Lecture-4 notes for SSY150: Multimedia and Video Communications

Video compression

Irene Y.H. Gu

Chalmers University of Technology, Sweden

April 16, 2020

1

Contents

2

1. Fundamental of video compression: how late compression rate is achieved?

High redundancy in consecutive video frames:
only small parts of images change due to the motion


encode and send change parts to the channel
(using motion compensation techniques) 3

2. Video coding modes

Switch between 2 modes:

- **Intra** mode (= 2D still image compression):

  + (DCT) encoding of an image frame is independent of other frames.
  + Applied to: 1st frame + refreshing frames (Intra and I frames)

- **Inter** mode:

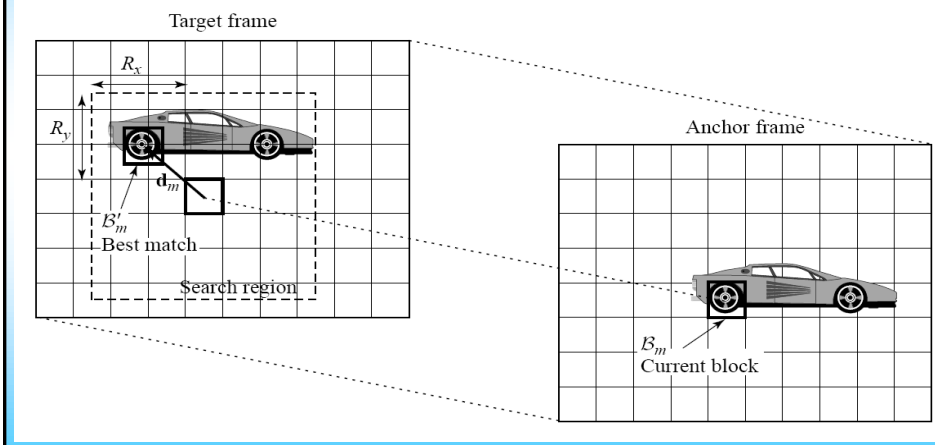
  a simple and effective method

  - block matching for motion compensation (MC)

Apply MC to **P** (predicted) frames and **B** (bidirectional predicted) frames 4

Inter mode: block matching for MC

Find the "best" matching blocks within the searching region



(Courtesy from Dr. Y.Wang)

3. Block matching for MC

The block-matching scheme for motion compensation:

$$I_t(x,y) = I_{t-k}(x-d_x, y-d_y)$$

need to find the 'best' displacement vector (d_x, d_y) for each block

Commonly used objective criteria:

(a) **Mean Square Error (MSE)**

For i-th block B_i, size (M_x, M_y)

$$MSE(d_x,d_y) = \frac{1}{M_x M_y} \sum_{(x,y)\in B_i} \left[I_t(x,y) - I_{t-k}(x-d_x, y-d_y) \right]^2$$

$$(d_x^*, d_y^*) = \arg\min_{(d_x,d_y)} MSE(d_x,d_y)$$

6

(b) **Mean Absolute Error (MAE)**

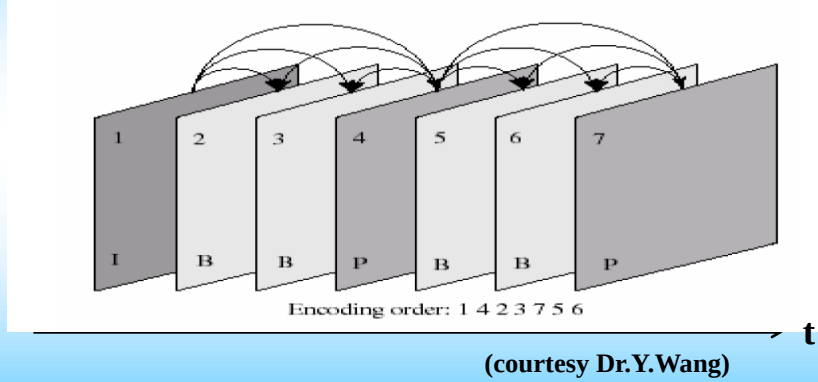
$$MAE(d_x, d_y) = \frac{1}{M_x M_y} \sum_{(x,y)\in B_i} \left| I_t(x,y) - I_{t-k}(x-d_x, y-d_y) \right|$$

$$(d_x^*, d_y^*) = \arg\min_{(d_x, d_y)} MAE(d_x, d_y)$$

7

4. I,P,B image frames, uni-directional, bi-directional MC

I,B,P frames in video coding:



(courtesy Dr.Y.Wang)

MC can be performed by forward and backward prediction using one or more frames

8

Uni-directional and bidirectional MC:

Unidirectional MC

MC is done by using weighted sum of one directional predicted blocks from one/several previous frames:

$$\boldsymbol{I}_t(x,y) = \sum_{K>0} w_k \ \boldsymbol{I}_{t-k}(x - d_{x_k}, y - d_{y_k})$$

Bi-directional MC

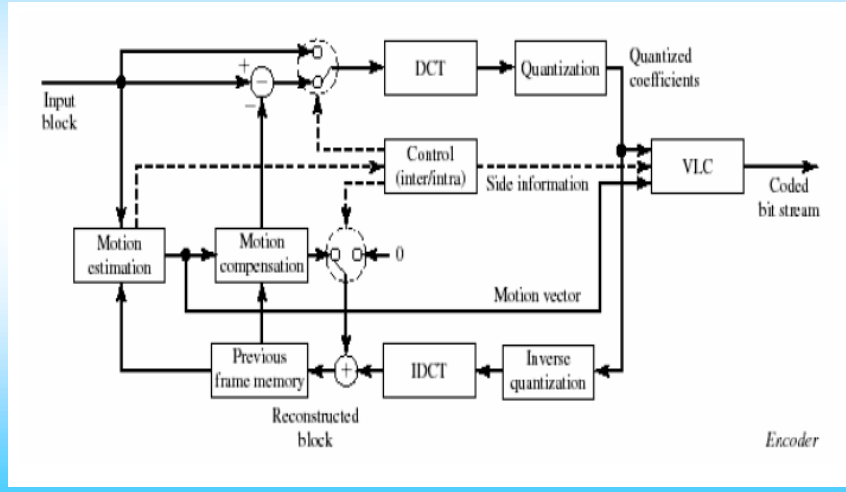
MC is done by weighted sum of forward/backward predicted blocks from one/several previous frames and one/several future frames:

$$\boldsymbol{I}_t(x,y) = \sum_{k} w_k \boldsymbol{I}_{t-k}(x - d_{x_k}, y - d_{y_k})$$

9

5. A typical video codec

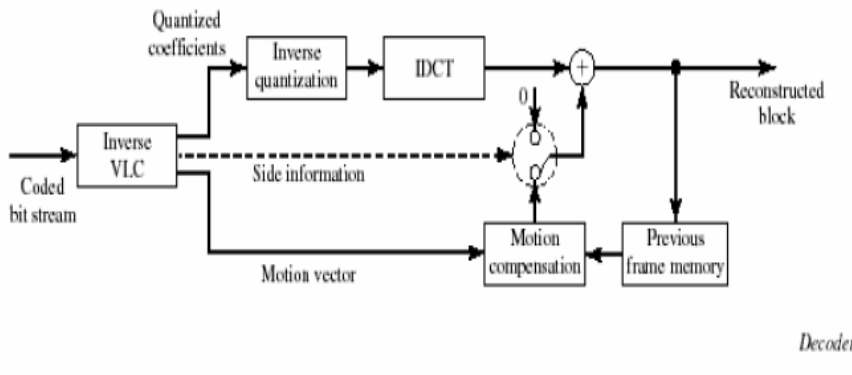
A typical encoder using block matching MC:



(courtesy Dr. Y. Wang)

10

A typical decoder with block matching MC



(courtesy Dr. Y. Wang)

11

6. Tradeoff: choosing inter or intra modes

Intra mode: apply to the 1st frame and refreshing (I) frames;
- low compression ratio ☹
- purely dependent on the current frame ☺

Inter mode: apply to P frames and B frames,
and only encode motion vectors
(and large residuals in the DCT domain);
- high compression ratio ☺
- motion errors can propagate through frames ☹

Selection of inter or intra mode:
tradeoff between rate and performance

12

7. Video coding standards (brief)

H.26x (ITU-T video coding standards):

(**ITU-T**: the ITU Telecommunication Standardization Sector)

H261, H263, H.264/AVC (advanced video coding standards)

MPEG-x: (MPEG audio and video coding & compression standards)

(**MPEG**: Moving Picture Experts Group)

MPEG-1, 2, 4, 7, 21 (for multimedia applications, especially object-based coding)

13

Basic features of video coding standards

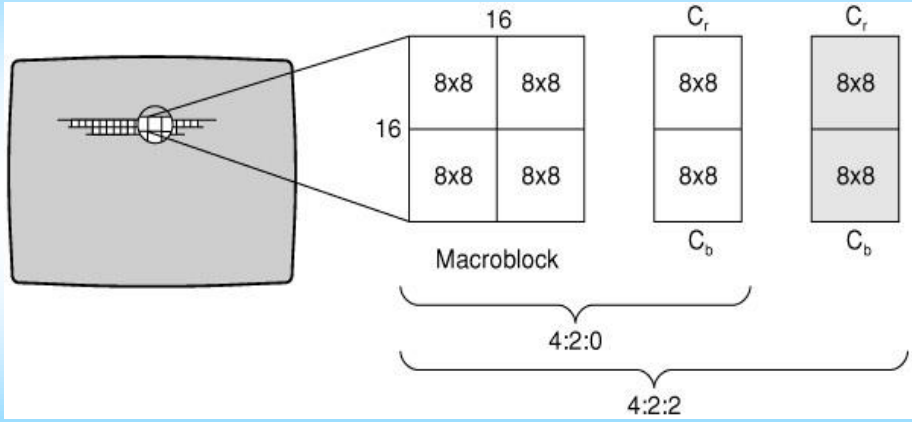
	MPEG-1 (1993)	MPEG-2 (1995)	MPEG-4 (2000)	H.261 (1993)	H.263 (1995)	H.264/ MPEG-4 AVC (2002)
Transform	8x8 DCT	8x8 DCT	8x8 DCT	8x8 DCT	8x8 DCT	4x4
MC Block Size	16 x 16	16 x 16 8x16	8 x 8, 16 x 16	16 x 16	8 x 8, 16 x 16	16x16, 16x8, 8x8, 8x4, 4x4
MC Accuracy	_ -pel	_ -pel	_ -pel	1-pel	_ -pel	1/8 -pel
Additional Motion Prediction Modes	- B-Frames	- B-Frames - Interlace	- B-Frames - Interlace - GMC (Global MC) - SPRITE Coding	-	- B-Frames	- B-Frames - Long term frame memory - in-loop deblocking filter - CAVLC/CABAC

MPEG-4: multimedia applications, object-based coding

14

Ref: T.Sikora, proc.IEEE, Vol.93, No.1, 2005

Macro-blocks and blocks in MPEG



15

Video Compression: H.264 / MPEG-4 AVC standard

- **Integer transform**
  (4x4 blocks, **to replace 8x8 DCT !**)

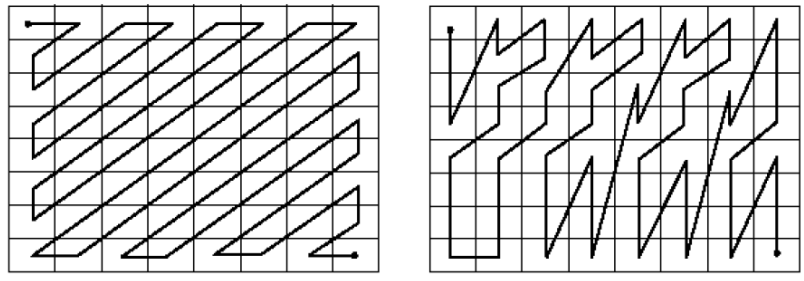
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

16

Zigzag scanning, alternative scanning

Two types of DCT and two types of scan pattern:

- **Frame DCT**: divides an MB into 4 blocks for Lum, as usual
- **Field DCT**: reorder pixels in an MB into top and bottom fields.



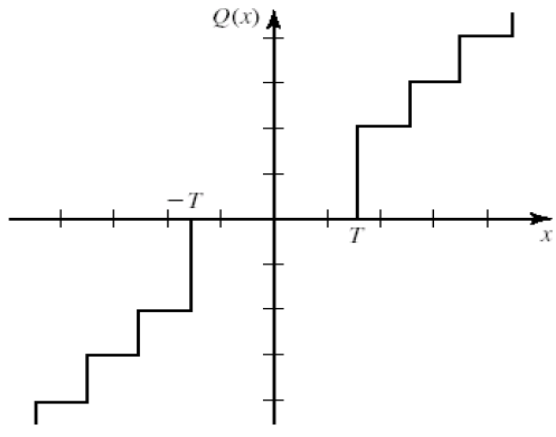
Zigzag scan Alternate scan

Figure 13.19 The zigzag scan as known from H.261, H.263, and MPEG-1 is augmented by the alternate scan in MPEG-2, in order to code interlaced blocks that have more correlation in the horizontal than in the vertical direction.

(Courtesy Dr. Y.Wang)

17

Scalar Quantization



Uniform,
Stepsize=8
with a dead zone

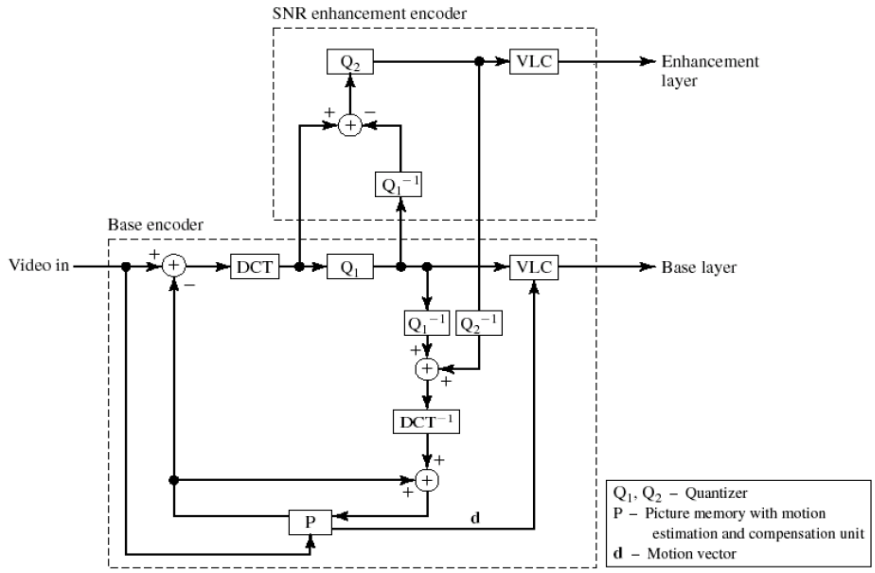
18

"Scalabilities" in the codec

- **SNR** scalability
- **Spatial** scalability
- **Temporal** scalability

19

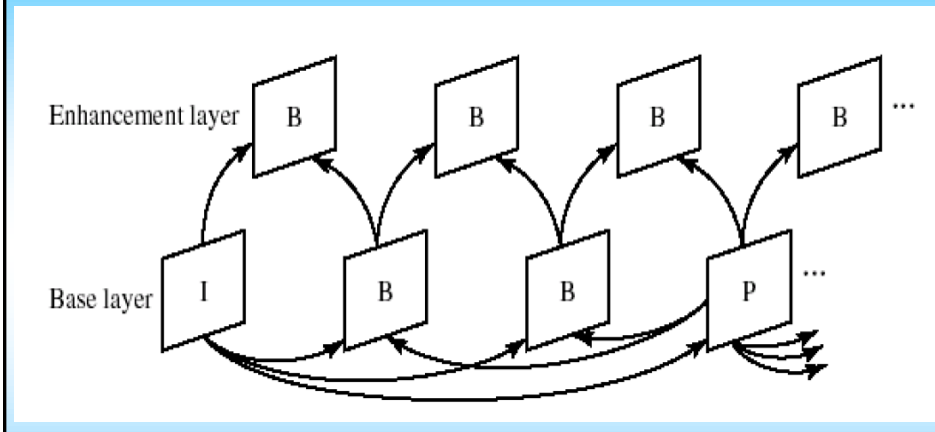
SNR (quality) Scalability



(Courtesy Dr. Y.Wang)

20

Temporal scalability: option-1

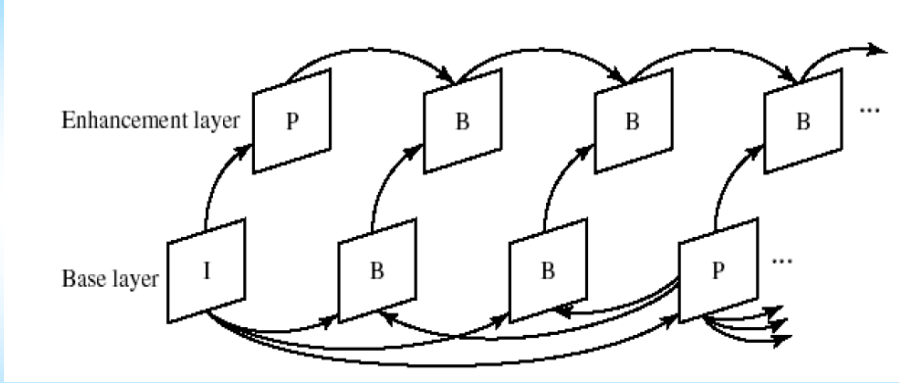


(courtesy Dr. Y.Wang)

- Use the base layer to predict B image frames in the enhanced layer.
- Errors in the enhanced layer do not propagation.

21

Temporal scalability: option-2



(courtesy Dr. Y.Wang)

Use both the base layer and the enhanced layer for image prediction in the enhanced layer

22

Spatial scalability

(courtesy Dr. Y.Wang)

23

MPEG-4: object-based coding

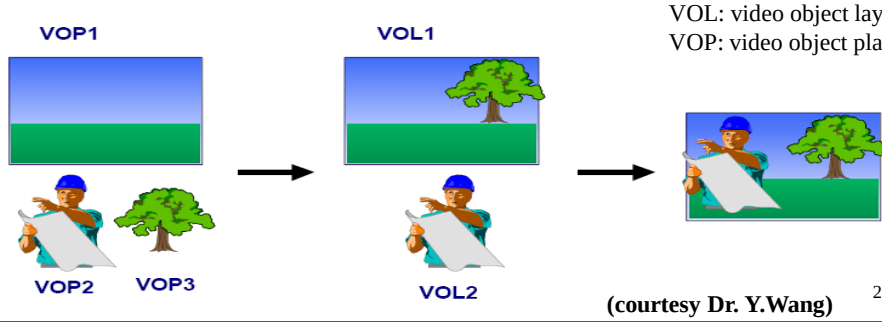
- Each image scene consists of multiple objects

  (require object segmentation /detection – not standardized)

  Example: using deep learning programs: U-Net, YOLOv3

- Objects are characterized by appearance, motion, shape, …

VO: video object
VOL: video object layer
VOP: video object plane



(courtesy Dr. Y.Wang)

24

Object detection by YOLO

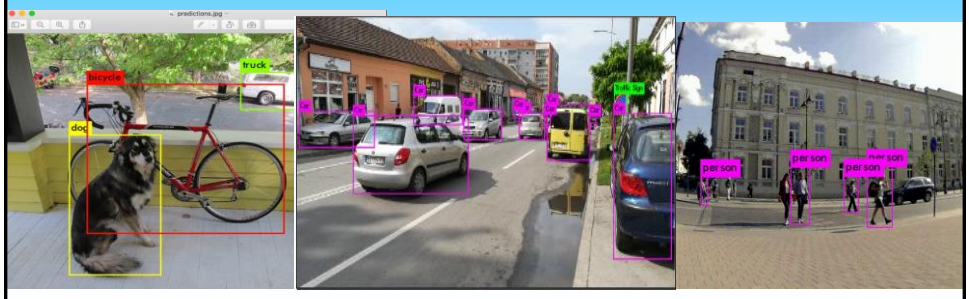
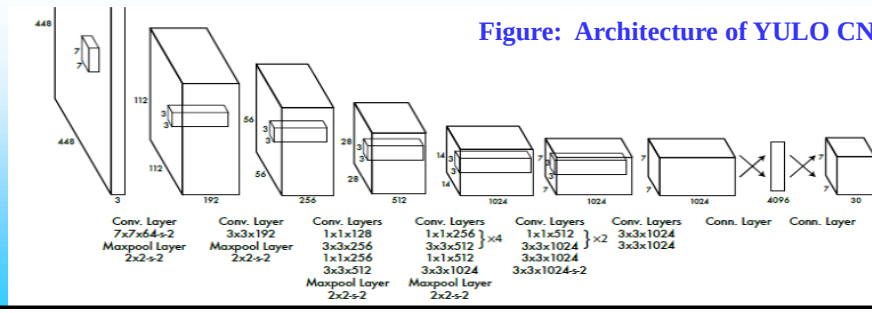
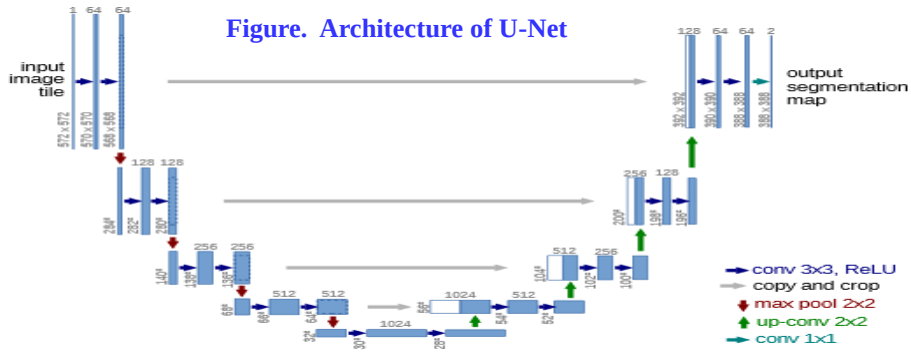


Figure: Architecture of YULO CNN

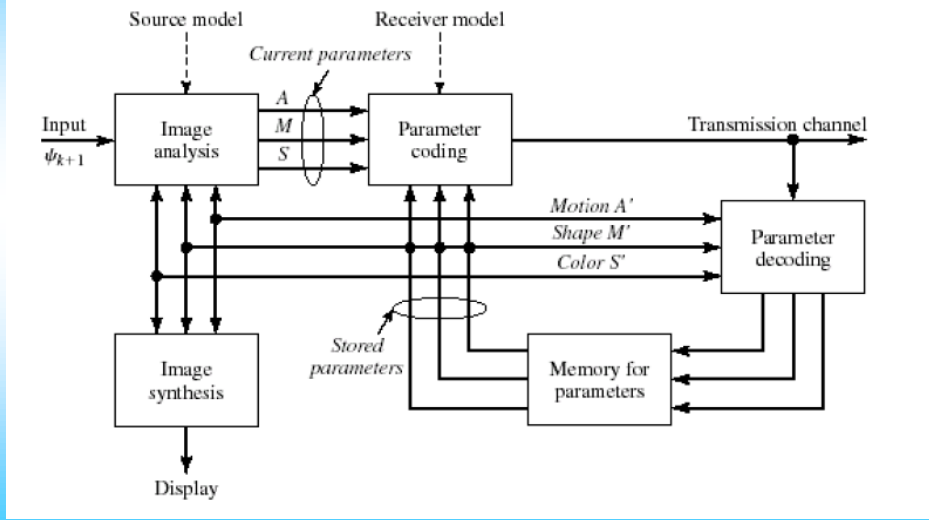


Semantic Segmentation by U-Net

Figure. Architecture of U-Net



Object-based coding



(Courtesy Dr. Y.Wang)

27

Visual Descriptors

Color:

Histogram, dominant color, etc.

Texture:

+ Homogeneity: energy in different orientation and frequency bands (Gabor transform)
+ Coarseness, directionarity, regularity
+ Histogram of intensity, histogram of oriented gradient (HoG)

Motion:

+ Camera motion
+ Motion trajectory of feature points in a non-rigid object
+ Motion parameters of a rigid object
+ Motion activity

Shape:

Boundary-based vs. region-based

28

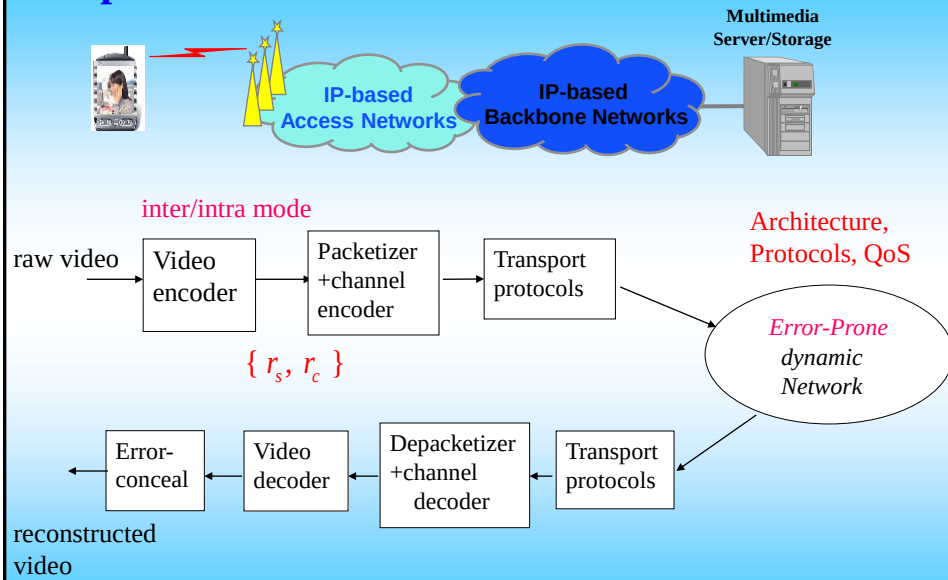
8. Why video communications: brief introduction + examples

Applications:

Video phone
Video conferencing
Video streaming
Distance learning
…

29

Compressed Video over Communication Networks



30

9: Lab-2: Video compression and coding using transforms, subband filters and motion compensation

Task-1: Learn how to handle video data in Matlab

e.g. read /save *.avi video files, play movies,
extract video frames, convert image formats in Matlab.

Task-2. Intra-mode compression

Apply block-based 2D DCT and
2D wavelet transform for single frame image compression

Task-3. Inter-mode compression based on MC

+ Compute motion blocks (apply a threshold to the difference image from frames)
+ Estimate motion vectors: Apply block matching with MSE/MAE criterion
+ Inter-frame compression through MC
use matched blocks from previous frames specified by MVs

Task-4: Compression of whole image frame:

Copy non-motion blocks from previous frame + MC image areas.

Task-5: Objective quality measures using PSNR, MSSIM

31

10. References

- **J.W.Woods, Subband image coding, Kluwer Academic Publisher,1991.**
- **D.S.Taubman, M.W.Marcellin, JPEG 2000, Kluwer Academic Pub.,2002.**
- **JVT /H.26L/ AVC Standards**
- **K.G.Nganm C.W.Yap, K.T.Tan,**
  Video coding for wireless communication systems, Marcel Dekker Inc., 2001
- **B.G. Haskell, A.Puri, A.N. Netraval,**
  Digital Video: an introduction to MPEG 2, Chapman & hall, 1997.
- **Proceedings of the IEEE, Vol.93, No.1, 2005:**
  Special Issue on advances in video coding and delivery

32