Overview : Image/Video Coding Techniques

A/Prof. Jian Zhang

NICTA & CSE UNSW

COMP9519 Multimedia Systems

S2 2009

THE UNIVERSITY OF NEW SOUTH WALES SYDNEY · AUSTRALIA

NICTA

Pixel Representation

- Y,U,V Colour Space
  - Colour can be represented by Red, Green and Blue components (*RGB*).
  - Transform to YUV or *YCbCr* with less correlated representation.

$$Y = 0.299R + 0.587G + 0.114B$$

$$U_t = \frac{B - Y}{2.03}$$

$$V_t = \frac{R - Y}{1.14}$$

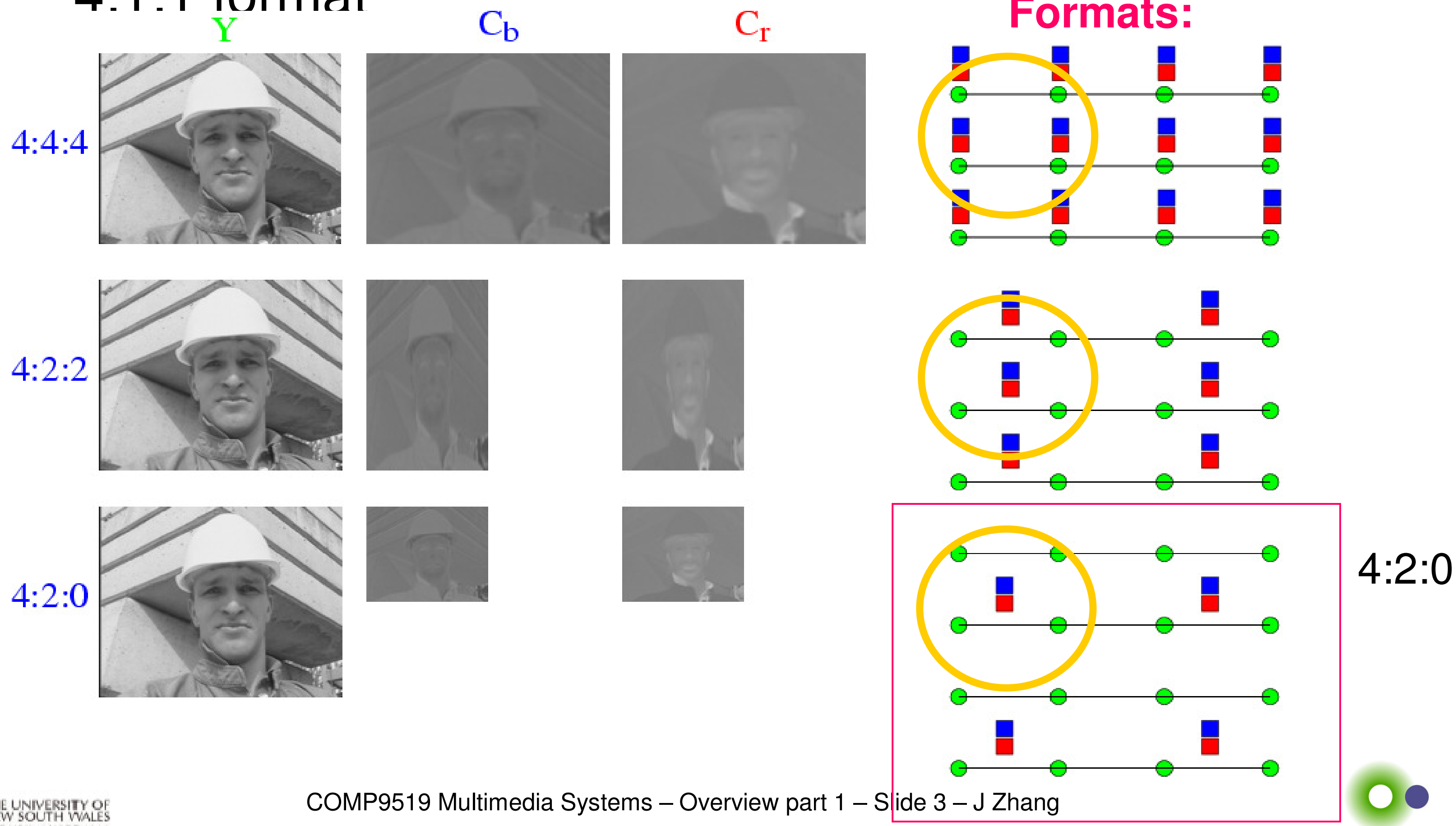
$$\begin{bmatrix} Y \\ U_t \\ V_t \end{bmatrix} = \underbrace{\begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Note:

The Y component represents luminance and the two chrominance components (U,V) contain considerably less information than the luminance component. For this reason, chrominance is often sub-sampled such as YUV 420. **The 420 format results in a ¾ data reduction in each U and V component.**

Basic Concepts

- Chrominance sub-sampling and formats
  - Y, Cb & Cr (YUV) Formats: 4:4:4 ->4:2:2 -> 4:2:0 or 4:1:1 format



Q1 – YUV Color Space

- You are given the task of designing an image encoder. You have the option of operating in either RGB colour space or YUV colour space. Which colour space would you choose for your image encoder ? Give reasons for your choice.

- YUV
- Compression
  - Chrominance component can be sub-sampled (e.g. 4:2:0) as the human visual system is more sensitive to luminance (Y) than chrominance (UV).

- Efficiency
  - Motion estimation can be done only on the Y component.

Q2 – YUV/RGB Color Space

- An RGB image is converted to YUV 4:4:4 format. "*The YUV 4:4:4 version of the image is of lower quality than the RGB version of the image*". Is this statement TRUE or FALSE ? Give reasons for your answer.

- FALSE
- Quality of YUV 4:4:4 = Quality of RGB
  - No sub-sampling
  - No data loss

Basic Transform coding

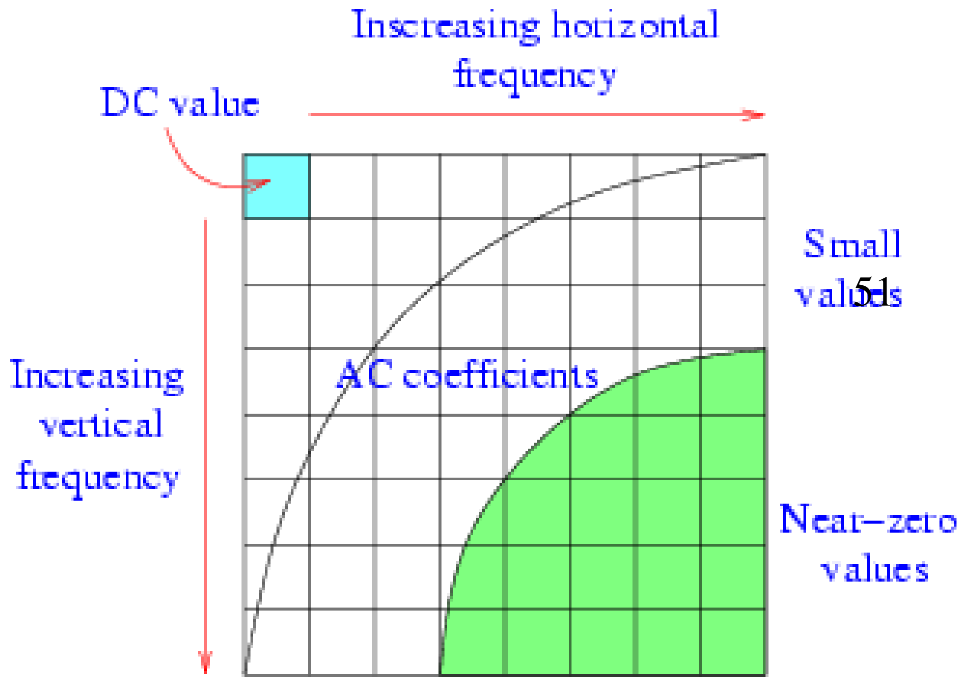
- Discrete Cosine Transform
  - For a 2-D input block U, the transform coefficients can be found as $Y = CUC^T$
  - The inverse transform can be found as $U = C^T YC$
  - The NxN discrete cosine transform matrix C=c(k,n) is defined as:

$$c(k,n) = \begin{cases} \dfrac{1}{\sqrt{N}} & for\, k = 0\, and\, 0 \le n \le N-1, \\ \sqrt{\dfrac{2}{N}} \cos \dfrac{\pi(2n+1)k}{2N} & for\, 1 \le k \le N-1\, and\, 0 \le n \le N-1. \end{cases}$$

Basic Transform Coding

- The distribution of 2-D DCT Coefficients

Ref: H. Wu



$$
\begin{bmatrix}
\boxed{68} & 3 & 5 & -2 & 0 & 0 & -2 & 0 \\
-10 & 0 & -4 & 3 & 0 & 0 & 0 & 0 \\
9 & 3 & 0 & 0 & 0 & -2 & 0 & 0 \\
3 & 2 & 0 & 3 & 0 & 2 & -2 & 0 \\
0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\
0 & 2 & -2 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Q3 – 8x8 Image Blocks

100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100

Block A

100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107
100	101	102	103	104	105	106	107

Block B

100	10	150	30	0	250	1	107
110	20	160	40	0	240	1	10
120	30	170	50	0	230	1	107
130	40	180	60	0	220	1	10
140	50	190	70	0	210	1	107
150	60	200	80	0	200	1	10
160	70	210	90	0	190	1	107
170	80	220	100	0	180	1	10

Block C

Q3 – 8x8 DCT Transformed Blocks

800	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Block A

828	-18	0	-2	0	-1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Block B

705	152	-24	68	-188	67	532	-139
-60	-113	-1	50	-14	-33	34	-13
0	0	0	0	0	0	0	0
3	-25	12	-6	8	-11	9	-4
0	0	0	0	0	0	0	0
13	-25	20	-16	15	-13	9	-5
0	0	0	0	0	0	0	0
43	-62	57	-51	44	-35	24	-12

Block C

Q3 – 8x8 Quantized DCT Transformed Blocks (Q-stepszie = 4)

200	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Block A

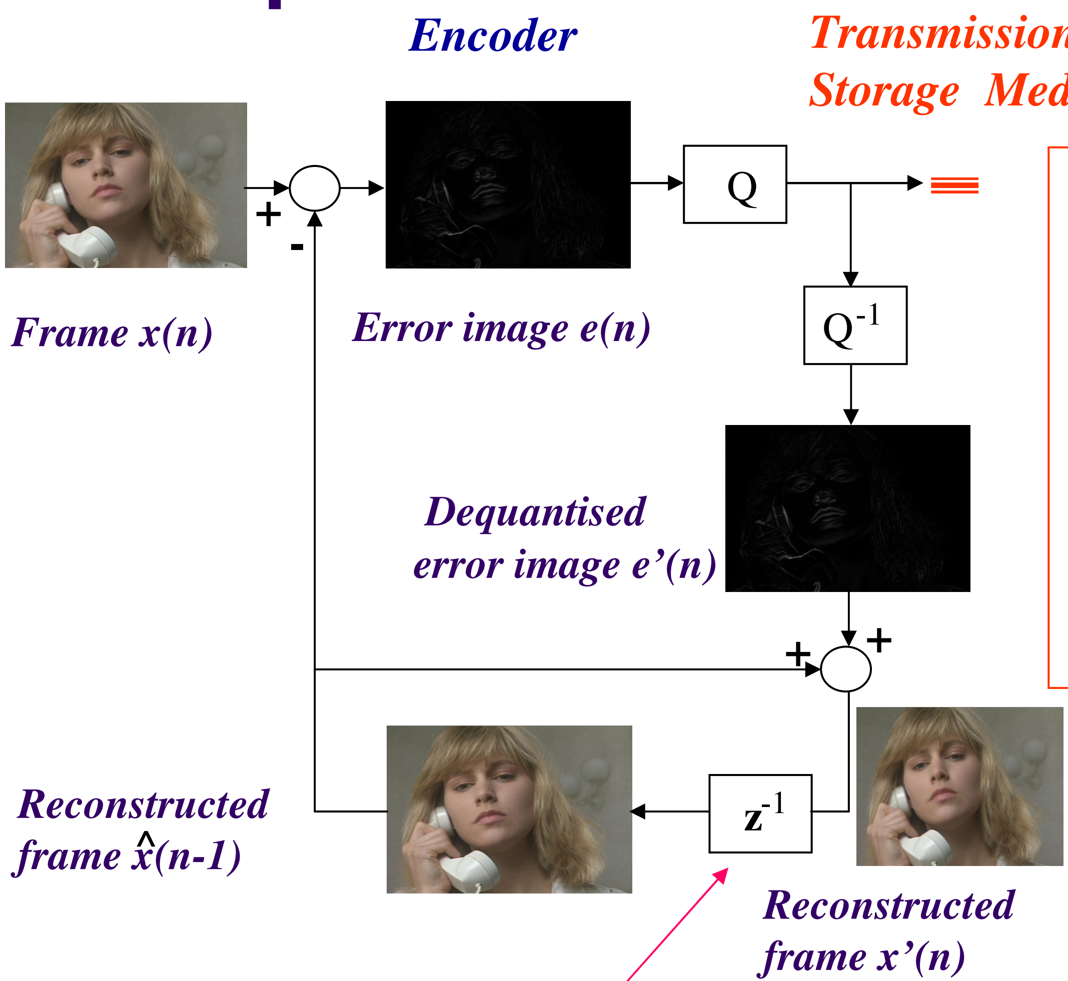
207	-5	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Block B

176	38	-6	17	-47	17	133	-35
-15	-28	0	13	-4	-8	9	-3
0	0	0	0	0	0	0	0
1	-6	3	-1	2	-3	2	-1
0	0	0	0	0	0	0	0
3	-6	5	-4	4	-3	2	-1
0	0	0	0	0	0	0	0
11	-15	14	-13	11	-9	6	-3

Block C

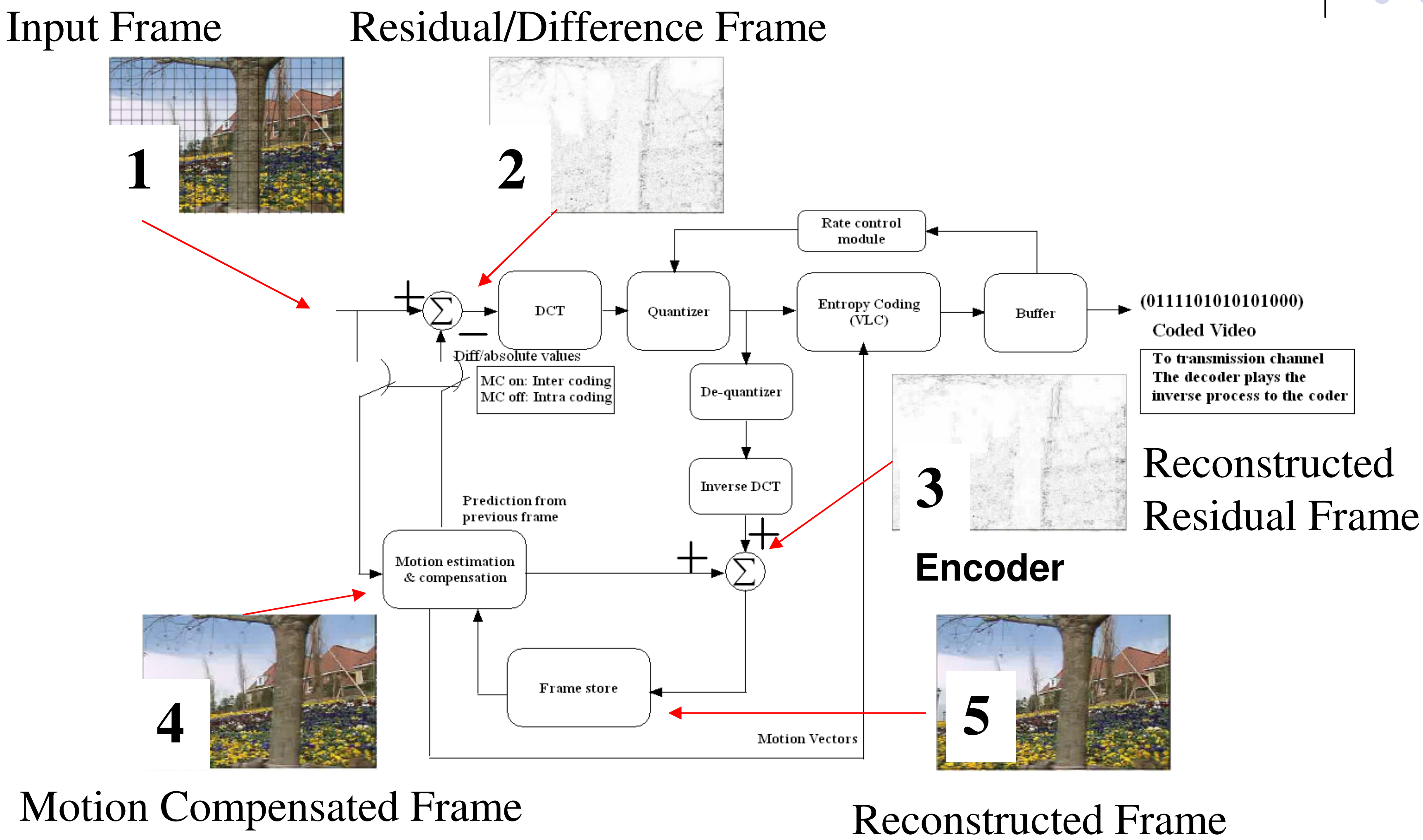
Simple Inter-frame Encoder



Step 1: Calculate the difference between the current ***x(n)*** and previous recon. ***$\hat{x}$(n-1)*** frames;
Step 2: Qantise and encode the difference image.
Step 3: Add the dequantised (residual) image to the previous frame ***$\hat{x}$(n-1)*** to reconstruct the current frame of image.

Delay to load the recon. Frame x'(n)

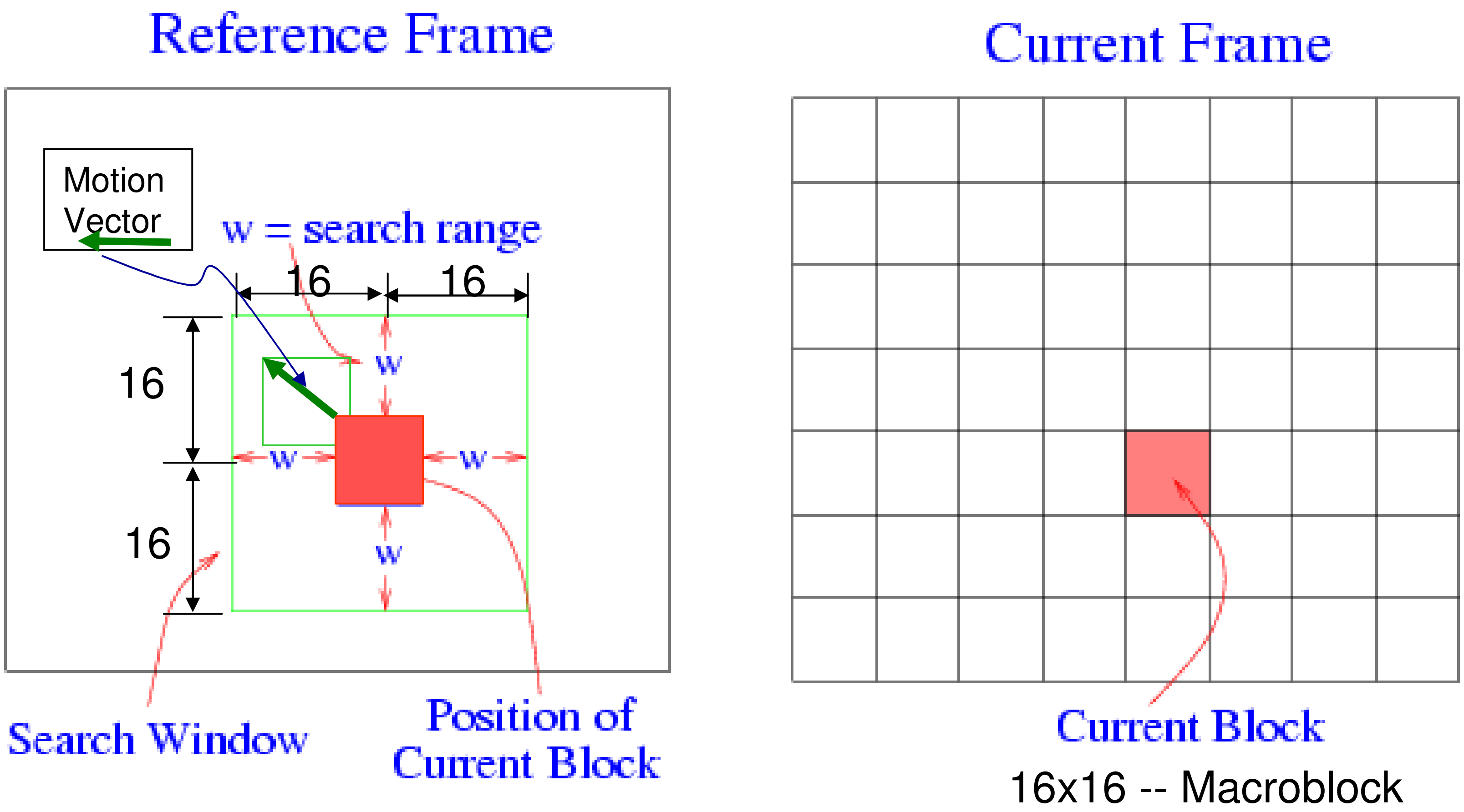
Q4 -- codec



Input Frame
Residual/Difference Frame
1
2
Rate control module
DCT
Quantizer
Entropy Coding (VLC)
Buffer
(0111101010101000)
Coded Video
To transmission channel
The decoder plays the inverse process to the coder
Diff/absolute values
MC on: Inter coding
MC off: Intra coding
De-quantizer
Inverse DCT
3
Reconstructed Residual Frame
Encoder
Prediction from previous frame
Motion estimation & compensation
Frame store
Motion Vectors
4
5
Motion Compensated Frame
Reconstructed Frame

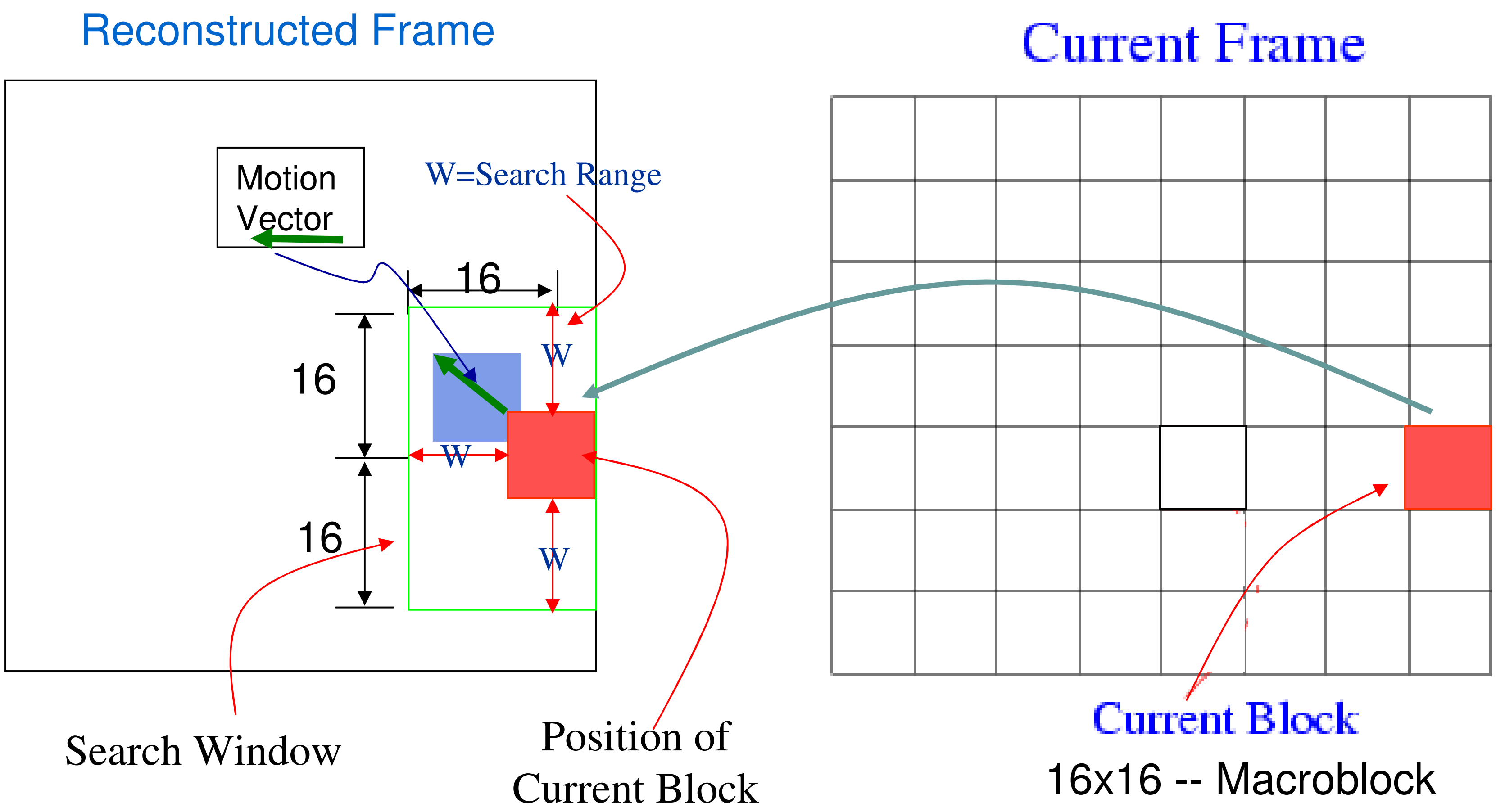
Block Based Motion Est.

- Block base search



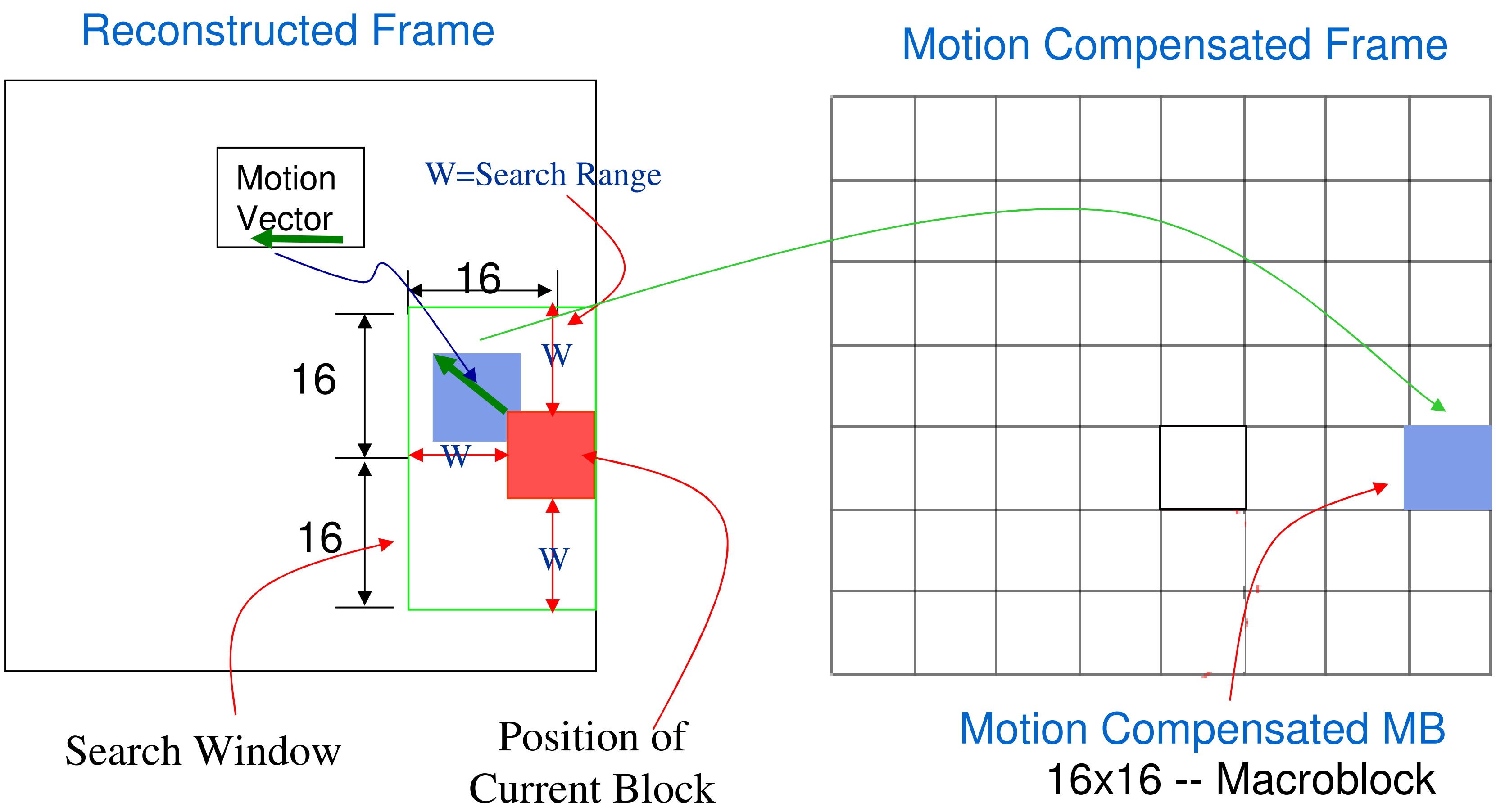
Block Based Motion Estimation

- Block base search

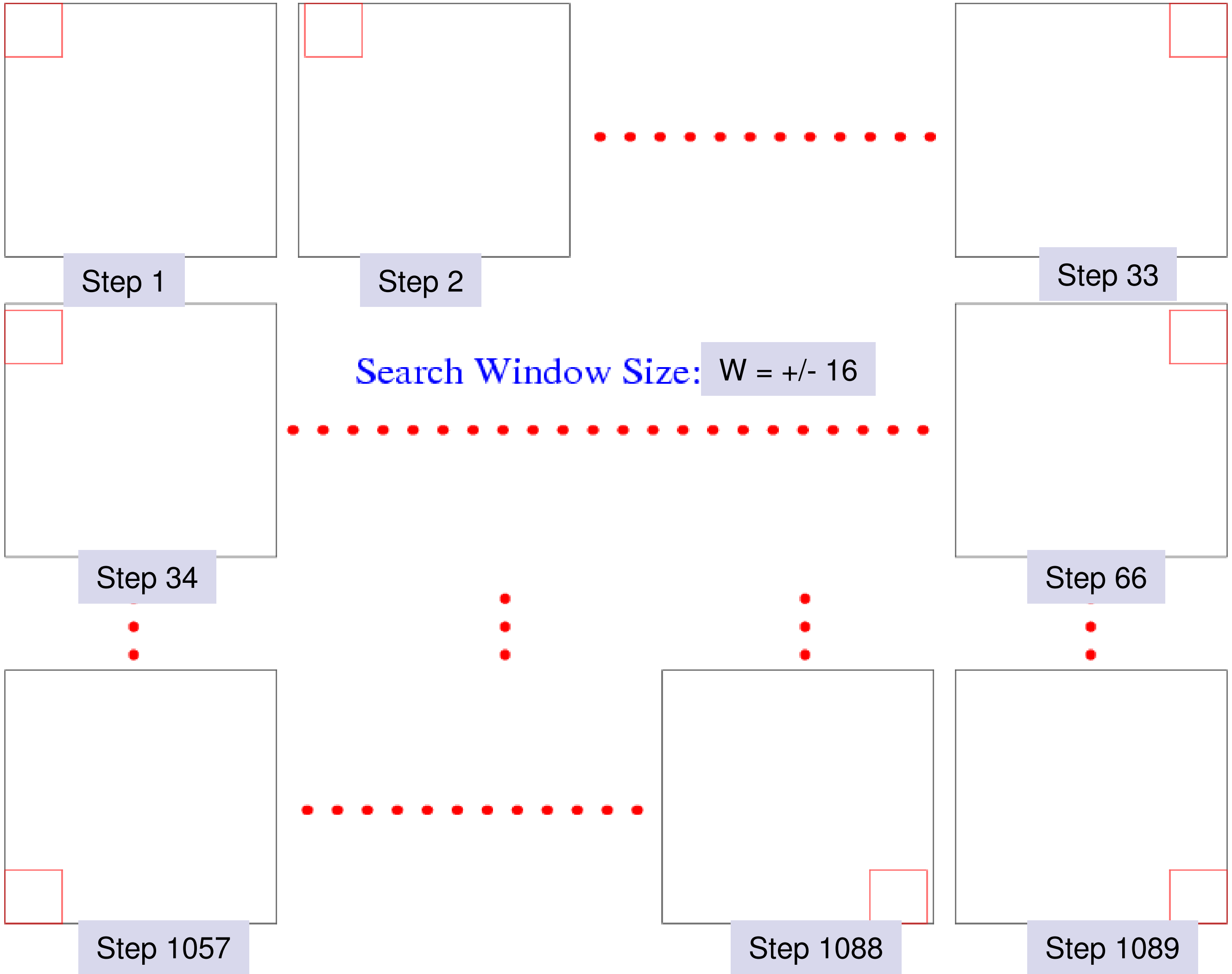


Block Based Motion Estimation

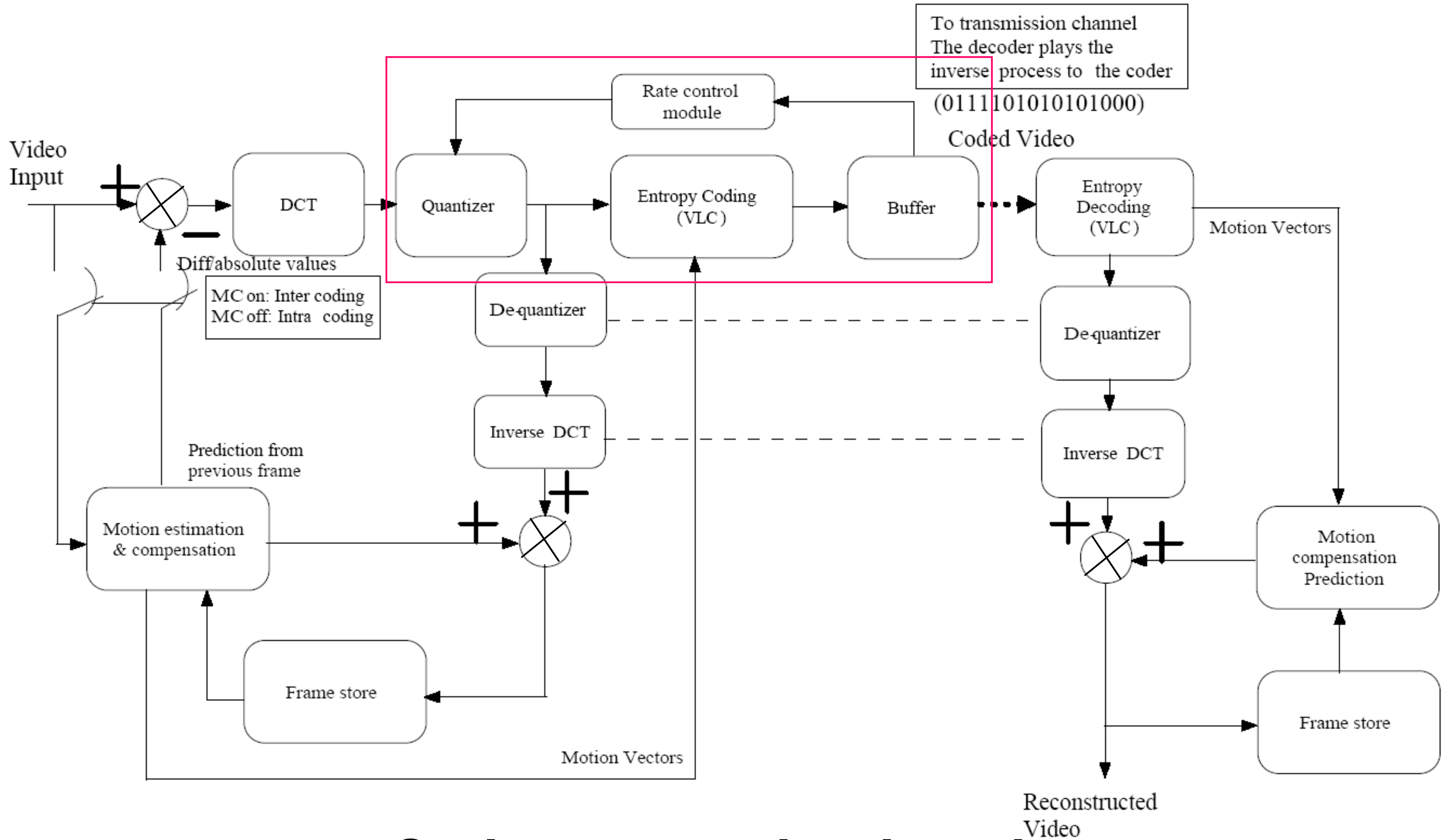
- Block base search



Full Search Algorithm

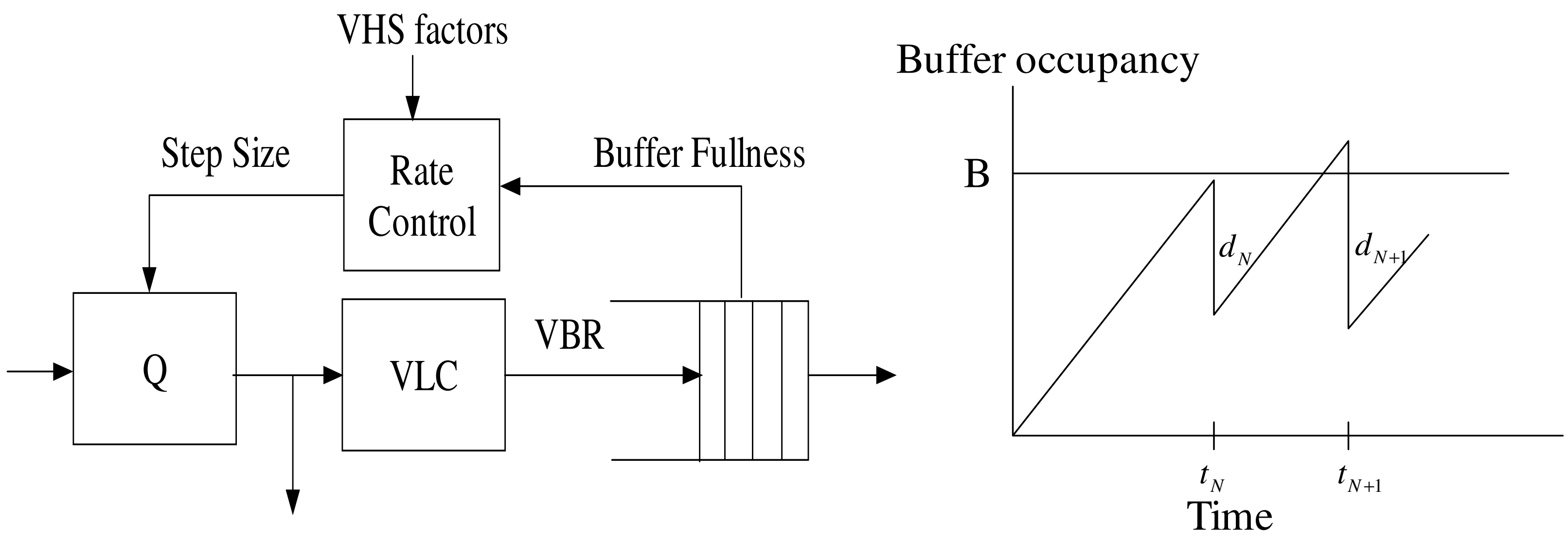


Motion Compensated DCT based Codec



Codec = encoder/decoder

Motion Compensated DCT based Codec – Rate Control



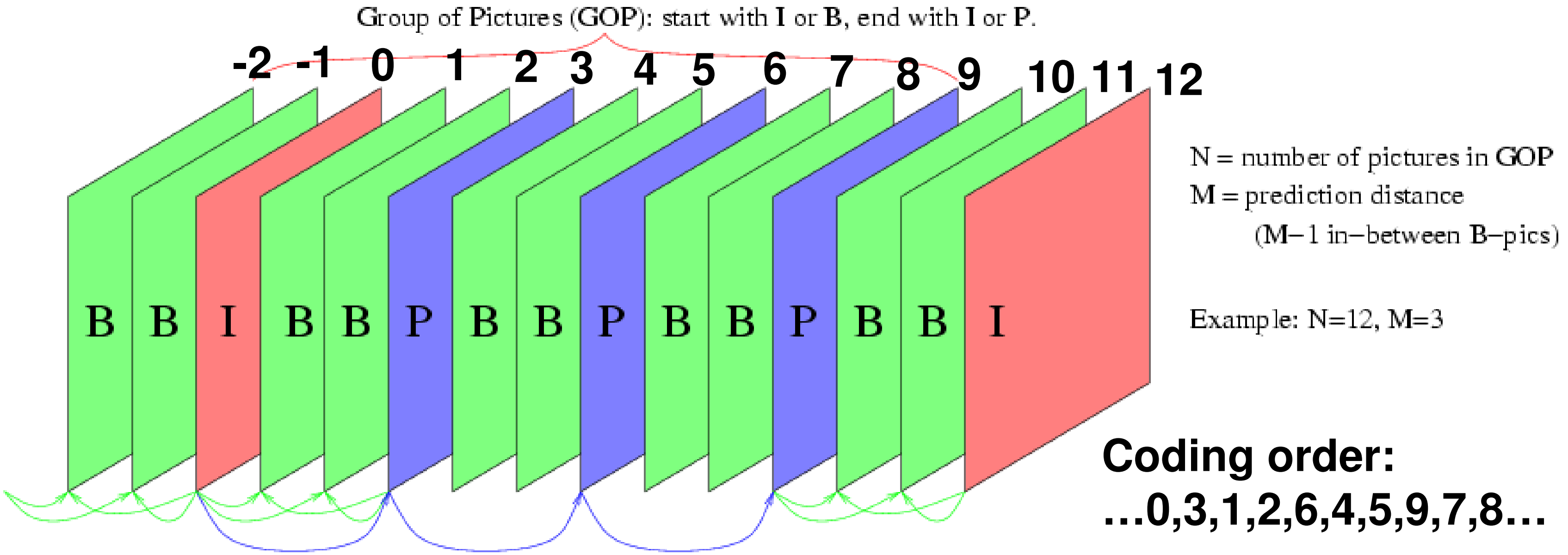
To maintain constraint bit rate operation, the variable rate output of the source coder is fed into a rate smoothing buffer, the fullness of which is used to control Coding parameters which trade-off bit rate and quality .

Hypothetical Reference Decoder is defined to verify the valiadiaty of the output of Codec.

The Key Operation in Video Codec – Motion Estimation

- Inter-frame coding (Zero MV) scheme achieves a low computation complexity while possibly results in a high error residuals due to no other search positions that are considered to find an optimal (low error residuals) position.
- Motion compensated coding scheme achieves a low error residuals while results in a high computationally complex due to an exhaustive search strategy within a defined search window
- Several sub-optimum fast search techniques have been developed. However, the quality-cost trade-off is usually worthwhile

Digital Video Coding (DVC) Standards– MPEG-1/2

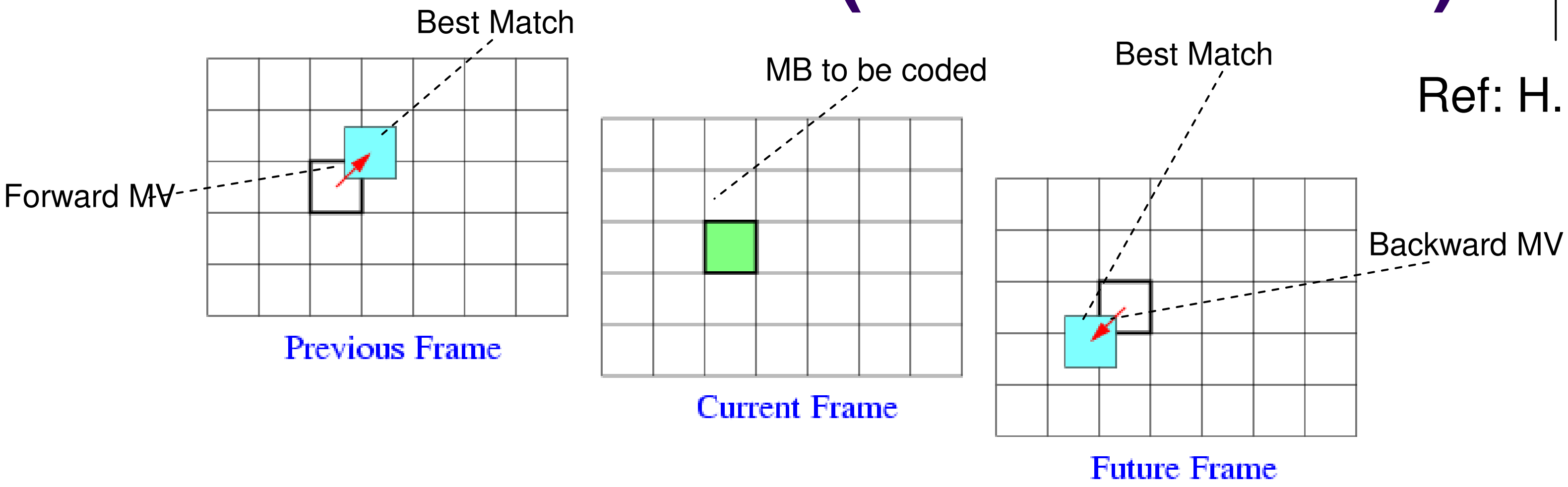


Coding order:
…0,3,1,2,6,4,5,9,7,8…

- **Intra coded picture (I-Picture):**
  - Coded on their own (all MBs are intra) and server as random access.
- **Predicted picture (P-Picture):**
  - Coded with reference (MC predictions) to the previous anchor I or P picture.
- **Bidirectionally predicted picture (B-Picture):**
  - Coded with reference to the previous and/or future anchor I or P pictures (forward or backward MC prediction and/or linear interpolation).

Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)

Ref: H. Wu



Forward prediction: Predict where the pixels in a current frame were in a past frame.

Backward prediction: Predict where the pixels in a current frame will go in a future frame.

Prediction for a macroblock may be backward, forward, or an average of both.

Advantages

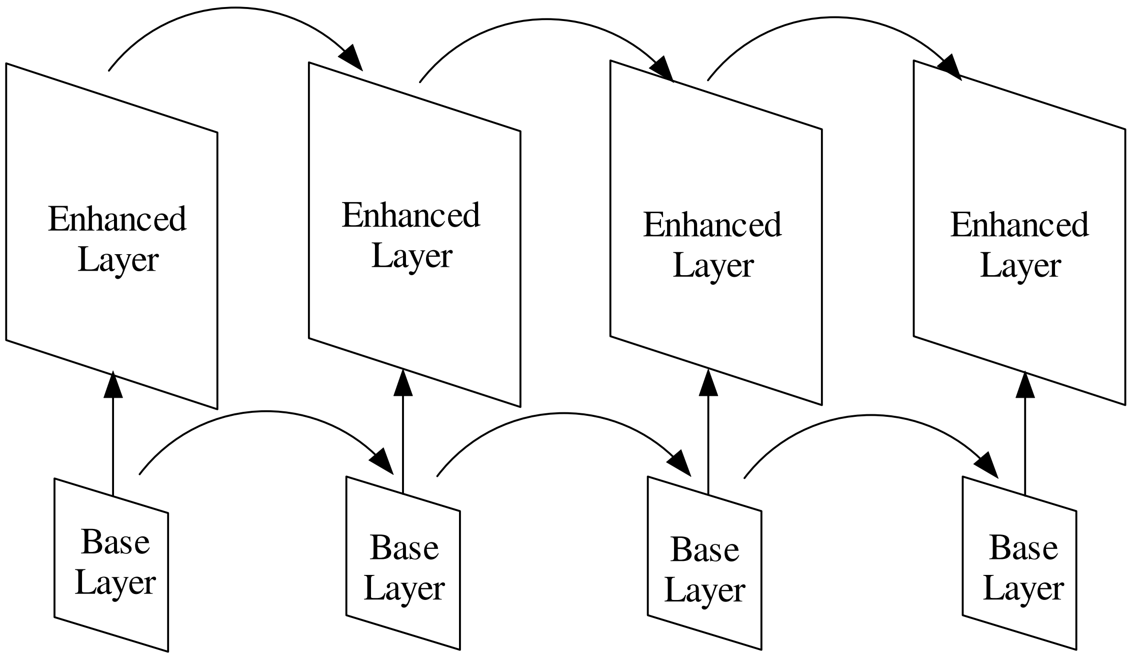
Main Advantage:

- High coding efficiency (gain/cost is significant)
- No uncovered background problem

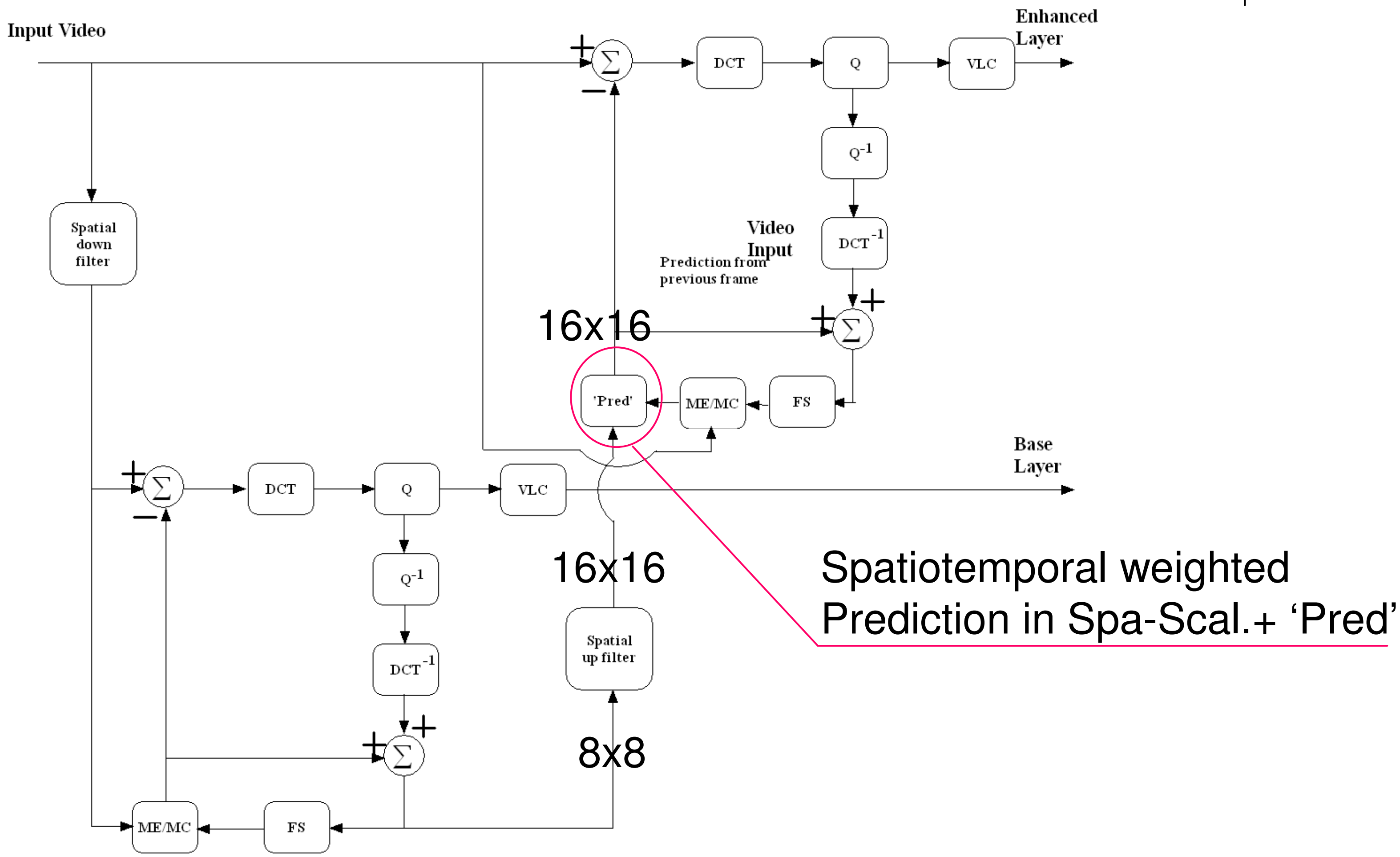
Major disadvantage: long delay and more memory to store two anchor frames

Digital Video Coding (DVC) Standards– MPEG-2 Scalability

- Spatial Scalability Types
  - Progress to progress
  - Progress to interlaced
  - Interlaced to progress
  - Interlaced to interlaced



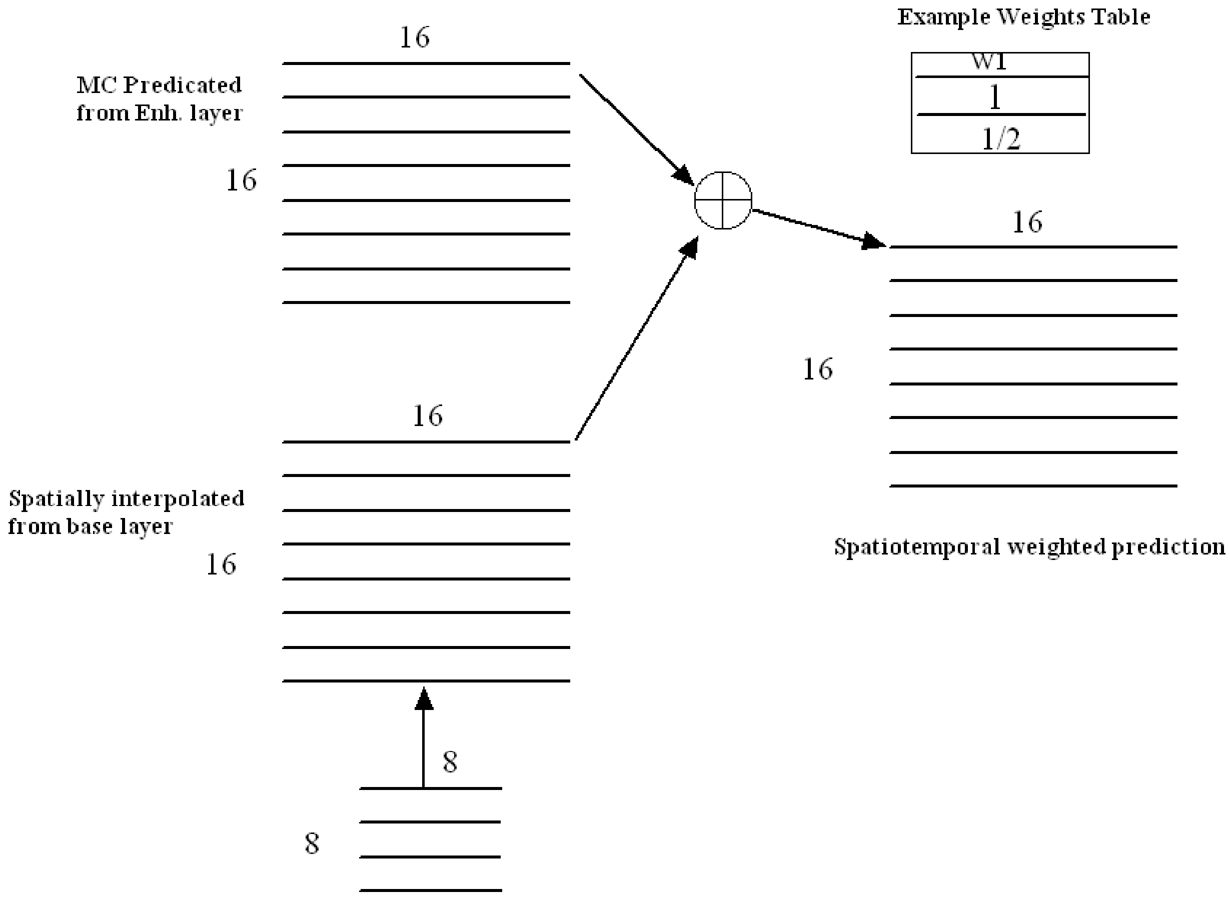
Digital Video Coding (DVC) Standards– MPEG-2 Scalability



2 layer spatially scalable coder

Digital Video Coding (DVC) Standards– MPEG-2 Scalability

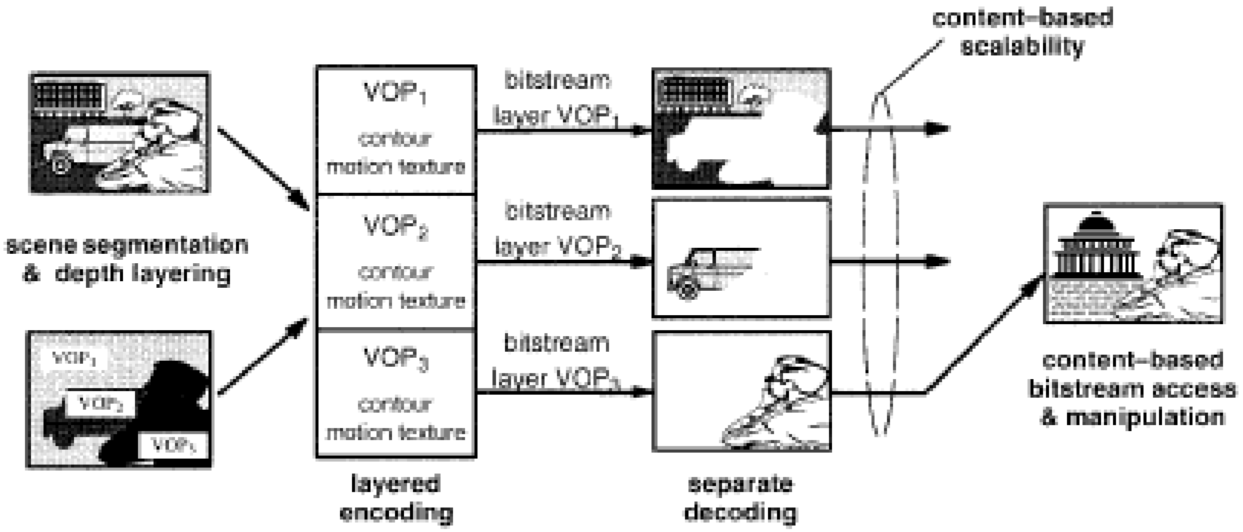
- Spatiotemporal weighted Prediction



MPEG-4 Visual Standard

- Access and manipulation of arbitrarily shaped images

Ref: Thomas Sikora



Object Based MPEG-4 Video Verification Model

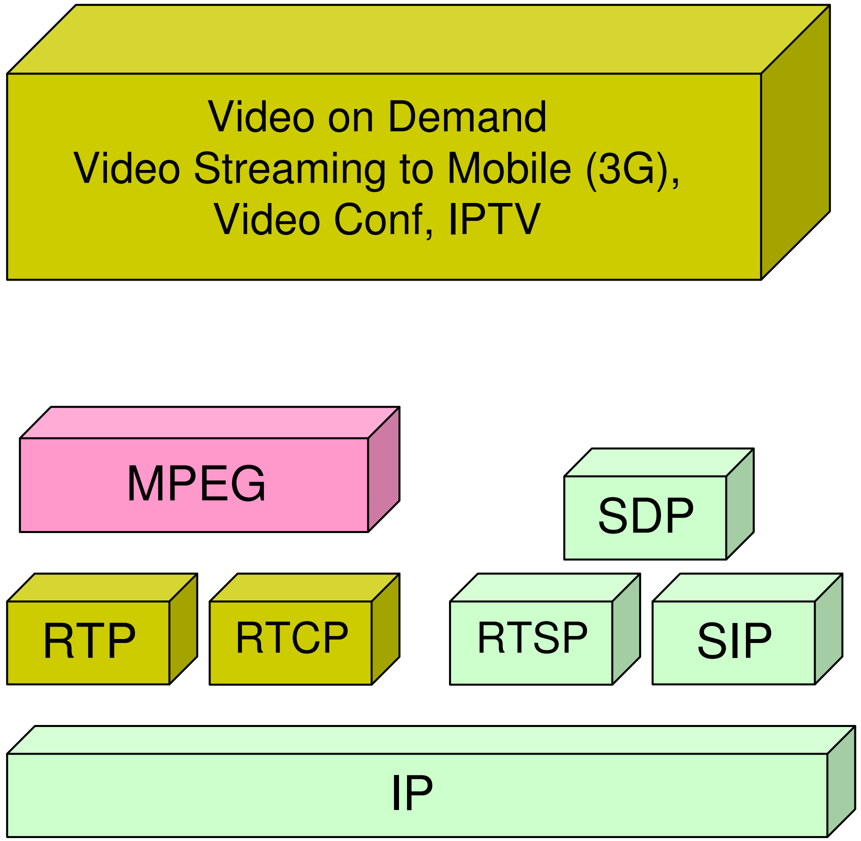
1. In MPEG-4, scenes are composed of different objects to enable content-based functionalities.
2. Flexible coding of video objects
3. Coding of a “Video Object Plane” (VOP) Layer

Internet Streaming Media

Reji Mathew & Jian Zhang
NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2009

Multimedia Streaming

- UDP preferred for streaming
- System Overview
  - Protocol stack
- Protocols
  - RTP + RTCP
  - SDP
  - RTSP
  - SIP
- Encoder Side Issues
- Receiver Side Issues
- File Format : MP4



Introduction

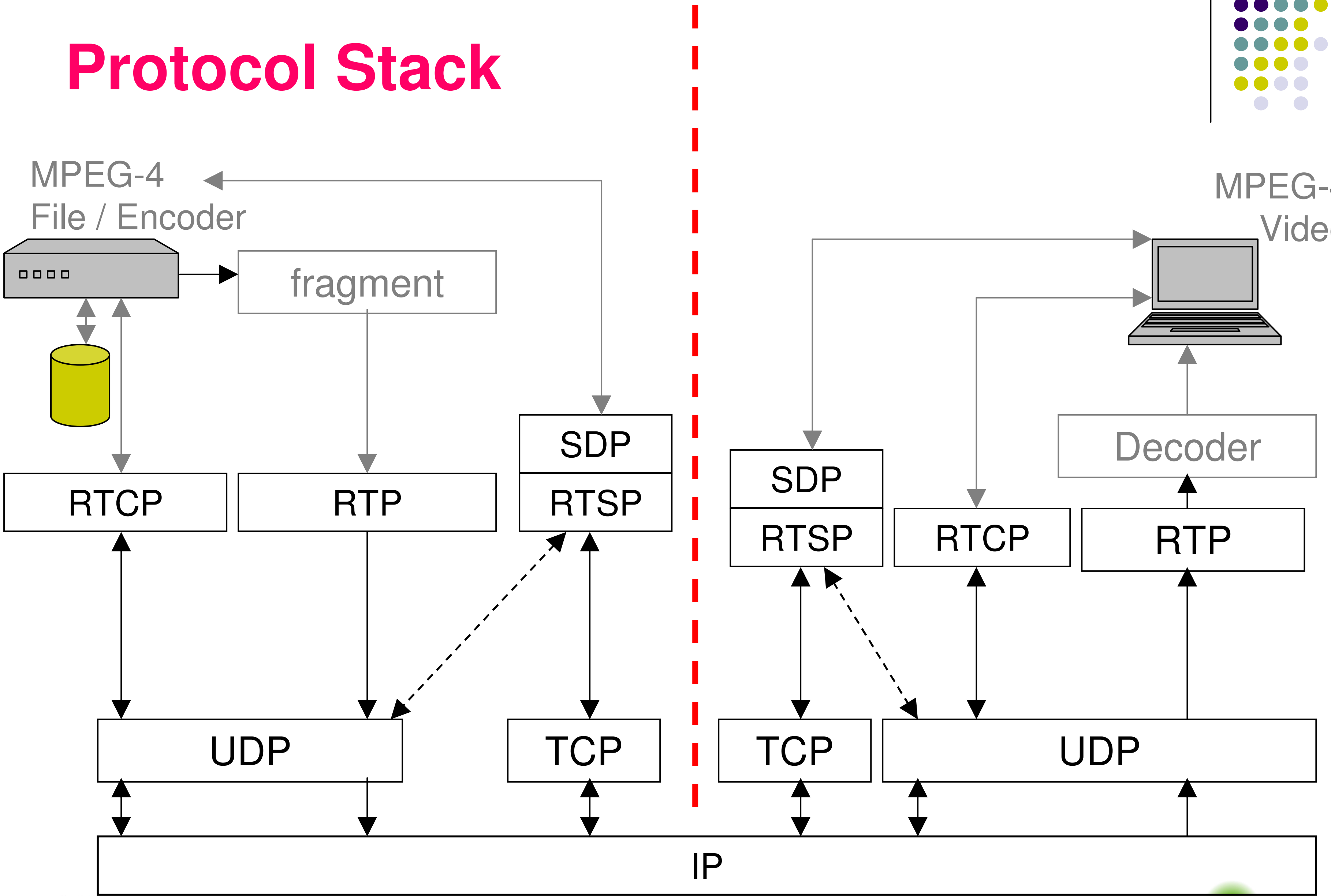
- Streaming video, TCP or UDP ?
  - What are the reasons for your preference ?

- Requirements for streaming video & audio ?
  - Or why not just stream over UDP ?

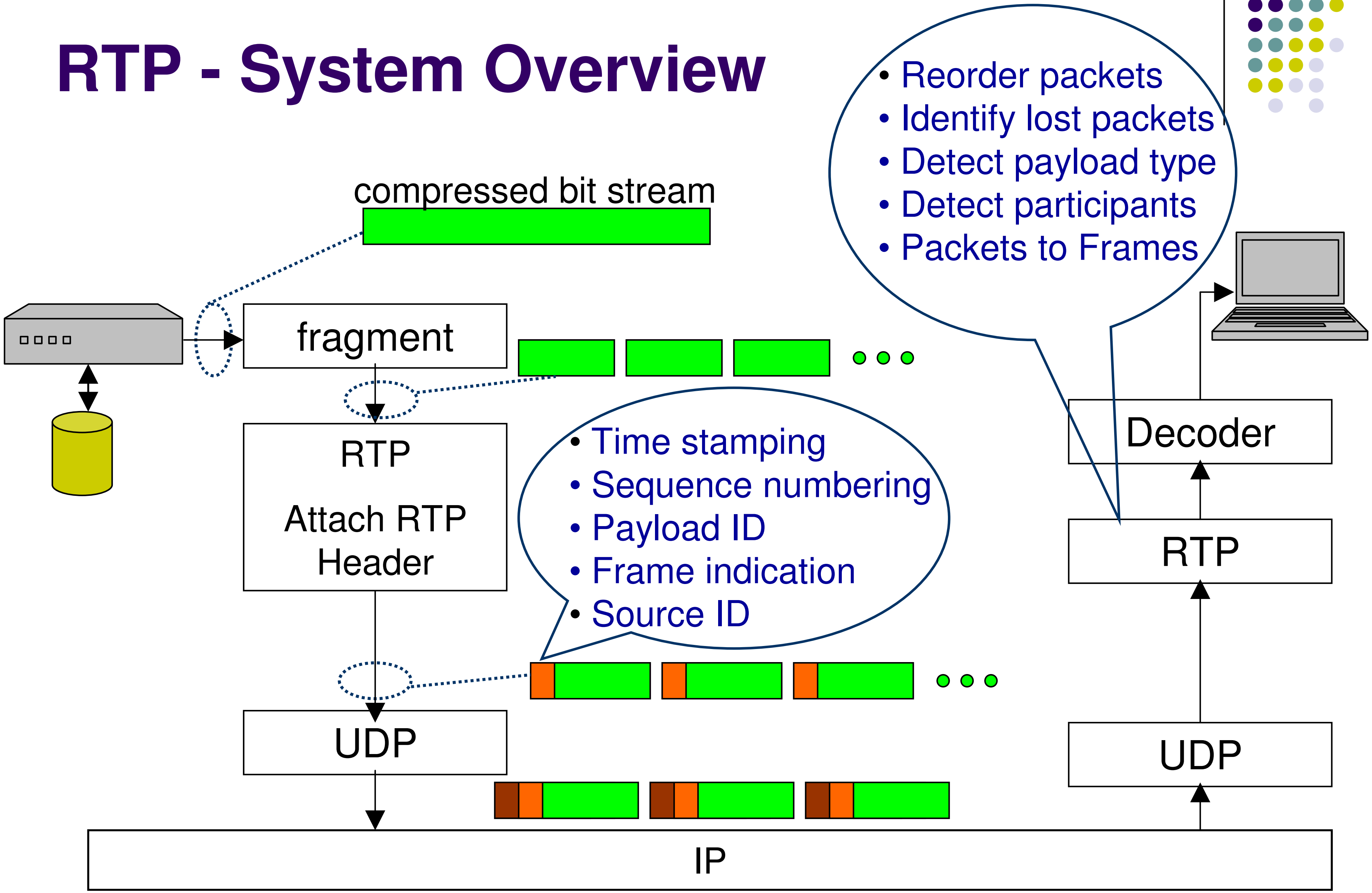
- Draw IETF protocol stack for streaming video ?
  - At Sender & Receiver,
  - show protocols for streaming over IP

- Explain the role or function of each IETF protocol ?
  - *Why do we need RTP, RTCP, SDP, RTSP ?*
  - *For information read introduction of corresponding IETF RFC documents*

Protocol Stack



RTP - System Overview



RTCP

- Need some feedback on network performance

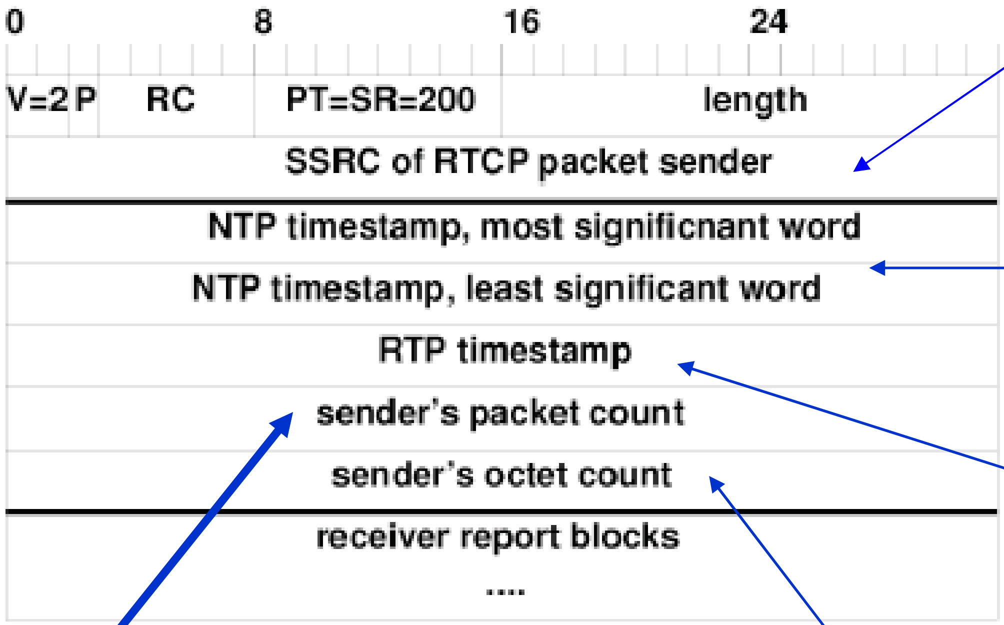
- Real-Time Control Protocol (RTCP)
  - Used in conjunction with RTP
  - Primary function – feedback on quality of data distribution

RTCP

- Control information exchanged between session members by RTCP packets
- Five RTCP packet types defined
- RTCP packets begin with a fixed part
- RTCP packets are sent periodically
- RTCP packets are stackable – compound packets
- Usually sent over UDP (same as RTP)
- RTCP packet types to carry control information
  - RTCP SDES : Source description items
    - sent by all session members
    - includes a persistent identifier of sources in the session
  - RTCP SR : Sender Report
    - sent by active senders
    - report of transmission and reception statistics
  - RTCP RR : Receiver Report
    - Sent by receivers (not active senders)
    - report of reception statistics
  - RTCP BYE : Packet to indicate end of participation
  - RTCP APP : Application specific functions

RTCP

- RTCP Sender Report Packet



SSRC ID for the originator of this Sender Report

Network Time Protocol timestamp (wallclock time) when this report was sent

Associated RTP time stamp (same instant as NTP timestamp).

Used for synch at the receiver

Total number of RTP packets sent since start of transmission

Total number of payload octets sent since start of transmission

RTCP

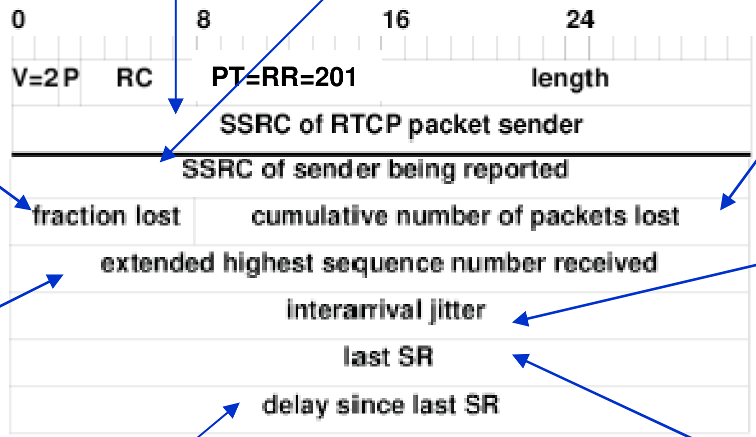
- RTCP Receiver Report Packet

SSRC ID for the originator of this RR

SSRC ID of sender

Fraction of RTP packets lost since last Receiver Report

Total number of RTP packets lost since start of transmission



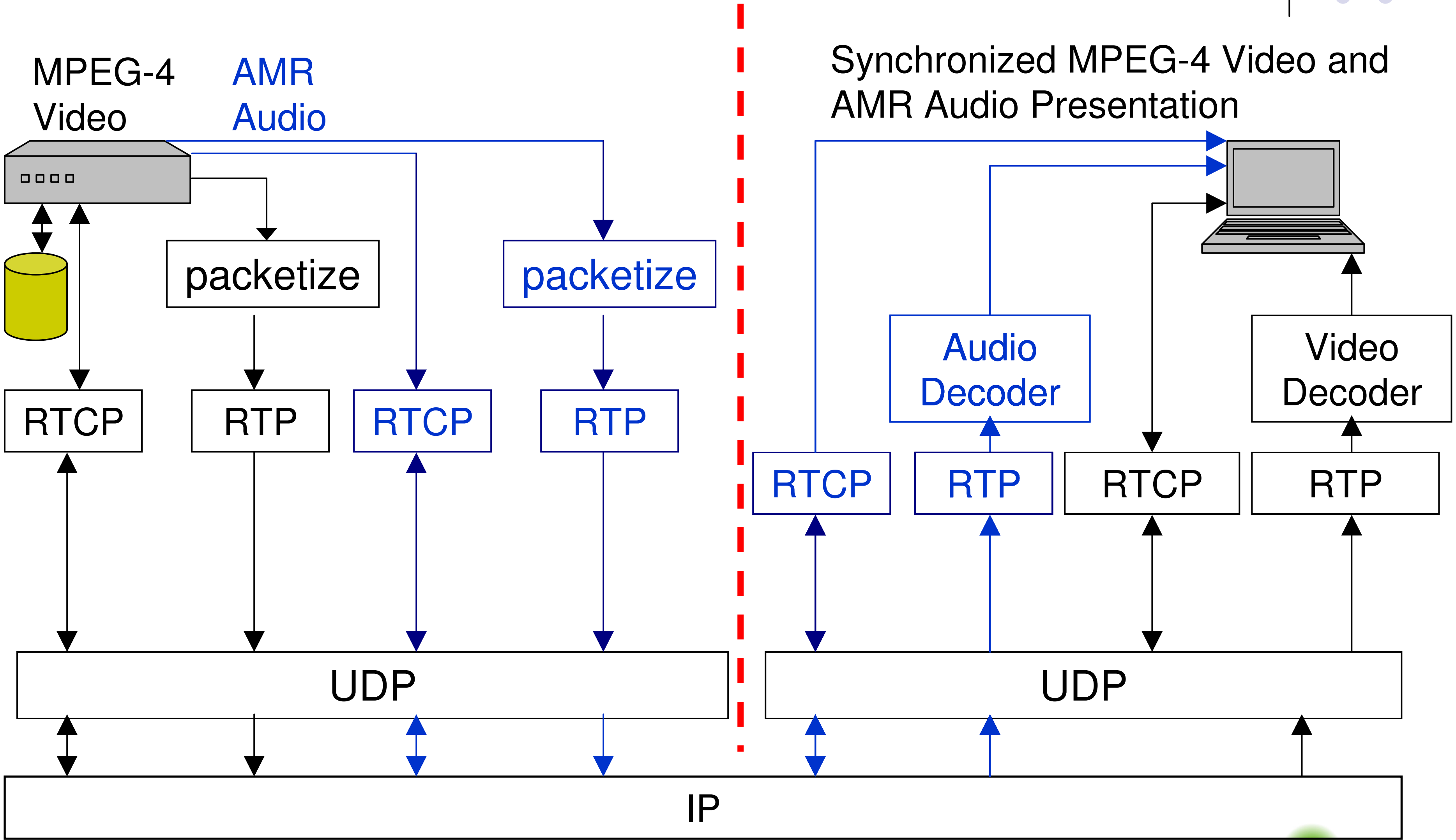
Highest RTP Sequence number of packets received

RTP packet inter-arrival time

Delay between receiving Sender Report and sending this Receiver Report

RTP timestamp of last Sender Report received from source

RTP + RTCP : System Overview



Further Protocols Enabling Streaming

- SDP – Description

```
v=0
o=mmvc 2890844526 2890842807 IN IP4 129.94.135.201
s=Camera ONE
i=Video stream for realtime surveillance
u=http://www.nicta.com/mmvc/demos/SurveillanceVideo.pdf
e=Jian.Zhang@nicta.com.au (Jian Zhang)
c=IN IP4 225.0.0.37/2
t=0 0
a=recvonly
m=video 20000 RTP/AVP 98
a=rtpmap:98 MP4V-ES/90000
a=fmtp:98 profile-level-id=1; config=000001b001000001…
a=orient:portrait
```

RTSP

- How to control streams
  - Start, Stop, Pause, Fast Froward, Rewind
  - “internet VCR”

- Solution – RTSP
  - Real Time Streaming Protocol
  - Establishes and controls one or more continuous media streams - such as audio and video

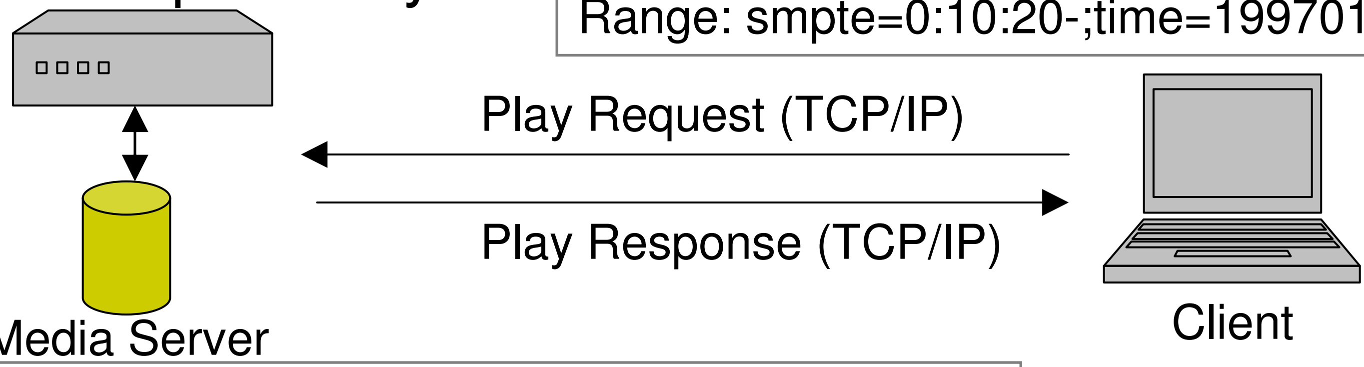
  - Similar in syntax and operation to HTTP/1.1
    - Client –Server protocol
    - Text based

  - IETF RFC 2326

RTSP

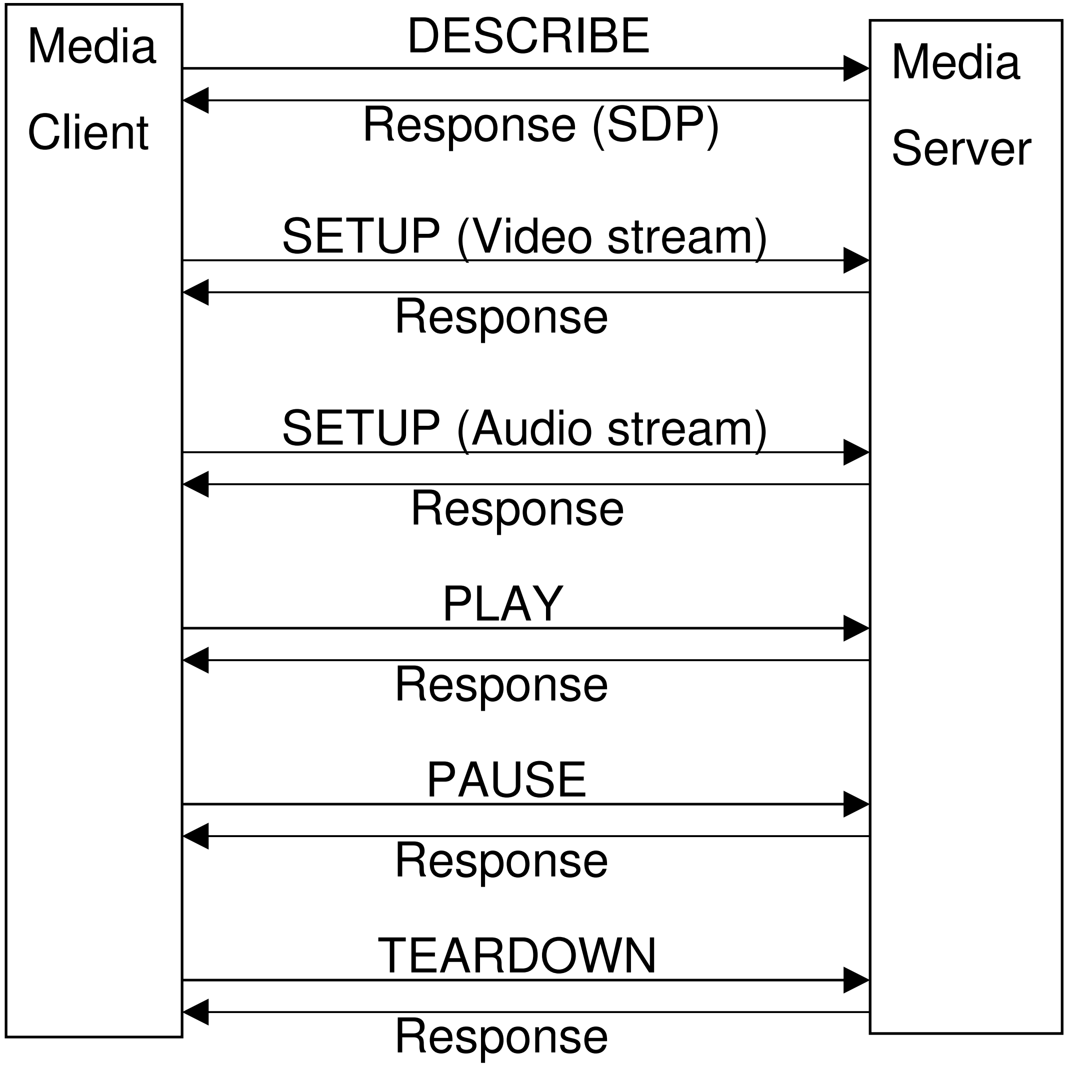
- Protocol Operation
  - Text based messages between client and server
  - Messages can be :
    - Requests
    - Responses
- Example : Play

PLAY rtsp://audio.example.com/twister.en RTSP/1.0
CSeq: 833
Session: 12345678
Range: smpte=0:10:20-;time=19970123T153600Z



RTSP/1.0 200 OK
CSeq: 833
Date: 23 Jan 1997 15:35:06 GMT
Range: smpte=0:10:22-;time=19970123T153600Z

RTSP : Signal Timing Diagram



System Overview

Receiver Side Issues

-Buffering input data

-Error concealment

-Inter media synchronization

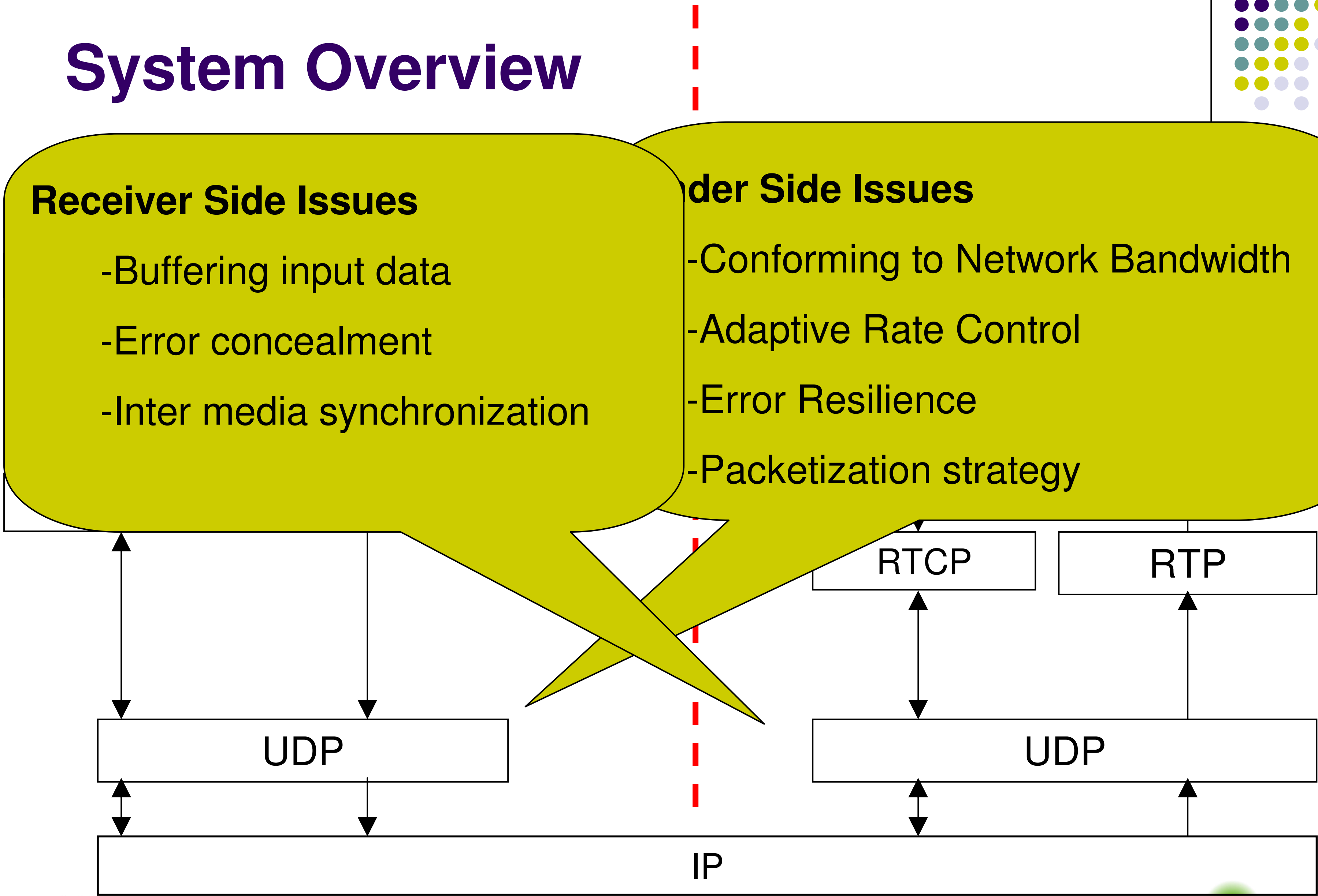
nder Side Issues

-Conforming to Network Bandwidth

-Adaptive Rate Control

-Error Resilience

-Packetization strategy



System Overview : *Sender Side Issues*

- Packetization Strategy
  - Example : fragment and packetize MPEG-4 video

- Must meet network MTU limits
  - Example : MTU = 1500 bytes
  - Packets larger than MTU will result in IP fragmentation
    - Resulting in overhead for each fragmented packet
    - Loss of one fragment will corrupt the whole packet

  - MPEG-4 : Try to allocate a single VOP (frame) per packet
    - If packet is greater than MTU then break into multiple packets
    - Can break any where - arbitrary byte positions
    - Best to break at GOB or slice boundaries

System Overview : *Sender Side Issues*

- Packetization Strategy (continued)

- Minimize number of packets required
  - Less packets means less bits spend on packet headers
  - Example :
    - 20 byte IP Header
    - 8 byte UDP Header
    - 12 byte RTP Header

- Minimize dependency between packets
  - So that even if one packet is lost, the next packet can still be decoded
  - Example :
    - MPEG-4 – use VOP boundaries to start / stop packets

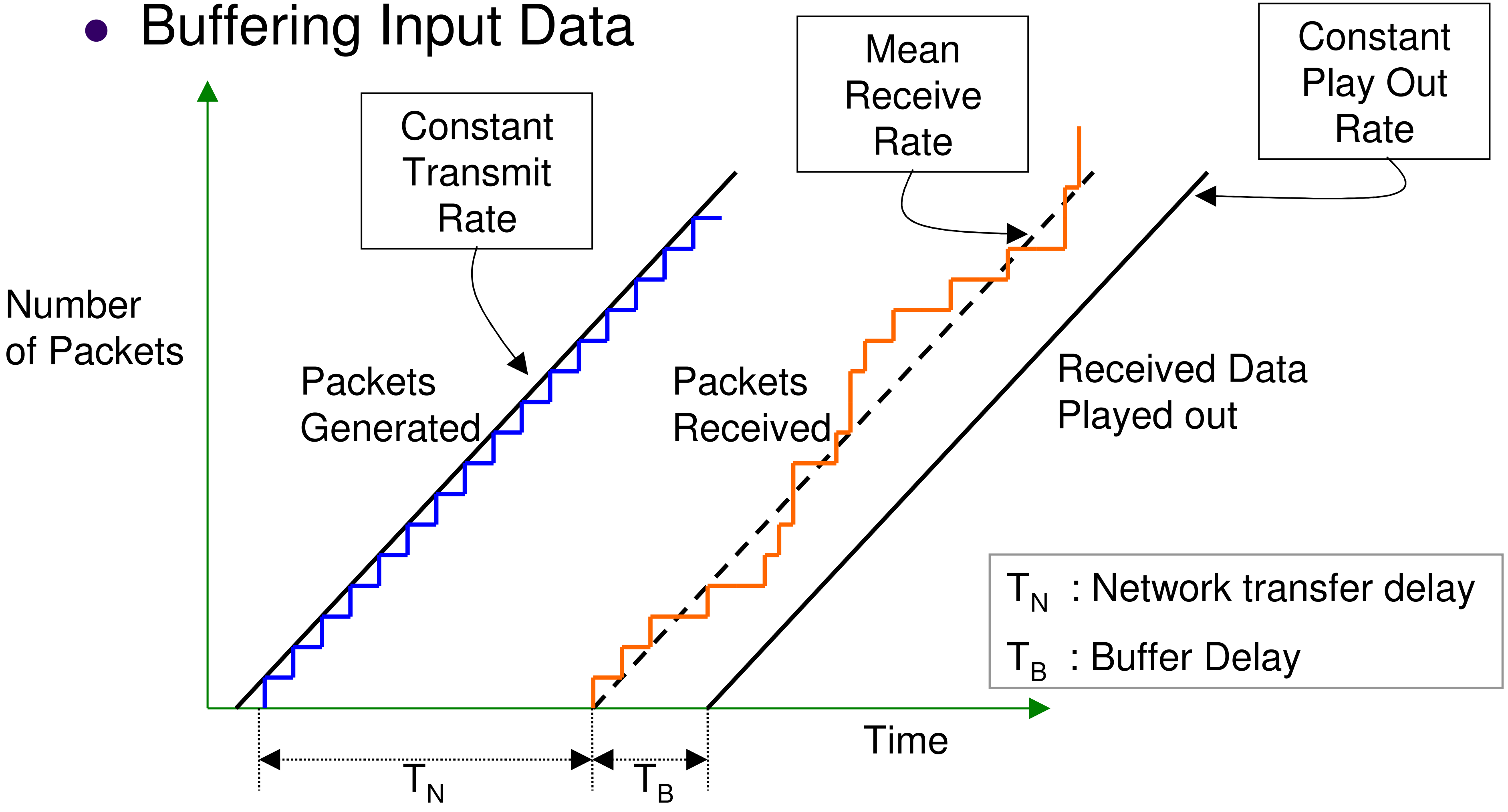
System Overview : *Sender Side Issues*

- Procedure to fragment and packetize live MPEG-4 coded video or stored MPEG-4 video.
  - Goal – try to packetize one Frame per RTP packet

1. Calculate Max allowable size of frame :
   - Max_FrameSize = MTU – Δ
   - Where Δ = RTP_header + UDP_header + IP_header
2. Find size of frame N from the bit stream : FrameSize
3. IF (FrameSize <= Max_FrameSize)
   - THEN include video frame in one RTP packet
4. ELSE
   - Break Frame into multiple packets
   - Best to break at GOB or slice boundaries
   - Example break at a slice boundary to form 2 smaller packets
5. N = N + 1 : go back to Step 2

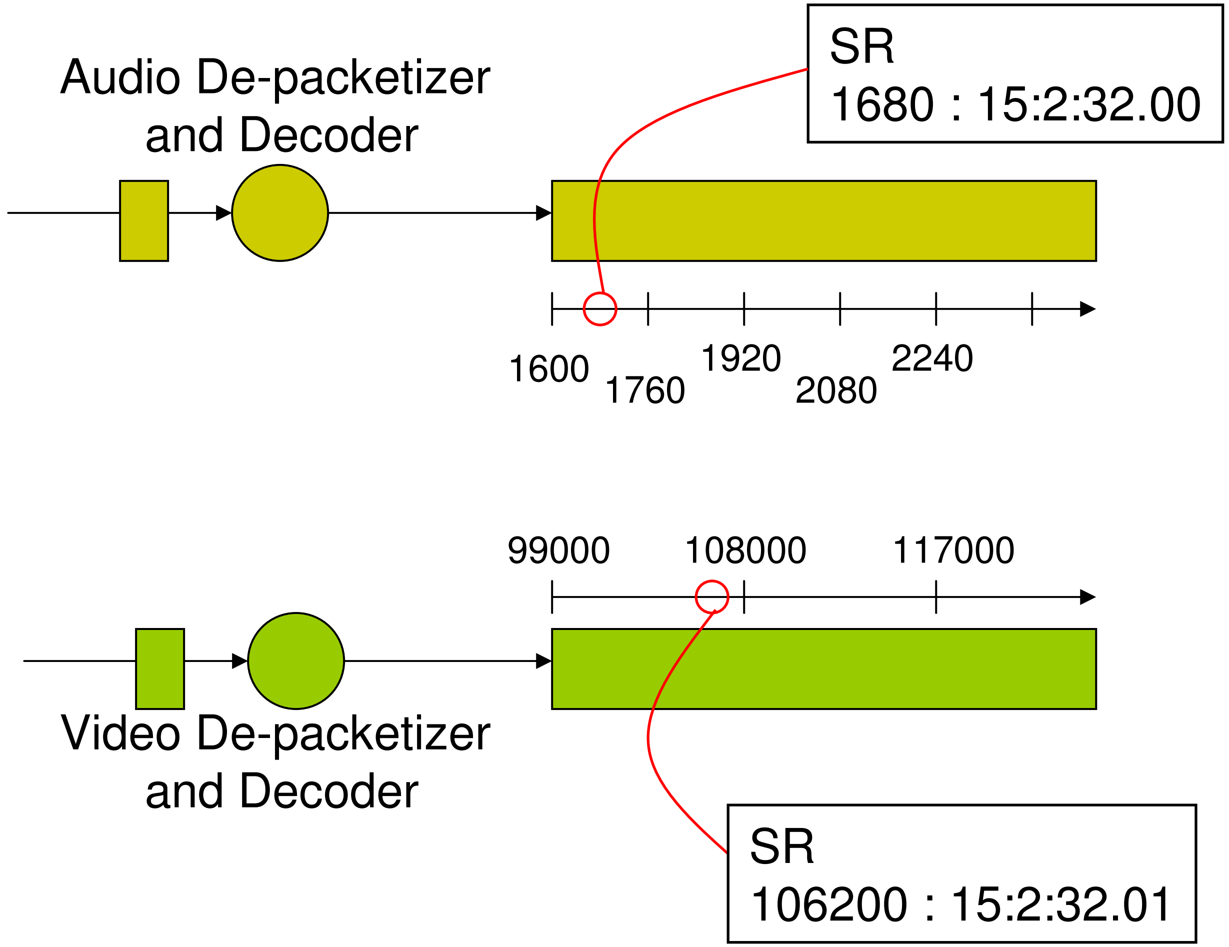
System Overview *: Receiver Side Issues*

- Buffering Input Data



System Overview : Receiver Side Issues

- Inter Media Synchronization



Audio clock = 8000Hz
(eg u-law)

1760 = 32.00 + 0.01
1920 = 32.00 + 0.03

2240 = 32.00 + 0.07

Video clock = 90000Hz
(eg MPEG-4)

108000 = 32.01 + 0.02
117000 = 32.01 + 0.12

Protocols Enabling Streaming

- The story so far,
  - RTP & RTCP
  - SDP
  - RTSP
  - Examples : Video on demand, IP TV

- Next – Session Initiation
  - SIP - Session Initiation Protocol
  - Example : multimedia conference initiation
  - What is role or function of SIP ?

SIP

- Session Initiation Protocol (SIP)
  - Another session control protocol
  - IETF RFC 3261 www.ietf.org/rfc/rfc3261.txt

  - Protocol that can establish, modify and terminate multimedia sessions.
  - Applications – IP telephony, multimedia conferences

  - SIP, like RTSP,
    - Uses text-based request/response transaction model
      - Requests contain Methods and Header fields
      - Responses include 3 digit status codes (eg “200 OK”)

    - Is Transport layer independent
    - Uses other protocols for media delivery (eg RTP).

File Format

- MPEG-1 & MPEG-2 content typically exchanged as files that represent a stream ready to be delivered
  - Embedded absolute time stamps
  - Fragmentation of media for some preferred transport
  - Random Access could be difficult

- MPEG-4 file format : MP4
  - What are the advantages of MP4 compare to stored MPEG-2 compressed video files?

  - What is “hinting” in the context of MP4?
    - What are some advantages of “hinting”, especially for streaming server operation ?

MP4 : Main File Format Concepts

- The Media data is stored separately from Meta data
  - Media data : Audio, Video samples
  - Metadata : Data describing the media

    Examples : timing info,

    number of bytes required for a frame

- Timing information specified by relative numbers (durations) rather than absolute numbers
  - Allows editing to be easier – eg insertion of a new frame

- Able to store media data distributed over several files
  - Use URLs to point to media data stored at various locations

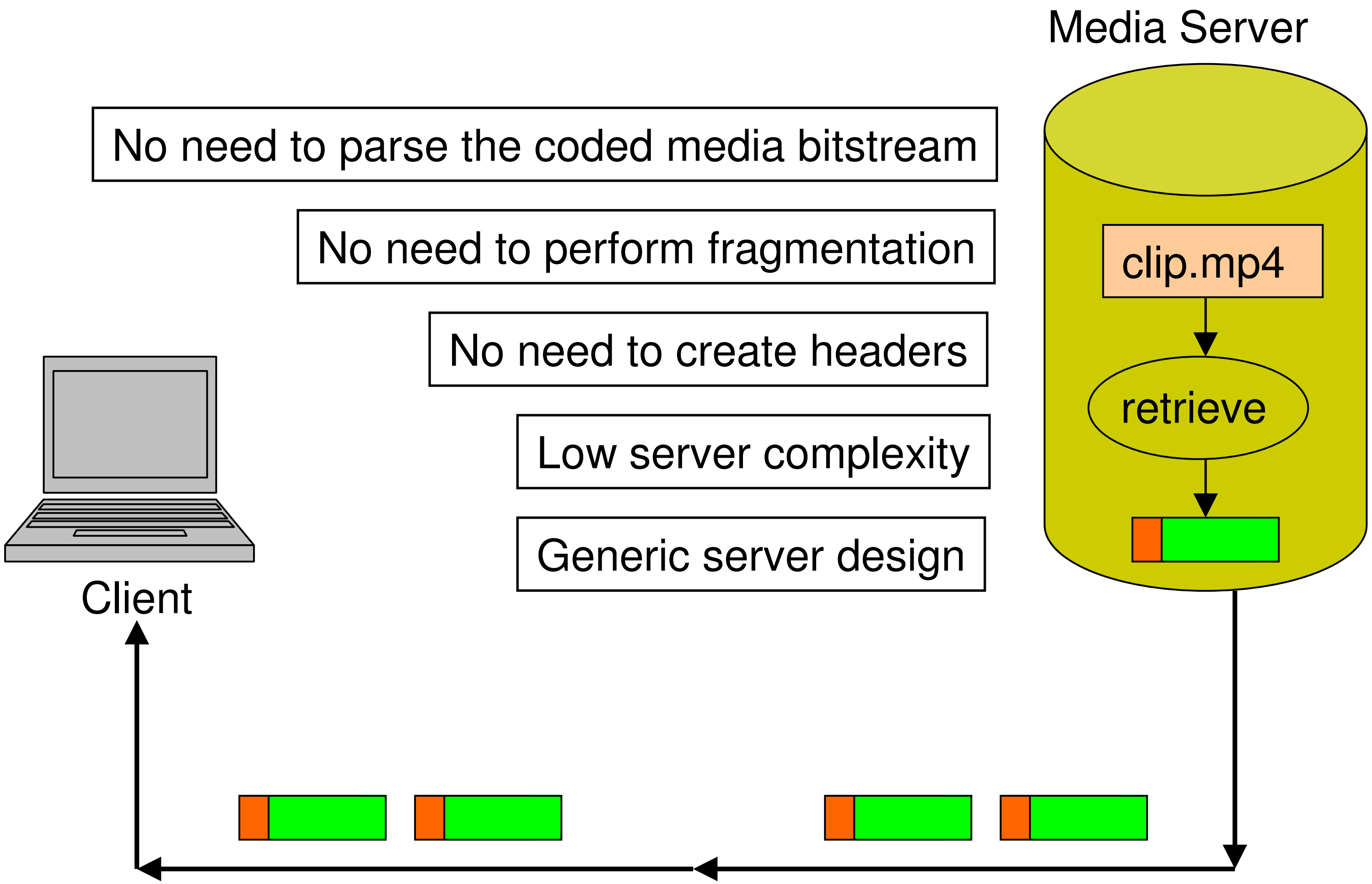
MP4 : Main File Format Concepts

- The Media data is stored separately from Meta data
- Timing information specified by relative numbers (durations) rather than absolute numbers
- Able to store media data distributed over several files

- Locating media data by means of data offsets and length information
  - Metadata tables mapping media sample number to location in a file

- Support streaming protocols through optional hint tracks
  - Metadata information for packetization and header data
  - Example – hints for RTP streaming stored as a separate track

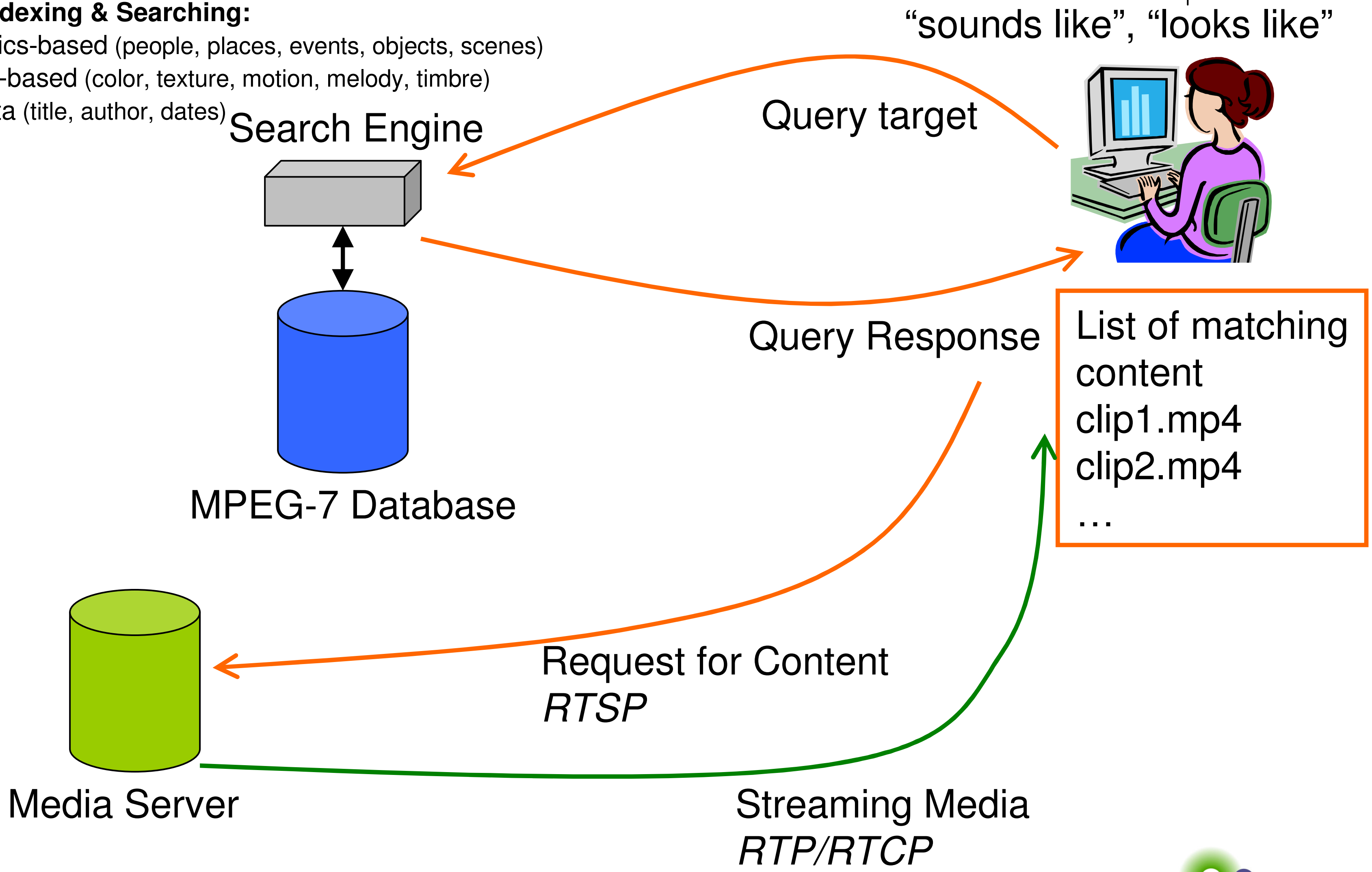
MP4: Hinting



Media Server
No need to parse the coded media bitstream
No need to perform fragmentation
No need to create headers
Low server complexity
Generic server design
clip.mp4
retrieve
Client

MPEG-7 System : Client Server architecture

- **MPEG-7 Indexing & Searching:**
  - Semantics-based (people, places, events, objects, scenes)
  - Content-based (color, texture, motion, melody, timbre)
  - Metadata (title, author, dates)



Overview : Multimedia Information Retrieval

A/Prof. Jian Zhang

NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2009

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY · AUSTRALIA

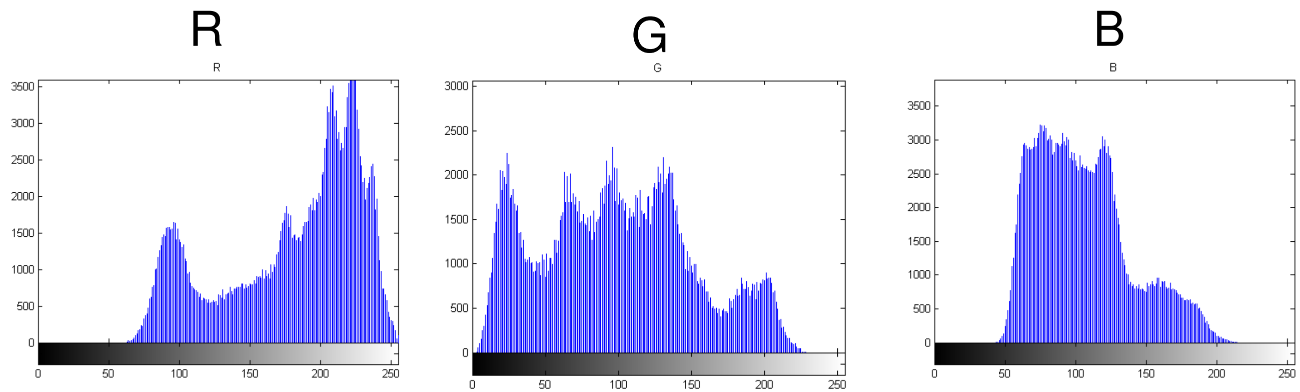
NICTA

Introduction

- The fundamental of Multimedia Database (Content) Management research covers:
  - Feature extraction from these multiple media types to support the information retrieval.
  - Feature dimension reduction – High dimensional features
  - Indexing and retrieval techniques for the feature space
    - Similarity measurement on query features
  - How to integrate various indexing and retrieval techniques for effective retrieval of multimedia documents.
  - Same as DBMS, efficient search is the main performance concern

Low level Feature Extraction -- Color Representation

- Color descriptors
  - Color histogram
    - It characterizes the distributions of colors in an image both globally and locally
    - Each pixel can be described by three color components.
      - A histogram for one component describes the distribution of the number of pixels for that component color in a quantitative level – a quantized color bin.
      - The levels can be 256, 64, 32, 16, 8, 4, 1 (8-bit byte)

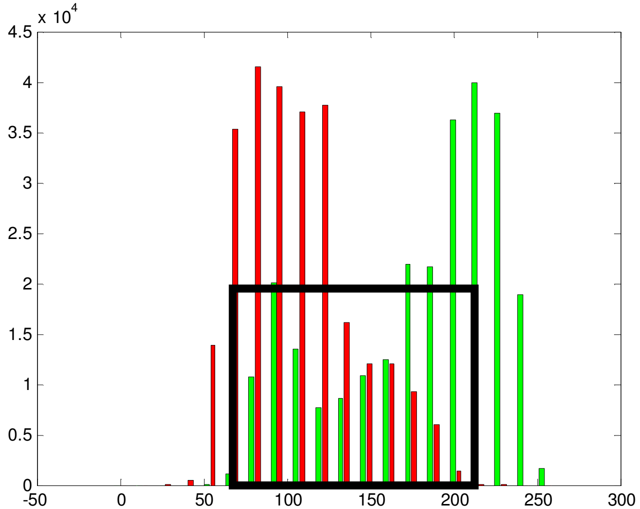


Low level Feature Extraction -- Color Representation

- Color Histogram Intersection

  - Histogram Intersection is employed to measure the similarity between two histograms

$$S(I_p, I_q) = \frac{\sum_{i=1}^{N} min(H_i(I_p), H_i(I_q))}{\sum_{i=1}^{N} H_i(I_q)}$$



Colors that are not present in the query image do not contribute to the intersection distance

Low level Feature Extraction -- Color Representation

- Color Coherence Vector (CCV)

  - A color's coherence is defined as the degree to which pixels of that color are members of large similar-color regions.

  - These significant regions are referred as coherent regions which are observed to be of significant importance in characterizing images

  - Coherence measure classifies pixels as either coherent or incoherent

  - A color coherence vector represents this classification for each color in the image.

Low level Feature Extraction -- Color Representation

- How to compute CCV

τ= 4, R=G=B

22	10	21	22	15	16
24	21	13	20	14	17
23	17	38	23	17	16
25	25	22	14	15	14
27	22	12	11	17	18
24	21	10	12	15	19

Blurred Image

1: 10-19
2: 20-29
3: 30-39

2	1	2	2	1	1
2	2	1	2	1	1
2	1	3	2	1	1
2	2	2	1	1	1
2	2	1	1	1	1
2	2	1	1	1	1

Discretized Image

B	C	B	B	A	A
B	B	C	B	A	A
B	C	D	B	A	A
B	B	B	A	A	A
B	B	A	A	A	A
B	B	A	A	A	A

Connected Components

Label	A	B	C	D
Color	1	2	1	3
Size	17	15	3	1

Connected Table

Color	1	2	3
α	17	15	0
β	3	0	1

Color Coherent Vector

Comparison

8.4 Color-based Image Indexing and Retrieval Techniques

- Similarity among colors
  - The limitation of using L-1 metric distance is that the similarity between different colors or bins is ignored (Cont.).
    - In the simple histogram measure, it might not be able to retrieve perceptually similar images due to these changes
  - Contributions of perceptually similar colors in the similarity calculation
    - Image distance and similarity have an inverse relationship.
    - The similar color measurement is a way to go !

Color-based Image Indexing and Retrieval Techniques

- Simple histogram distance measure
  - The distance between the histogram of the query image and images in the database are measured
    - Image with a histogram distance smaller than a predefined threshold are retrieved from the database
    - The simplest distance between images I and H is the L-1 metric distance as

$$D(I,H) = \text{sum } |I-H|$$

    - More popular way is to calculate the L-2 metric distances as

$$D(I,H) = \text{Sqrt } (\text{sum } (I-H|)^2)$$

Color-based Image Indexing and Retrieval Techniques

- Example 2 – Niblack's similarity measurement

X – the query histogram; Y – the histogram of an image in the database
Z – the bin-to-bin similarity histogram. Z is Transpose matrix = {|I-H|}

The Similarity between X and Y ➔, $\| Z \| = Z^t A Z$

Where A is a symmetric color similarity matrix with $a(i,j) = 1 - d(c_i,c_j)/d_{max}$

c_i *and* c_j *are the ith and jth color bins in the color histogram*

$d(c_i,c_j)$ *is the color distance in the mathematical transform to Munsell color space and* d_{max} *is the maximum distance between any two colors in the color space.*

- The similarity matrix A accounts for the perceptual similarity between different pairs of colors.

8.4 Color-based Image Indexing and Retrieval Techniques

- Cumulative histogram distance measure

  - Instead of bin-to-bin distance without considering color similarity, a cumulative histogram of image M is defined in terms of the color histogram H(M):

$$Ch_i = \sum_{j<=i} h_j$$

The cumulative histogram vector matrix $CH(M)=(Ch_1,Ch_2....Ch_n)$

  - The drawback of this approach is that the cumulative histogram values may not reflect the perceptual color similarity

Similarity Measurement

Consider two images A and B, and histogram similarity matrix C

A

1	2	1	3	3
1	2	1	2	3
1	1	1	2	2
1	3	1	3	2
1	3	1	2	2

B

1	2	1	2	1
2	2	1	2	1
3	1	1	1	1
3	3	3	3	3
2	2	2	3	3

C

1	0.5	0
0.5	1	0.5
0	0.5	1

- For each image, calculate histogram, cumulative histogram and CCV
- For the two images, calculate L1 (also consider L2) histogram distance, L1 cumulative histogram distance, histogram intersection, Normalized CCV distance and Niblack's histogram similarity value
- Suppose that the threshold for the size of the connected component is 3.

Similarity Measurement – Histogram

Bin		1	2	3
Histogram		11	8	6
Cumulative Histogram		11	19	25
CCV	α	11	6	3
	β	0	2	3
Image A				

Bin		1	2	3
Histogram		9	8	8
Cumulative Histogram		9	17	25
CCV	α	8	6	8
	β	1	2	0
Image B				

Similarity Measurement – Histogram

- L1 Histogram Distance

$$D = |11-9| + |8-8| + |6-8| = 4$$

- L2 Histogram Distance

$$D = \sqrt{(11-9)^2 + (8-8)^2 + (6-8)^2} = 2\sqrt{2} = 2.82$$

- L1 Cumulative Histogram Distance

D = |11-9| + |19-17| + |25-25| = 4

- Histogram Intersection

D = [min(11,9) + min(8,8) + min(6,8)] / (11 + 8 + 6)
= [9 + 8 + 6] / 25
= 23 / 25
= 0.92

Similarity Measurement – Histogram

- Normalized CCV

D = |(11-8)/(11+8+1)| + |(0-1)/(0+1+1)|
+ |(6-6)/(6+6+1)| + |(2-2)/(2+2+1)|
+ |(3-8)/(3+8+1)| + |(3-3)/(3+3+1)|
= 3/20 + 1/2 +5/12
= 1.1817

- Niblack's similarity measure

Transpose(Z) = [|11-9|, |8-8|, |6-8|] = [2, 0, 2]

$$D = Z^T C Z = [2 \quad 0 \quad 2] \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = 8$$

Image Retrieval based on Texture

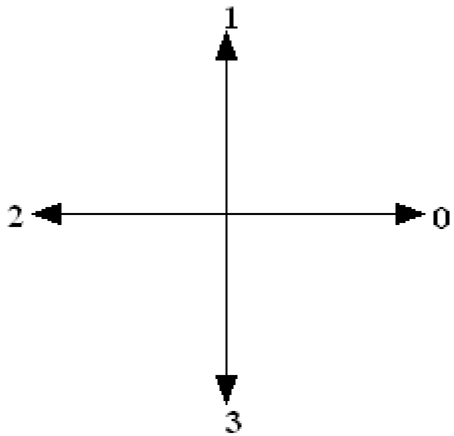
- The second-order statistics take into account the relationship between the pixel and its neighbors
  - The Grey-level Co-occurrence Matrix (GLCM) is used to calculate the second-order statistics.
  - Suppose the following 4x4 pixel image with 3 distinct grey-levels:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

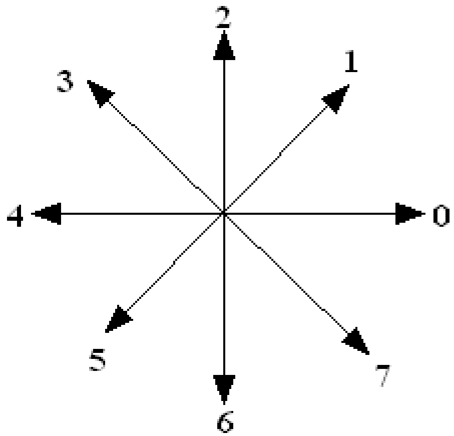
  - And d = (dx, dy) = (1,0) means that compute the co-occurrences of the pixels to the left of the current one.

Image Indexing and Retrieval based on Shape

- Boundary-based methods -- Chain Code
  - Chain codes are used to represent a boundary by a connected sequence of straight-line segments of special length and direction
  - Typically, this representation is based on 4- or 8-connectivity of the segments. The direction of each segment is coded by using a numbering scheme



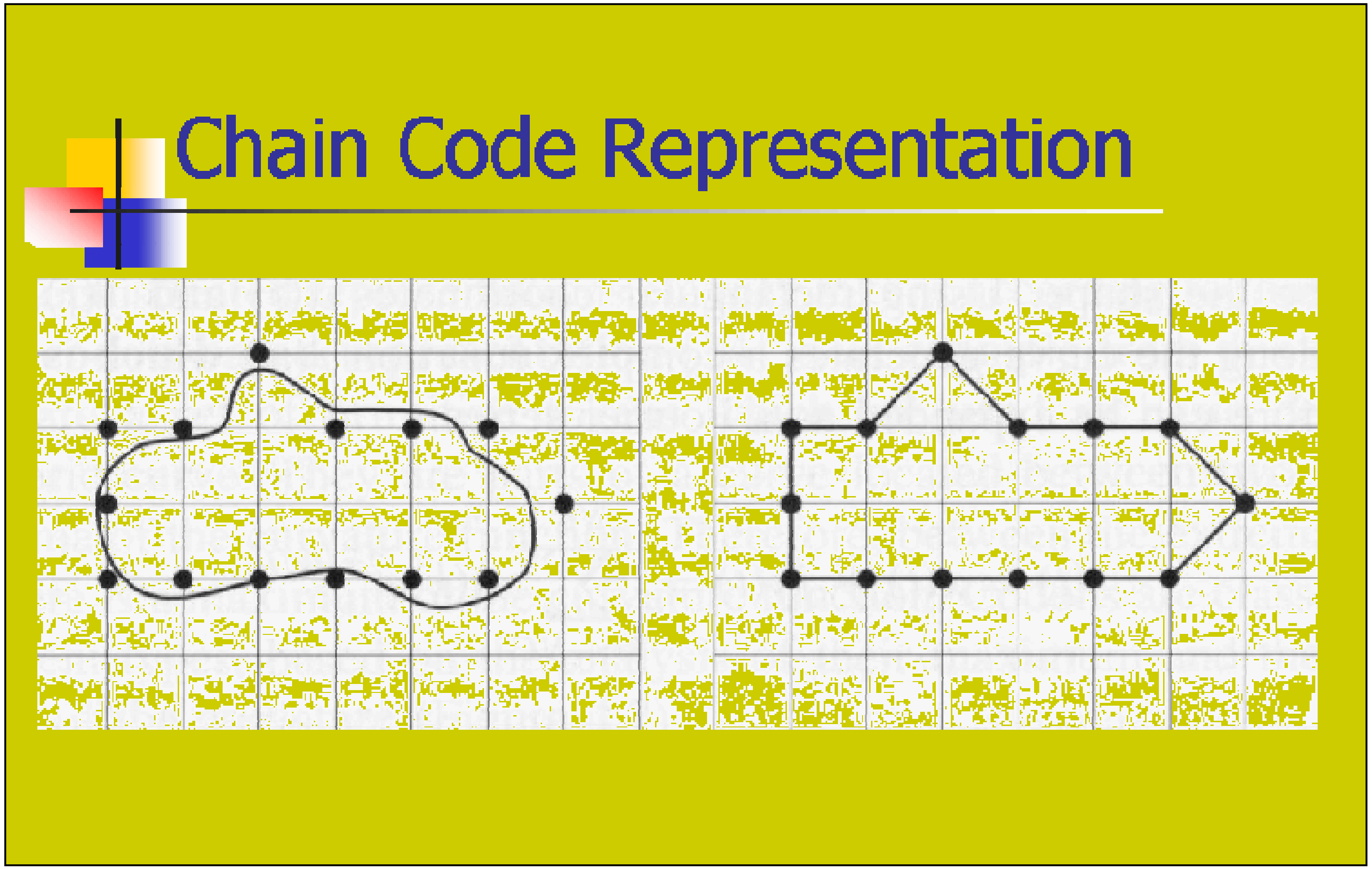
Direction numbers for 4-directional chain code



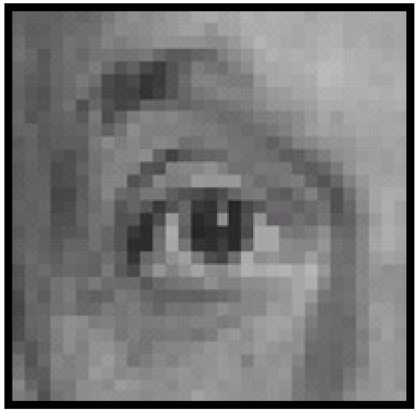
Direction numbers for 8-directional chain code

Image Indexing and Retrieval based on Shape

- Boundary-based methods -- Chain Code

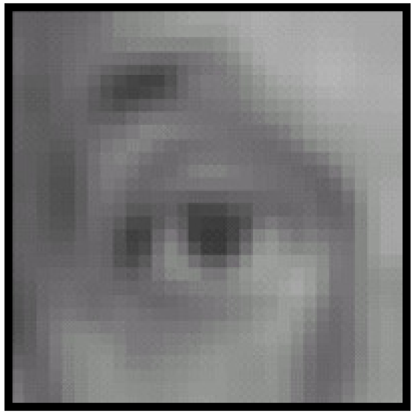


Linear Image Filters and Convolution



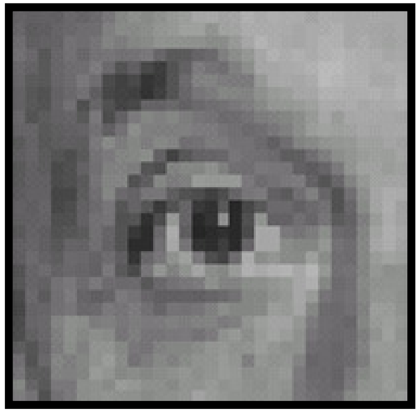
Original

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$



Blur (with a box filter)

Linear Image Filters and Convolution



Original

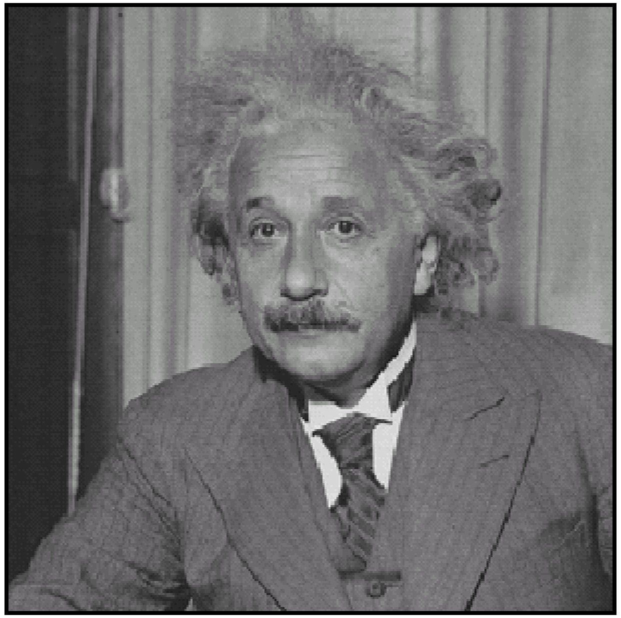
$$\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right)$$

Sharpening filter

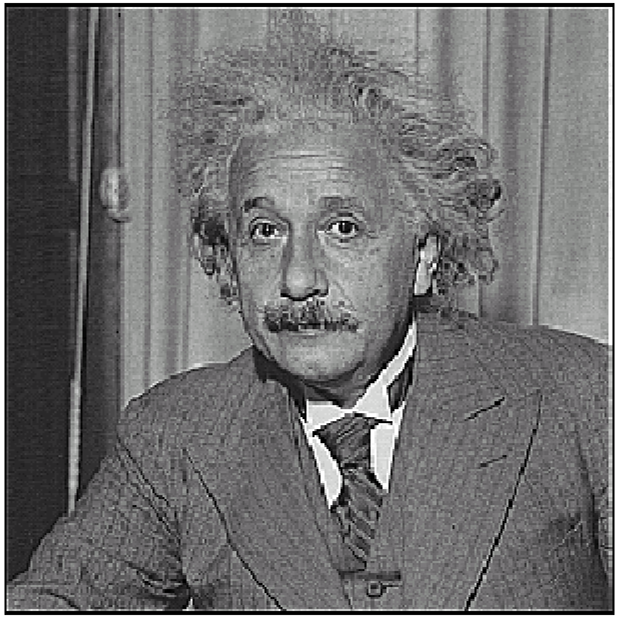
- Accentuates differences with local average
- Also known as Laplacian

Linear Image Filters and Convolution

Sharpening



before



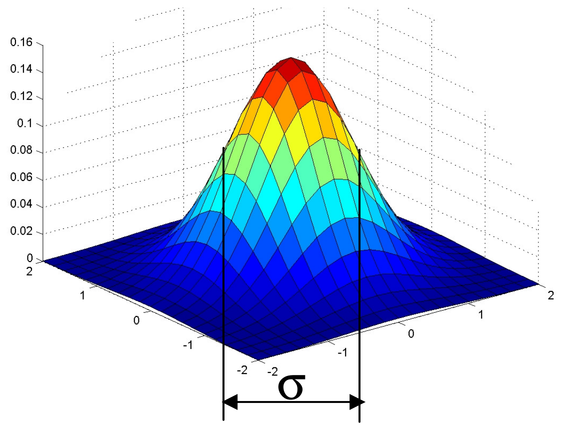
after

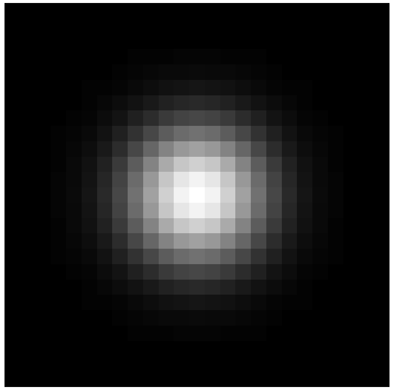
Linear Image Filters and Convolution

- Gaussian Kernel

Slide credit: Christopher Rasmussen

  - Idea: Weight contributions of neighboring pixels by nearness





5 x 5, σ = 1

0.003	0.013	0.022	0.013	0.003
0.013	0.059	0.097	0.059	0.013
0.022	0.097	0.159	0.097	0.022
0.013	0.059	0.097	0.059	0.013
0.003	0.013	0.022	0.013	0.003

$$G_\sigma = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

  - Constant factor at front makes volume sum to 1.

Linear Image Filters and Convolution

- Gaussian Vs Average

Gaussian Smoothing

Smoothing by Averaging

Smoothing with a Gaussian

- Advantage: the noise or the nature of the object observed might be of a Gaussian probable form
- In most cases on a image, a single point of light viewed in a de-focussed lens looks like a fuzzy blob
- A Gaussian gives a good model of a fuzzy blob
- Filter weights decrease monotonically from central peak, giving most weight to central pixels

Consultation time

- A in-lecture consultation time will be 22 October 2009 at this lecture room (6-8 PM)
- A special consultation time will be 13 Nov. 2009. The special consultation (2-4 PM) at Level 4, L5 Building. A notice will be sent from comp9519@cse.unsw.edu.au
- Please review the tutorial notes (related parts as well)
- Good Luck for your final Exam.