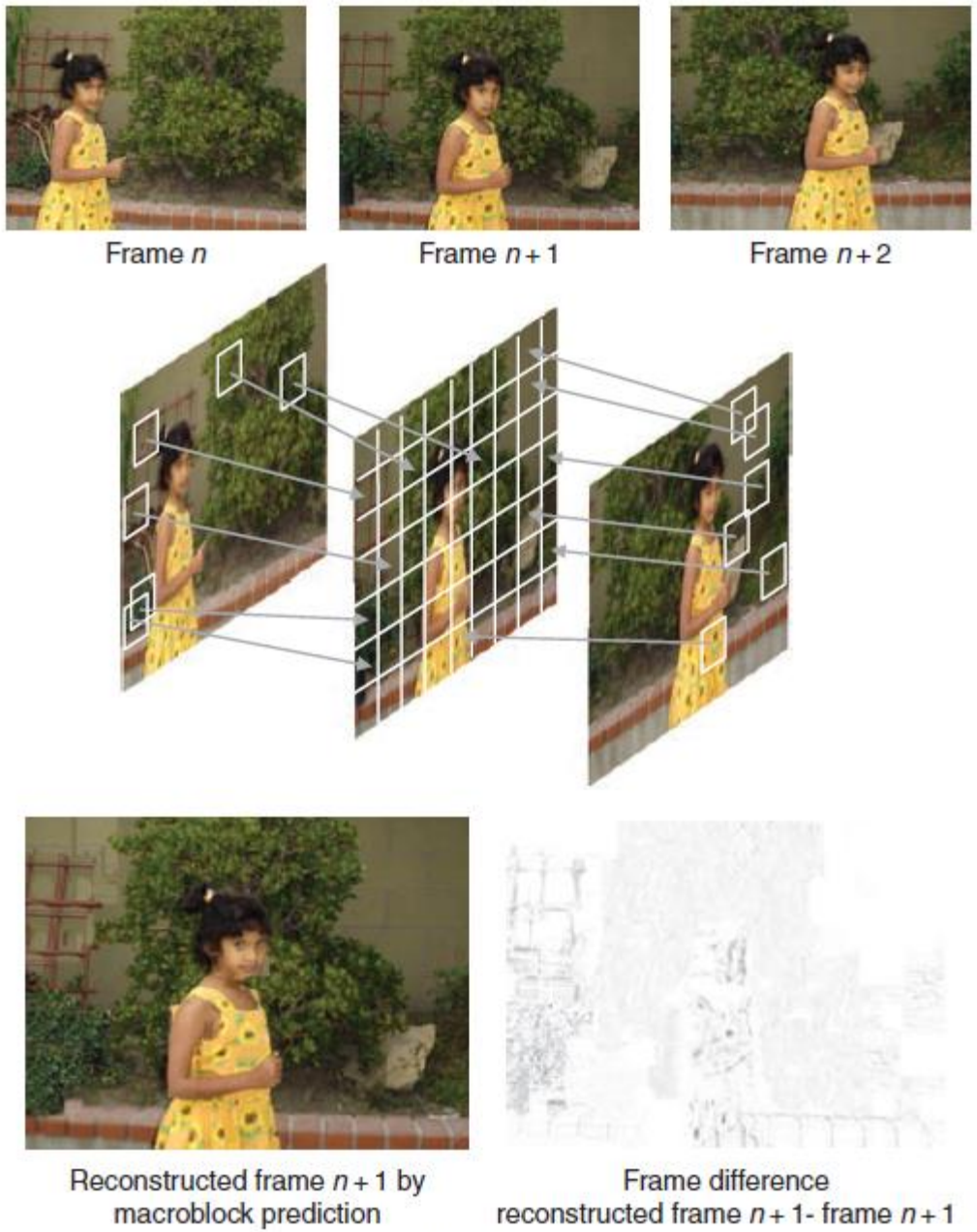

Frame n
Frame n+1
Frame n+2
Reconstructed frame n+1 by macroblock prediction
Frame difference reconstructed frame n+1- frame n+1

Figure 8-13

pakorn.wa@kmitl.ac.th

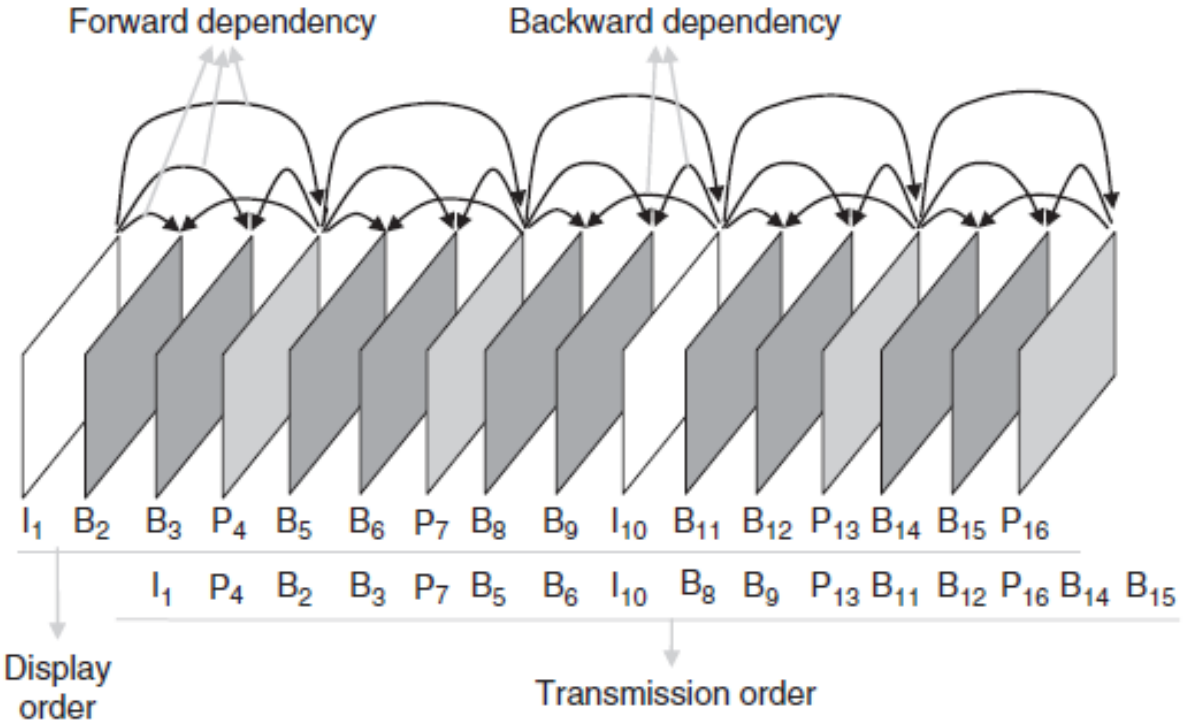


Figure 8-14 B frame predictions. A sequence of frames is shown encoded as I, P, and B frames. B frames have forward and backward dependencies. This imposes a change in the transmission/coding order shown on the bottom line.

Multiframe Prediction

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

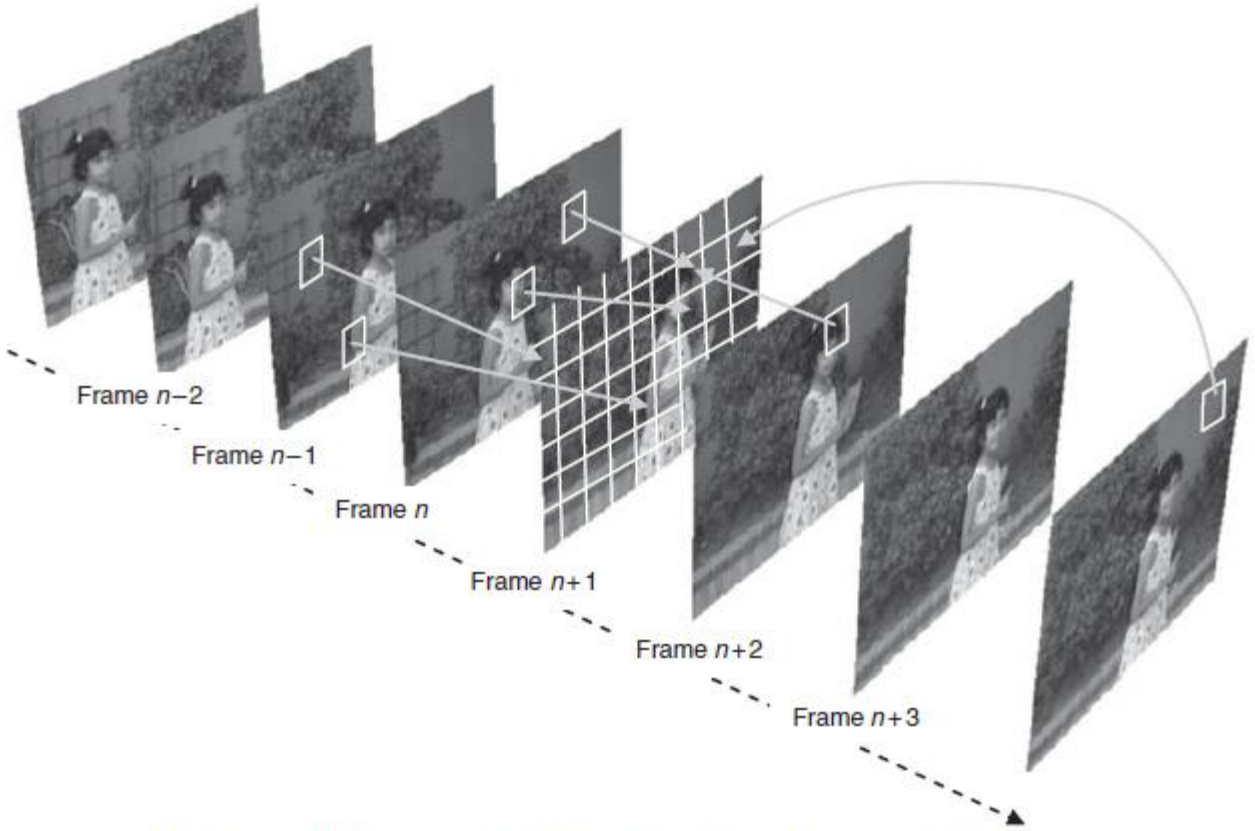


Figure 8-15 Multiframe prediction. A number of reference frames from the past and the future are used to predict frame n + 1. Regions from other multiple frames can do a better job at prediction.

Video Structure – Group of Pictures (GOP)

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

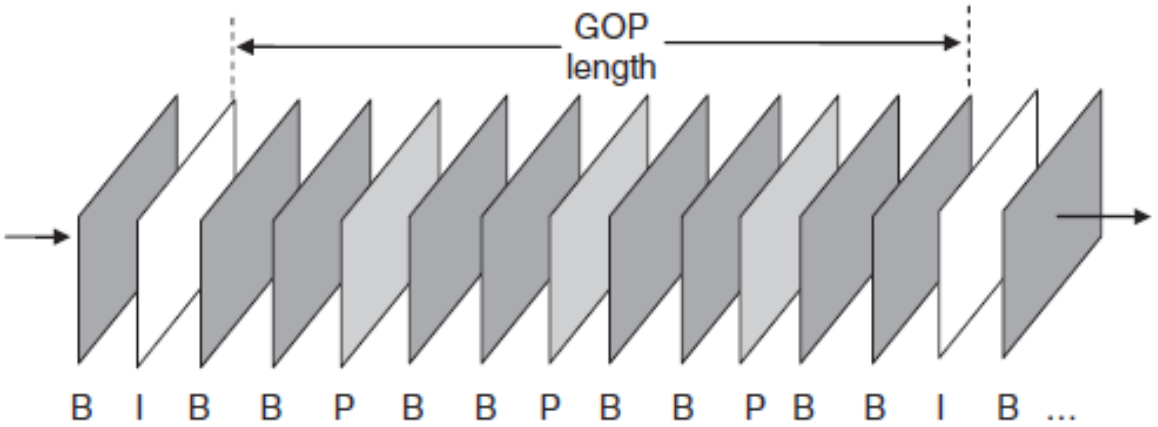


Figure 8-16 Group of Pictures (GOP). An example of a GOP used in MPEG-1 is shown. Here, the GOP length is 12 with three P frames between I frames and two B frames between P frames.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

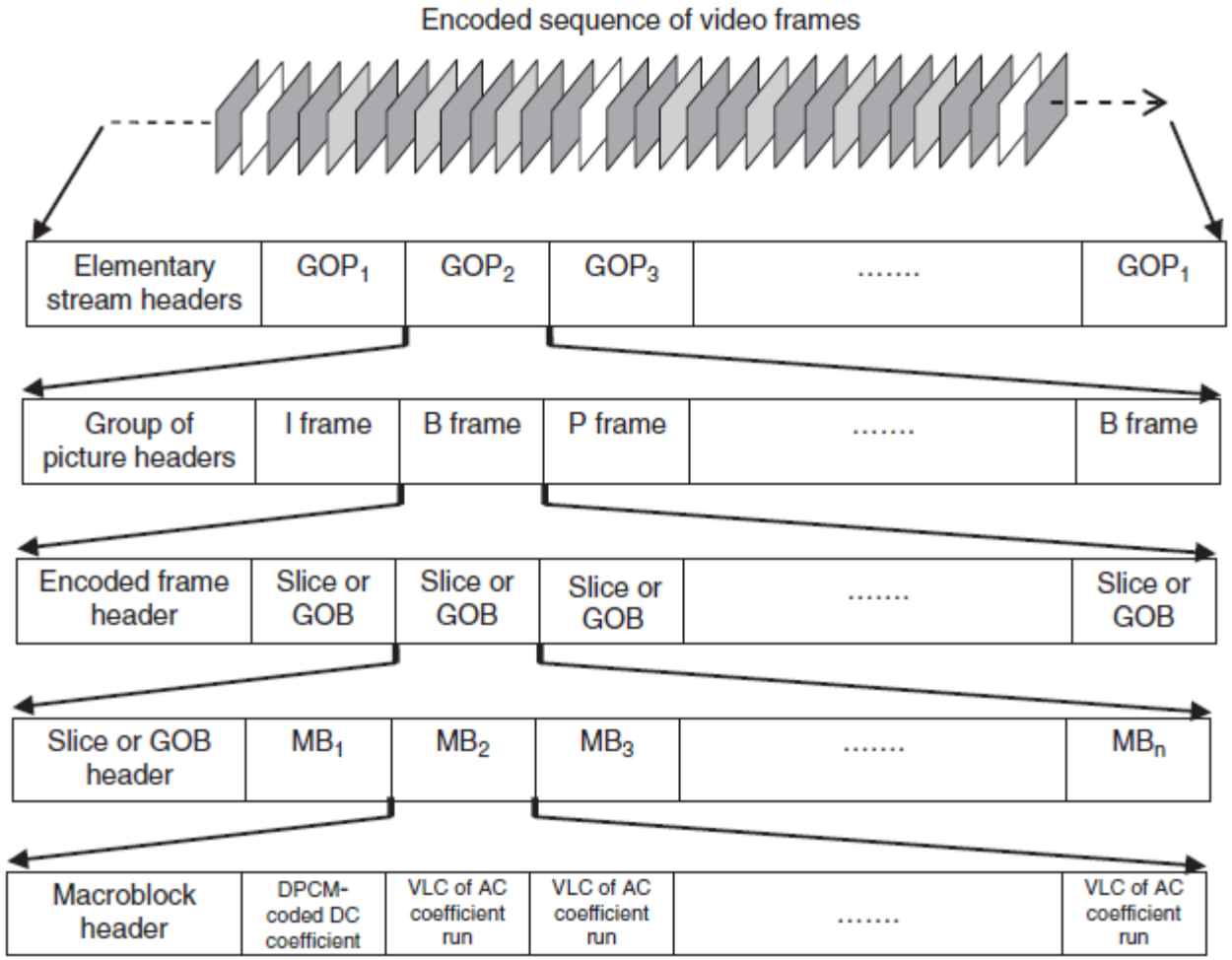


Figure 8-17 Video structure. A typical bit stream structure is hierarchically encoded with GOPs, frames, slices or GOBs, macroblocks, and runs of coefficients.

Complexity of Motion Compensation

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Sequential of Brute Force Search

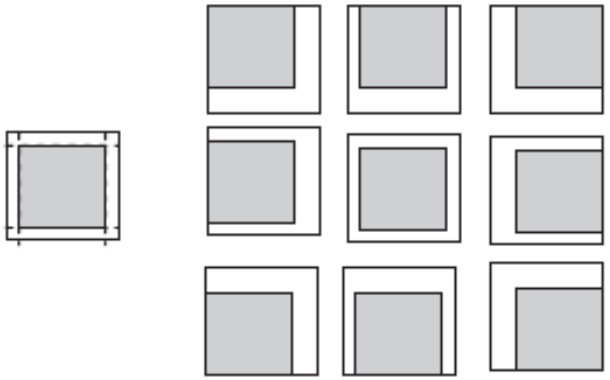


Figure 8-18 Search space example with $k = 1$. There are nine candidate positions computed as $(2k + 1)^2$. In general, the search or the value ranges from 0 to 31 and depends on the motion in the video.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Logarithmic Search

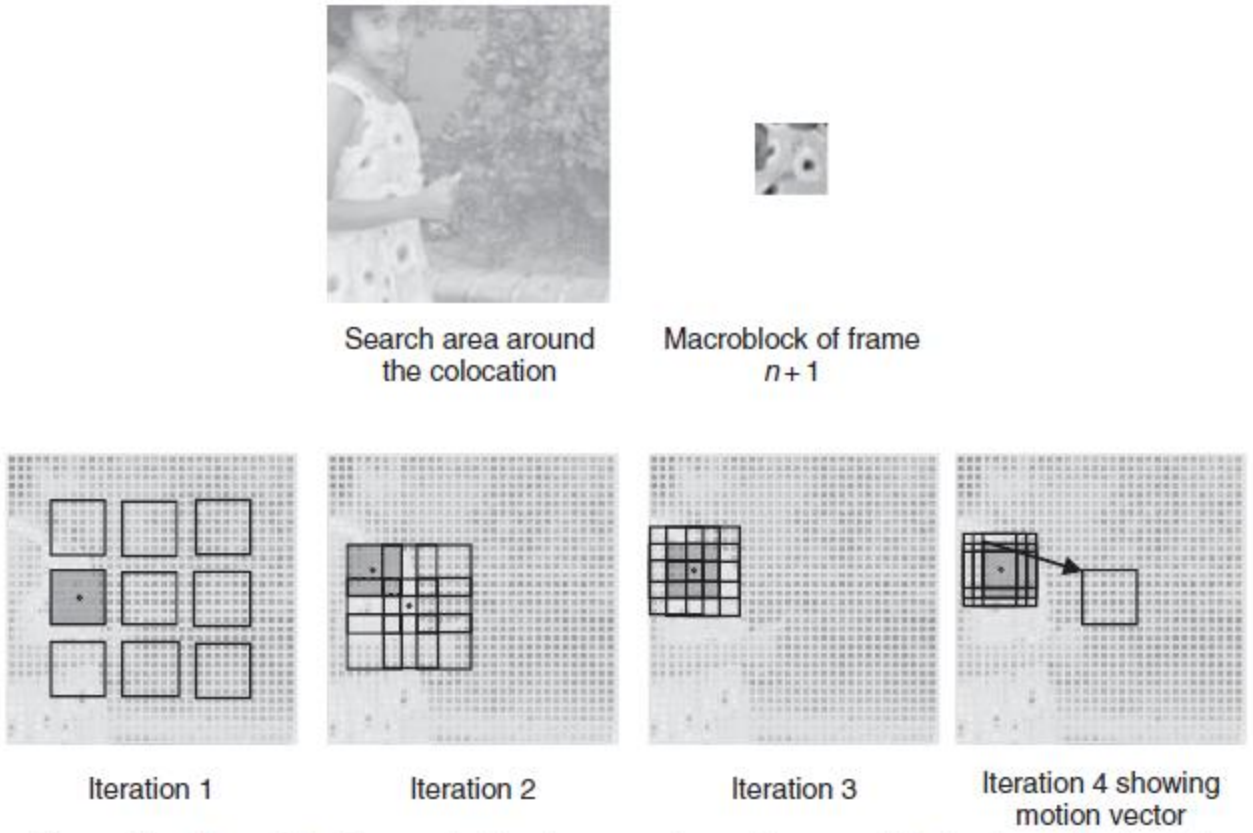


Figure 8-19 Logarithmic search. The top row shows the macroblock whose best match needs to be searched in the search area with k = 16 of frame n. The bottom row shows four iterations of the logarithmic search. At each iteration, the best-matched MAD block is shown shaded. Each iteration converges closer to the best match. See the color insert in this textbook for a full-color version of this image.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- Hierarchical Search

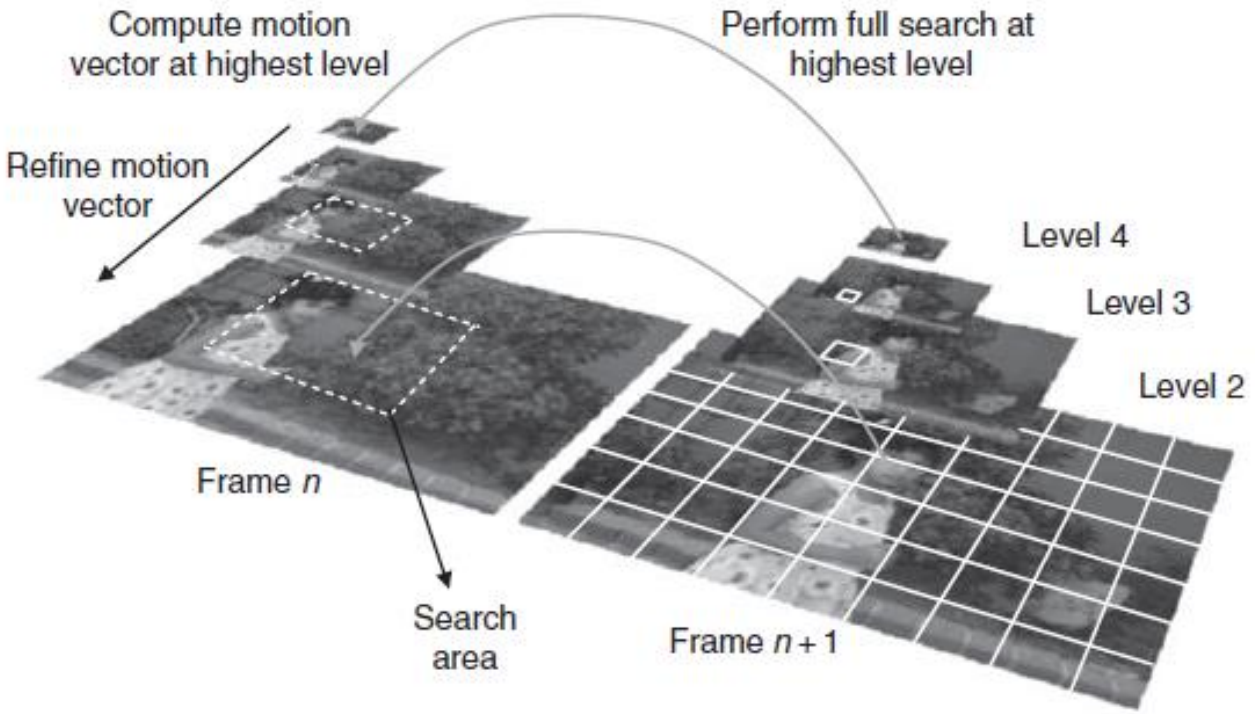


Figure 8-20 Hierarchical motion vector search. A hierarchy of subsampled images is obtained for the target and reference images. A motion vector search for a sample macroblock proceeds by performing a full motion search at the highest level to obtain a motion vector. The motion vector is further refined by using the lower levels in the reference pyramid. See the color insert in this textbook for a full-color version of this image.

Video-Coding Standards

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

- H.261
  - Designed to support video teleconferencing over ISDN lines in 1990
  - Designed for the QCIF resolution (176 × 144) but can also support CIF-sized video (352 × 288)
  - Compression occurs by compressing frames in the intramode (I frames) and in the intermode (P frames) – no support for B frames

- H.263
  - Extend the H.261-based videoconferencing
  - Support a wider range of picture formats, including 4CIF (704 × 576) and 16CIF (1408 × 1152)
  - Use I, P, and B frames.

pakorn.wa@kmitl.ac.th

- MPEG-1
  - Designed to aid storage of noninterlaced digital video at 1.5 Mbps
  - Used in VCD format
  - Designed to encode SIF-sized frame resolution at 1150 Kbps
  - The picture quality overall does compared to VHS video
  - Use intra- and intermodes, where the intermodes use both P and B frames
  - Usage of slices while encoding predictive frames
  - A frame is divided into slices of macroblocks where each slice might contain a variable number of macroblocks.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

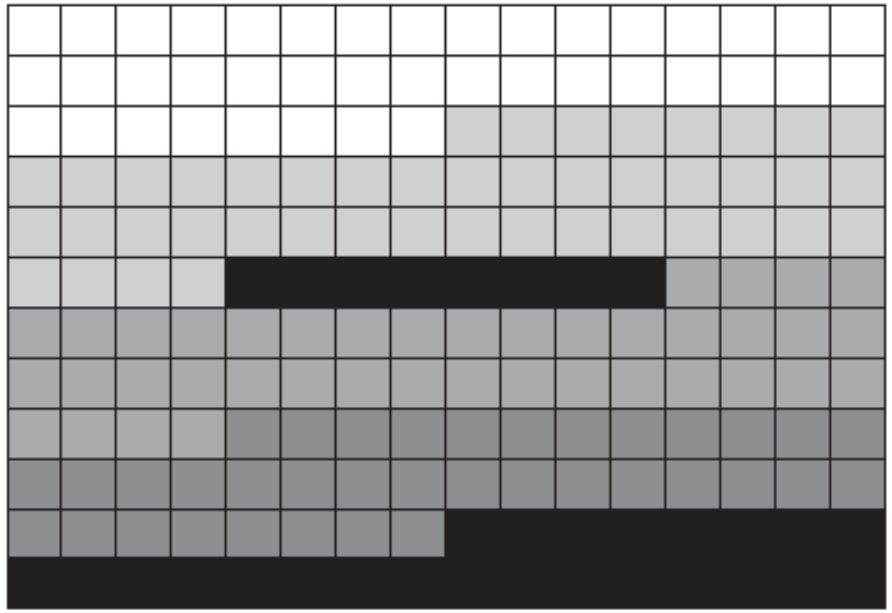


Figure 8-22 Slices in an MPEG-1 frame. Each slice is coded independently and can be quantized differently depending on the entropy of the slice.

pakorn.wa@kmitl.ac.th

- MPEG-1 (Cont'd)
  - Each slice is encoded independently with different quantizer scales
  - Use subpixel (half pixel) motion compensation with motion vectors generated using bilinear interpolation methods.
  - Random access of frames is allowed for application

pakorn.wa@kmitl.ac.th

- MPEG-2
  - Designed for higher bandwidths and higher quality supporting bit rates from 4 to 9 Mbps
  - Using I, P, and B frames similar to MPEG-1 with half-pixel approximation for motion vectors
  - Support for interlaced video
  - Also designed as a transmission standard providing support for a variety of packet formats with error-correction capability across noisy channels
  - Video was scalable encoding

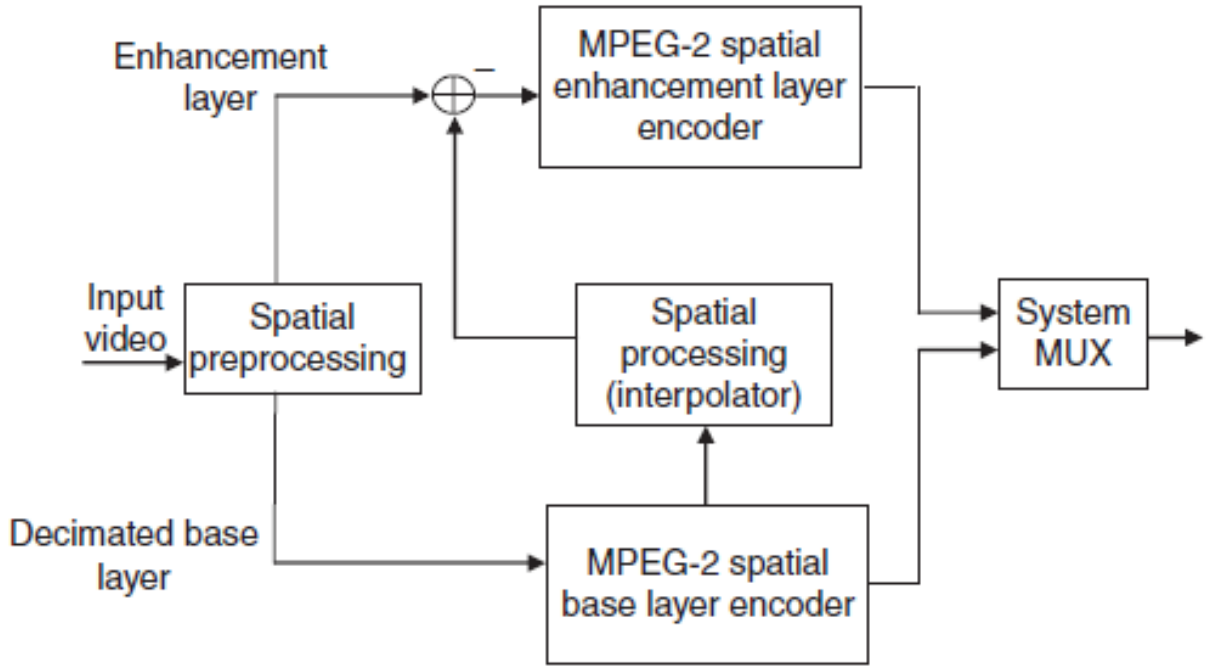


Figure 8-23 Scalable MPEG-2 video encoding. The input video is broken into two layers, a base layer containing low frequency and an enhancement layer containing high frequency. Both layers are coded independently and then multiplexed into the coded bit stream.

pakorn.wa@kmitl.ac.th

- MPEG-4 – VOP and Object Base Coding, SP and ASP
  - Have a broader scope than just encoding video or audio
  - It is a true multimedia format that has video, audio, 2D/3D graphics, and system level parts
  - Support both progressive and interlaced video encoding
  - The video compression obtained by MPEG-4 ASP (Advanced Simple Profile) is better than MPEG-2 by 25% for the same video quality.
  - The standard is object based and supports multiple video streams that can be organized into a hierarchical presentation using scene layout descriptions.
  - A video can consist of different video object planes (VOPs), and the image frames in each VOP can have arbitrary shapes, not necessarily rectangular

- MPEG-4 (Cont'd)
  - Compression provides temporal scalability, which utilizes advances in object recognition. By separating the video into different layers, the foreground object, such as an actor or speaker, can be encoded with lower compression which background objects can be encoded with higher compression
  - Provide a synchronized text tract for courseware development and a synchronized metadata track for indexing and access at the frame level.

- H.264 or MPEG-4-AVC
  - AVC: Advanced Video Codec
  - H.264 provides for smaller block usage up to 4 × 4
  - Allow variable block-sized motion compensation
  - Perform quarter pixel motion compensation (QPEL)
  - Provide for multiple reference picture motion compensation – frame prediction can use any frame(s) from the past or future
  - The entropy coding done on motion vectors as well as the quantized error images uses context-adaptive binary arithmetic coding (CABAC)

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

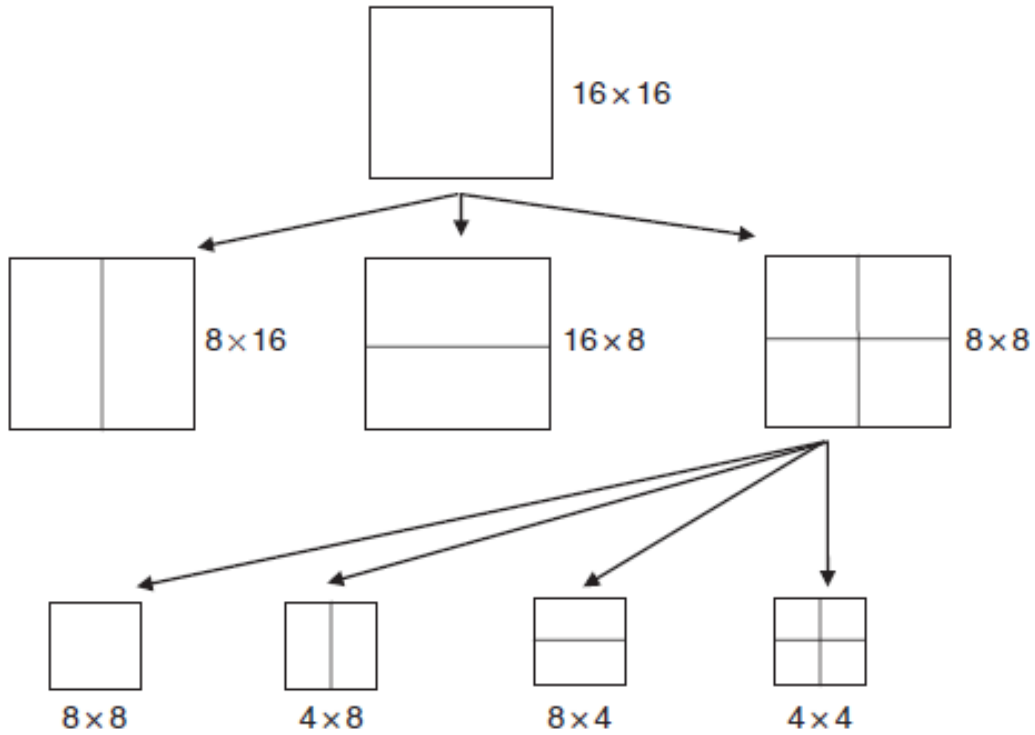


Figure 8-24 Hierarchical breakdown of macroblocks used for motion compensation in H.264. The top macroblock is a 16 × 16 used in MPEG-2. For H.264, this is approximated by 8 × 16, 16 × 8, 8 × 8, 4 × 8, 8 × 4, and 4 × 4 blocks.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

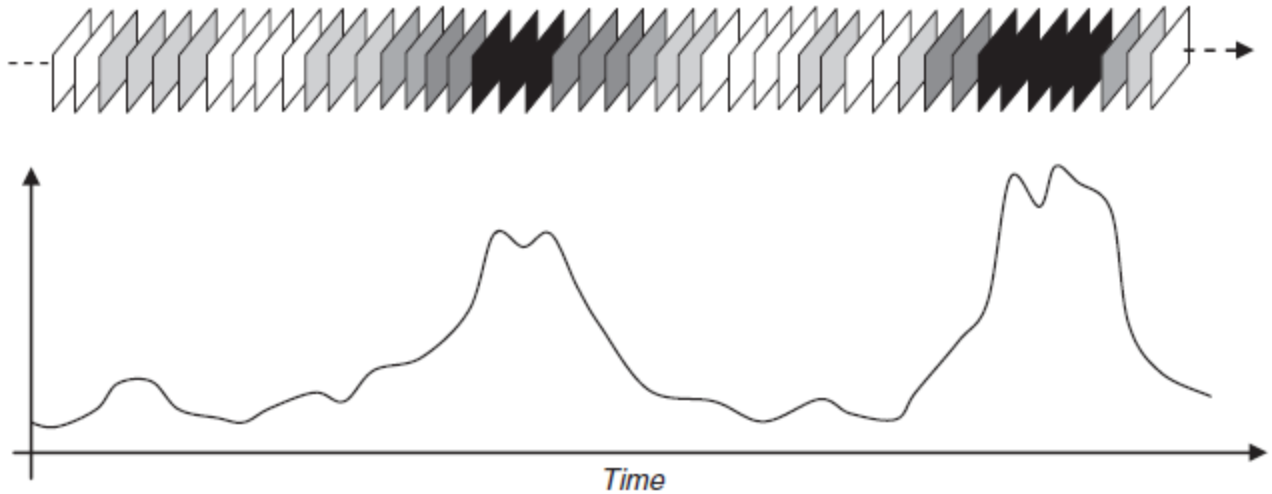


Figure 8-25 Bit budget graph. The top frames are color coded to show motion activity. The darker the frame, the more motion. The lower graph shows the bit requirements over time. In high-motion areas, more bits are required during compression compared with low-motion frames, inherently creating VBR. A CBR output would tend to more regularly distribute bits, making the graph more flat.

KMITL
พระจอมเกล้าลาดกระบัง
pakorn.wa@kmitl.ac.th

Q & A