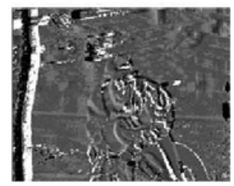
Lecture 5
VIDEO CODING STANDARDS

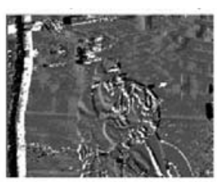
- Interframe compression methods
- Overview of standards
- ITU-T standards
- ISO standards

Introduction

- Simplest way ~ compress each frame image individually
  - e.g., “motion-JPEG”
  - Only spatial redundancy is explored and reduced
- How about temporal redundancy?
  - Pixel-by-pixel difference could still be large

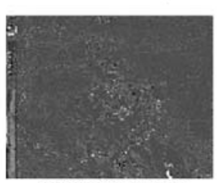


“Horse ride”



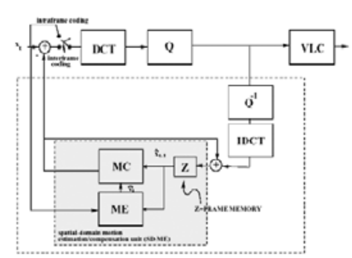
Pixel-wise difference w/o motion compensation

Motion estimation



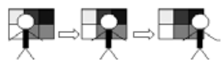
Residue after motion compensation

Hybrid MC-DCT encoder



Problems to be solved

- Not all regions are easily inferable from previous frame
  - Occlusion ~ solvable by backward prediction using future frames as ref.
  - Adaptively deciding using prediction or not



- Drifting and error propagation
  - Solution: Encode reference regions or frames from time to time
- Random access
  - Solution: Encode frame without prediction from time to time
- How to allocate bits?
  - According to statistics
  - Consider constant or variable bit-rate requirement
    - *Constant-bit-rate (CER) vs. Variable-bit-rate (VER)*

Importance of standards

- Multimedia technologies are used in many audiovisual applications for which interoperability is a major requirement.
- The interoperability requirement is solved by specifying standard algorithms.
- To allow evolution and competition, standards shall provide interoperability by specifying the minimum possible set of elements, for example only the bitstream syntax and the decoder (*not the encoder*).

Standards are also repositories of the best technology and thus an excellent place to check the technological evolution and trends !

Success of standards depends on much more than technology !

Standardization (1/2)

- Two major international organizations have been responsible for providing standards on audio-visual coding
  - Inernational Organization for Standardization (ISO)
  - International Telecommunication Union (ITU)

Standardization (2/2)

ITU-T H.261	Video for ISDN applications @ p × 64 kbps (p = 1,…, 30) (1990)
ITU-T H.263	Video for PSTN applications @ less than 64 kbps (1995)
ISO MPEG-1	Video for optical storage media @ 1.2 Mbps (1993)
ISO MPEG-2	High quality generic video @ 4-20 Mbps (1995) includes digital SDTV, HDTV, and DVD
ISO MPEG-4	Object-based video @ 10 kbps – 2 Mbps (1998)

Video formats

	ITU-R 601 525/60	*ITU-R 601 625/50*	*CIF*	*QCIF*	*SIF525*	*SIF625*
No of active pels/line						
Lum (Y)	720	720	352	176	360	360
Chroma (U,V)	360	360	176	88	180	180
No of active lines						
Lum (Y)	480	576	288	144	240	288
Chroma (U,V)	480	576	144	72	120	144
Interlace	2:1	2:1	1:1	1:1	1:1	1:1
Temporal rate	60	50	15-30	5-30	30	25
Aspect ratio	4:3	4:3	4:3	4:3	4:3	4:3
Raw data (Mbps)	165.9	165.9	37.3	9.35		

Standards	Application	Video Format	Raw Data Rate	Compressed Data Rate
H.320 (H.261)	Video conferencing over ISDN	CIF QCIF	37 Mbps 9.1 Mbps	>=384 Kbps >=64 Kbps
H.323 (H.263)	Video conferencing over Internet	4CIF/ CIF/ QCIF		>=64 Kbps
H.324 (H.263)	Video over phone lines/ wireless	QCIF	9.1 Mbps	>=18 Kbps
MPEG-1	Video distribution on CD/ WWW	CIF	30 Mbps	1.5 Mbps
MPEG-2	Video distribution on DVD / digital TV	CCIR601 4:2:0	128 Mbps	3-10 Mbps
MPEG-4	Multimedia distribution over Inter/Intra net	QCIF/CIF		28-1024 Kbps
GA-HDTV	HDTV broadcasting	SMPTE296/295	<=700 Mbps	18--45 Mbps
MPEG-7	Multimedia databases (content description and retrieval)			

What is MPEG?

- Moving Picture Experts Group
- ISO/IEC JTC1/SC29/WG11 (ISO/IEC Joint Technical Committee 1, Sub Committee 29, Work Group 11)
- Deals with the Representation of Moving Video and Audio
  - **and 'supporting' technologies, e.g. MPEG-2 Systems, DSM-CC, MPEG-4 Scene Description, MPEG-7 DDL, MPEG IPMP,…**
  - **3-5 one week meetings for year, with 300 - 400 experts (~150 companies and research inst)**
    - USA: AT&T, TI, Motorola, Microsoft, IBM, DEC, ...
    - Asia: Sony, JVC, Mitsubishi, Matsushita, Daewoo, Samsung...
    - Europe: Philips, Thomson, Alcatel, Siemens, Bosch, Deutsche Telekom, BT, CNET...

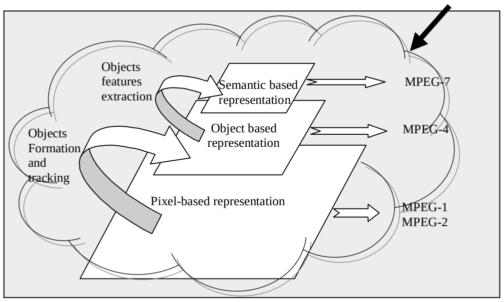
MPEG standards

- MPEG-1 (Nov.1992), MPEG-2 (Nov.1994) standards
  - **Widely adopted in multimedia industry**
  - **Digital TV, CD-i, Video-on-Demand, Archiving, Music on the Internet (MP3)**
- MPEG-4 (Oct. 1998) standard
  - **Object-based audiovisual coding**
- MPEG-7 (July 2001) standard
  - **Object-based audiovisual description**
- Under development: MPEG-21 standard
  - ***"Multimedia Framework"***
- Under development: MPEG-A standard
  - ***"Multimedia Application Formats"***

Relation between MPEG standards

- Data representation pyramid

MPEG21



What is ITU-T?

- Telecommunication standardization sector of ITU (Formerly CCITT)
- Study group 15 is responsible for video conference and video telephony
- Over 20 companies and research centers around the world actively participate
- Meets 3-5 times every year
- Proposes « Recommendations »

ITU-T Recommendations

- H.120 (1984)
  - Video-conferencing over H1 channels (1.5-2 Mb/s)
- H.261 (1988)
  - Video-conferencing over ISDN lines (p*64 Kb/s; p1…30)
- H.263 (1996)
  - Video-telephony over POTS and Internet (8-64 Kb/s)

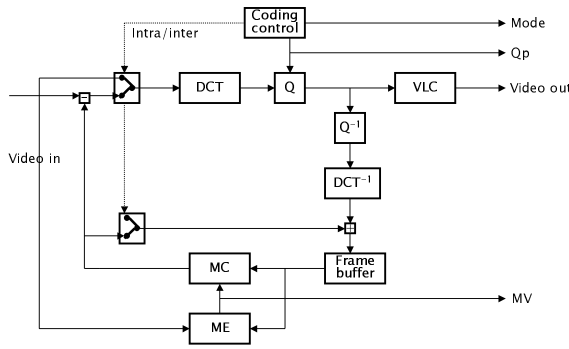
Video Compression Basics

- Spectral redundancy reduction
- Spatial redundancy reduction (compression)
  - transform coding
- Temporal redundancy reduction (compression)
  - motion estimation/compensation
  - Predictive coding
- Entropy coding
  - from symbols to bits through statistics
- Bitstream syntax (specific vocabulary…)

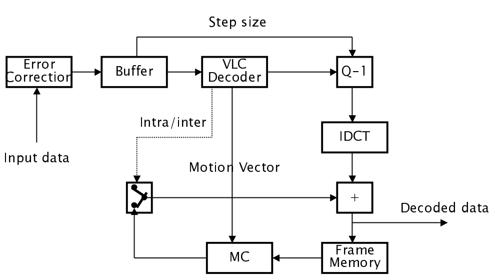
ITU-T recommendation H.261 (p x 64)

- Developed for transmission of video at a rate of multiples of 64Kbps
- Only the decoder is standardized along with the bitstream syntax
- The video data is split into Intra and Inter frames
- Intra frames are coded with a JPEG-like DCT based approach
- For inter frames a decşsşon has to made for each MB (Inter/Intra/MC/No MC)
- Both spatial and temporal coding
- Handles video of size CIF and QCIF
- Average frame rate 5-15 f/s

H.261 coder structure (encoder)

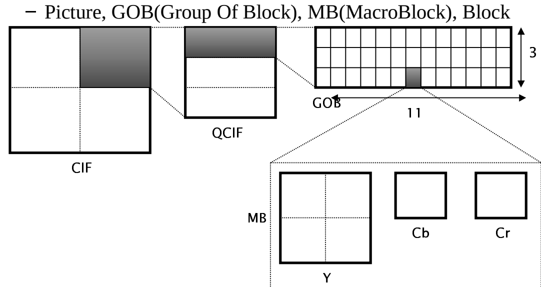


H.261 coder structure (decoder)

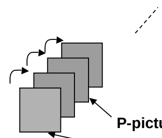


H.261 frame format

- 4 layers in the compressed stream
  - Picture, GOB(Group Of Block), MB(MacroBlock), Block



- INTRA I-frames

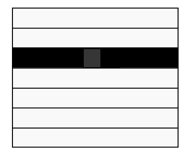


- Predicted P-frames:
  - backward predicted from previous anchor picture (I or P)

H261 Syntax

A sequence is composed of

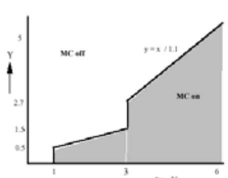
- Frames structured as
  - ...IPPPP…I...
  - Frame
    - *Group of Blocks (Basic resync point)*
      - *Macroblock (motion)*
        - *Block (DCT)*



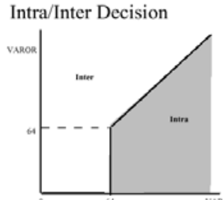
Mode decision

- MC/No MC Decision

$$X = \frac{\sum |bd|}{256} \; ; \quad Y = \frac{\sum |dbd|}{256}$$



- Intra/Inter Decision

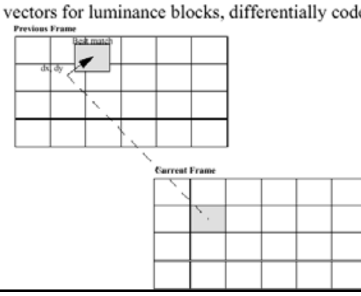


$$VAR = (\sum (pel - mc\_pel)^2) / 256$$

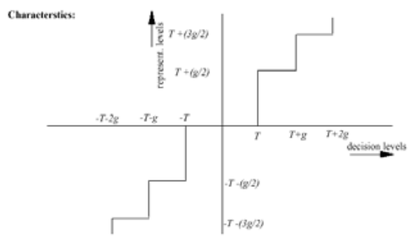
$$VAROR = (\sum (pel - blk\_avg)^2) / 256$$

Motion estimation and compensation

- Motion Estimation/Compensation on 16x16 Luminance blocks
- Search with integer pel accuracy with range up to +- 7 pels
- Motion vectors for luminance blocks, differentially coded



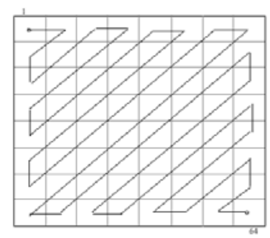
Quantization of DCT coefficients



- The intra DC coefficient is quantized with a stepsize of 8 and no dead-zone.
- Other coefficients are quantized with a central dead-zone *T* about zero and a stepsize *g*, an even value in the range 2 to 62 (31 stepsizes are allowed).

Scanning of DCT coefficients

- Convert 2D coefficient block to 1D coefficient block
- Scan order frequency dependent for efficient VLC
- Zig-zag scan of 8x8 DCT coefficient blocks



ITU-T Recommendation H.263

- Intended for low bitrate videotelephony, started in 1994, completed 1995
- Builds on H.261; adds tools to improve coding efficiency
- Standard specifies bitstream syntax and decoder
- 1995-97: Extension of H.263 to H.263+ video

New features of H.263 (1/4)

- Bit rate: The target bit rate is up to 64Kbps
  - Generally it aims good quality at 28.8Kbps
  - Note that H.261, the target bit rate is 64 ~ 1920Kbps
- Picture format - Three new format : sub-QCIF, 4CIF, 16CIF

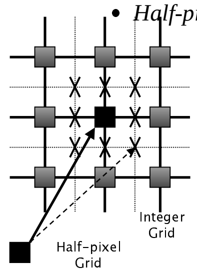
Format	Image size	
	Y(pixels * line)	Cb and Cr(pixels * line)
Sub-QCIF	128 * 96	64 * 48
QCIF	176 * 144	88 * 72
CIF	352 * 288	176 * 144
4CIF	704 * 576	352 * 288
16CIF	1408 * 1152	704 * 576

New features of H.263 (2/4)

- Group of blocks structure
  - Picture, Group of block(GOB), MacroBlock, Block
  - Each GOB contains one MB row
    - Ex.
      - For QCIF, each GOB has 11MBs (33MBs in H.261)
- Precision of motion compensation
  - *Half-pixel accurate motion estimation*
    - Improve motion prediction significantly
    - Integer-pixel accuracy is used in H.261

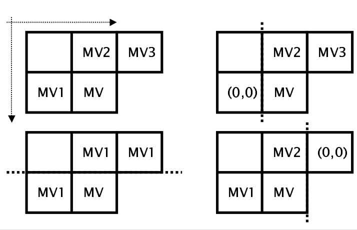
New features of H.263 (3/4)

- *Half-pixel accurate motion estimation*
  - ◆ Don't restrict to integer pixel grid
  - ◆ More finer resolution
  - ◆ Use interpolation to calculate half-pixel value



New features of H.263 (4/4)

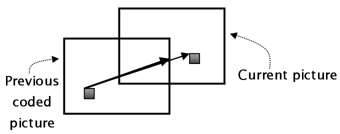
- *Improved motion vector scheme*
  - $MV = MV_d + MV_p$ (MV_p : the predictor)
  - MV_p = median(MV1, MV2, MV3)



H.263 coding options (1/8)

- 4 advanced coding options
  - Unrestricted motion vectors mode (UMV-mode)
  - Advanced prediction mode (AP-mode)
  - PB-frames mode
  - Syntax-based arithmetic coding mode (SAC-mode)

H.263 coding options (2/8)

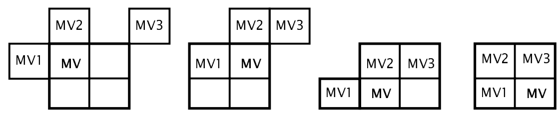


❖ the closest edge pixel is used instead

- Unrestricted motion vectors mode
  - Motion vectors over picture boundaries
  - Extension of the motion vector range
    - [-16, 15,5] → [-31, 31]

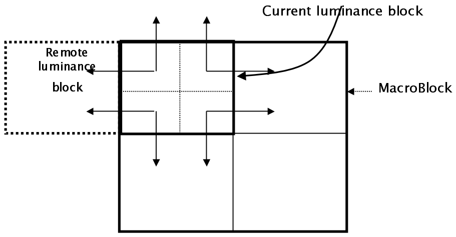
H.263 coding options (3/8)

- Advanced prediction mode
  - 4 motion vectors per macroblock
    - By default, each MB has just one motion vector
    - In this option, each block in one MB has one motion vector
    - Three MVs are used for predictor



H.263 coding options (4/8)

- Advanced prediction mode (cont'd)
  - Overlapped block motion compensation mode

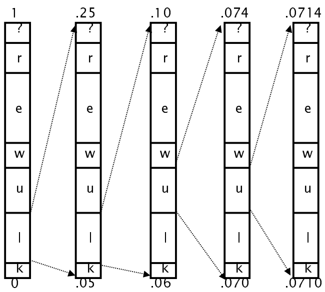


H.263 coding options (5/8)

- Advanced prediction mode (cont'd)
  - Overlapped block motion compensation mode
    - each pixel in an luminance prediction block is a weighted sum of three predictors
    - p(x,y) = ((q(x,y) * H0(I, j) + r(x,y) * H1(I, j) + s(x, y) * H2(I, j) + 4) / 8
      - p(x, y) : Each pixel in a luminance prediction block
      - q(x, y), r(x, y), s(x, y) : pixels from the referenced picture obtained from three MV
      - H0, H1, H2 : weight factor matrix

H.263 coding options (6/8)

- Syntax-based arithmetic coding mode

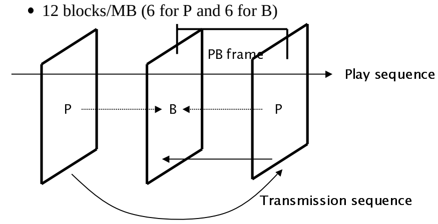


Si	Pi	Subinterval
k	0.05	[0.00, 0.05)
l	0.20	[0.05, 0.25)
u	0.10	[0.25, 0.35)
w	0.05	[0.35, 0.40)
e	0.30	[0.40, 0.70)
r	0.20	[0.70, 0.90)
?	0.10	[0.90, 1.00)

Arithmetic coding for string “lluu”

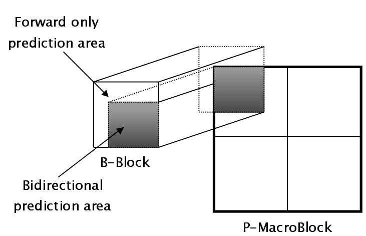
H.263 coding options (7/8)

- PB-Frames mode
  - PB-Frame
    - Consists of P & B pictures interleaved as one unit
    - B picture : Bidirectionally predicted picture
    - 12 blocks/MB (6 for P and 6 for B)



H.263 coding options (8/8)

- PB-Frames mode (cont’d)
  - Only parts of a B-block are bidirectionally predi

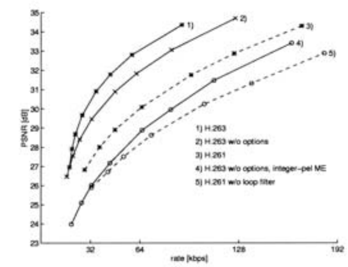


Performance comparison with H.261 (1/4)

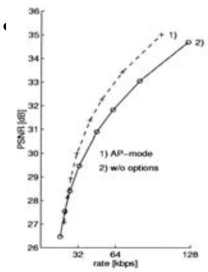
- Simulation setup
  - H.263 codec : TMN-Test model for H.263
    - Telenor Research, Norway, 1995
  - H.261 codec : H.261 software codec
    - Portable Video Research Group, Stanford, 1995
  - Performance metric
    - PSNR(Peek Signal-to-Noise Ratio)

Performance comparison with H.261 (2/4)

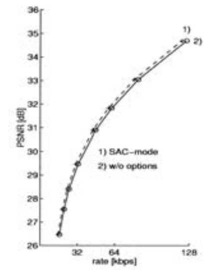
- Ov



Performance comparison with H.261 (3/4)



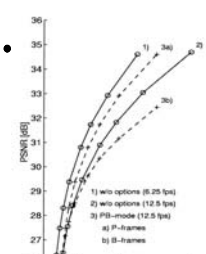
Performance of
Advanced Prediction mode



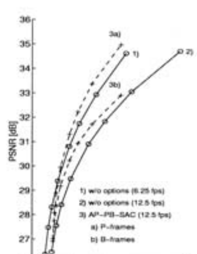
Performance of
SAC mode

Performance comparison with H.261 (4/4)

-)



Performance of
PB-Frames mode



Performance of
AP-PB-SAC mode

Summary

- H.263
  - Video coding mechanism for low-bit-rate
  - Used in video/teleconferencing
  - Performance is generally better than H.261
    - Half-pixel motion estimation
    - Improved motion vector scheme
    - 4 advanced coding options
  - Complexity is the problem
    - H/W approach (VLSI, ASIC, etc)

The H. 263+ Standard: Optional Modes

- **Unrestricted Motion Vector (UMV) mode, Annex D: MV range extended to ±256 depending on the picture size, reversible VLCs used for MVs**
- **Advanced Intra Coding (AIC) mode, Annex I: Inter block prediction from neighboring intra coded blocks, modified quantization, optimized VLCs**

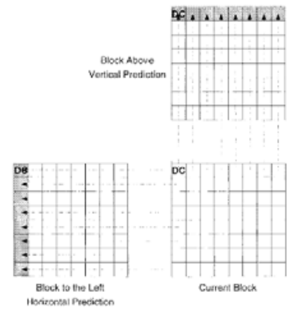


Fig. 6. Neighboring blocks used for intra prediction in the advanced intra coding mode.

•**Deblocking Filter (DF) mode, Annex J: Deblocking filter inside the coding loop.**

•**Reference Picture Selection (RPS) mode, Annex N: Reference picture selected for prediction to suppress temporal error propagation**

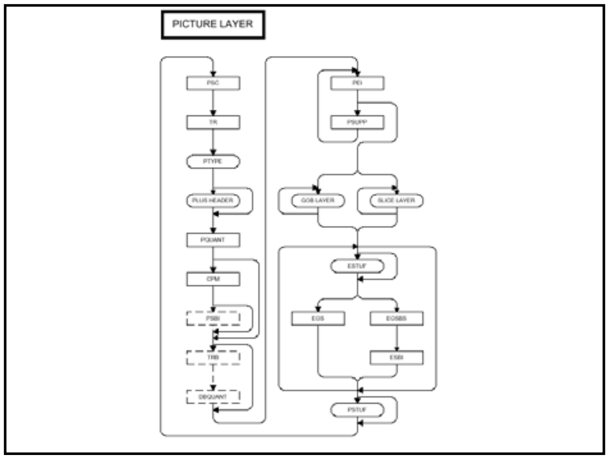
•**Modified Quantization (MQ) mode, Annex T: Quantizer allowed to change at the macroblock layer, finer chrominance quantization employed, range of representable quantized DCT coefficients extended to [- 127, +127]**

•**Slice Structured (SS) mode, Annex K: Slices used instead of GOBs**

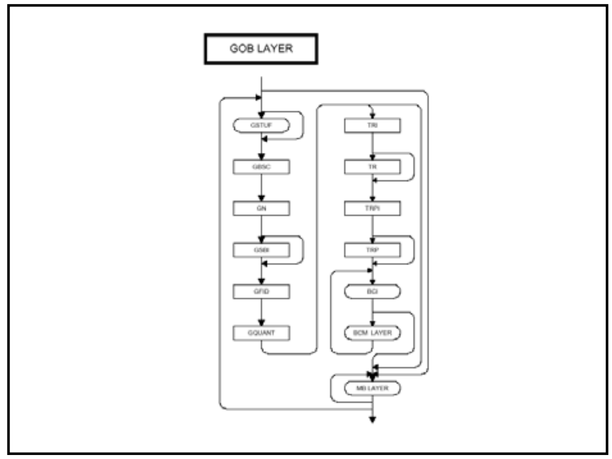
•**Temporal, SNR, and Spatial Scalability mode, Annex O: Syntax to support temporal, SNR, and spatial scalability**

•**Alternative Inter VLC (AIV) mode, Annex S: The intra VLC table designed for encoding quantized intra DCT coefficients in the AIC mode used for inter coding**

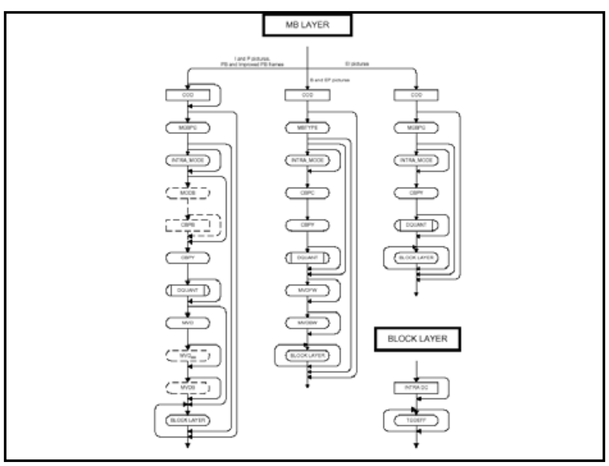
•**Improved PB- frames (IBP) mode, Annex M: Forward, backward and bi- directional prediction supported, delta vector not transmitted**



PICTURE LAYER
PSC
TR
PTYPE
PLUS HEADER
PQUANT
CPM
PSBI
TRB
DBQUANT
PEI
PSUPP
GOB LAYER
SLICE LAYER
ESTUF
EOS
EOSBS
ESBI
PSTUF



GOB LAYER
GSTUF
GBSC
GN
GSBI
GFID
GQUANT
TRI
TR
TRPI
TRP
BCI
BCM LAYER
MB LAYER



MB LAYER
COD
MCBPC
INTRA_MODE
MODB
CBPB
CBPY
DQUANT
MVD
BLOCK LAYER
MBTYPE
CBPC
MVDFW
MVDBW
BLOCK LAYER
INTRA DC
TCOEFF