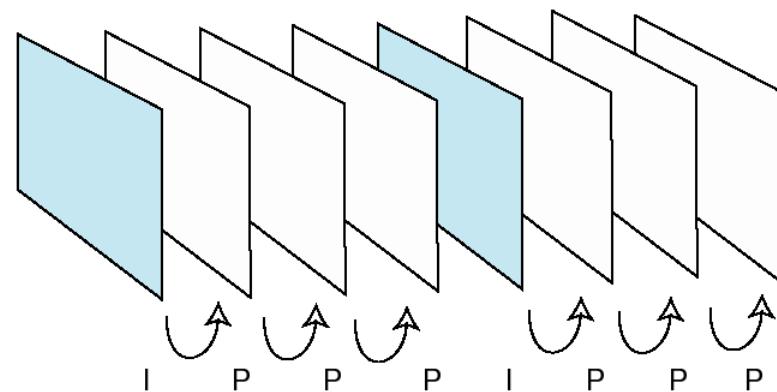


## Standards para compressão de video

- H.261 (ITU antigo CCITT) em 1990
  - Videoconferência, etc (usando a rede telefónica p/ 64Kbps)
- MPEG 1 (ISO) em 1991
  - Armazenamento (em CD-ROM) de video (qualidade VHS) e audio com débito de 1,5Mbps
- MPEG 2 (H.262) em 1994
  - Armazenamento de video e audio de alta qualidade com débito de 10 a 20Mbps (>50Mbps)
  - Inclui transmissão HDTV
- H.263 em 1996 (extensões H.263+ H.263++)
- MPEG 4 em 1998
  - Transmissão de video e audio digital com débito baixos (< 64Kbps)
  - Inclui transmissão de VOD e video na internet
- H.264 em 2002 (também na norma MPEG4, part10) - AVC

## H.261

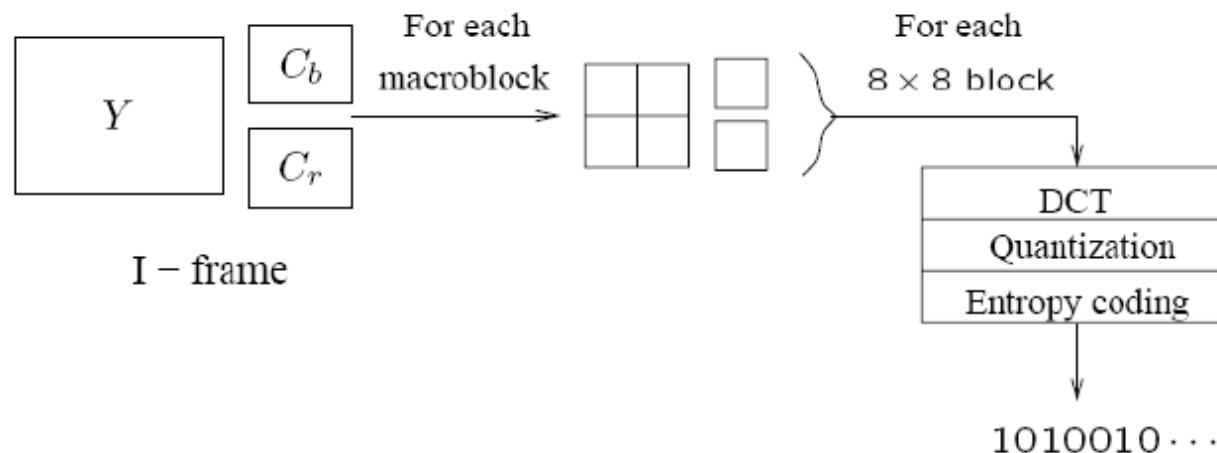
- Transmissão de vídeo na ISDN em tempo real  
(atraso máximo de 150 ms)
- Formatos de vídeo suportados:
  - QCIF(176x144)
  - CIF (352x288) (opcional)
  - Sub QCIF (128x96) (opcional)
- Subsampling 4:2:0
- Uma sequência é composta por dois tipos de frames:
  - I-frames (intraframes)
  - P-frames (interframes)



- O intervalo entre I-frames é variável, determinado pelo codificador;
- Na norma H.261 tipicamente envia-se duas I-frames por seg.

# H.261

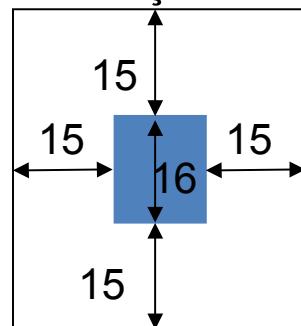
- I-frames (intraframes)
  - Remoção de redundância espacial
  - Tratadas como imagens independentes;
  - DCT para cada bloco de 8x8 do MB;
  - Quantificação (usa um coeficiente igual para o MB ou GOB);
  - zigzag;
  - Compressão por código de Huffman



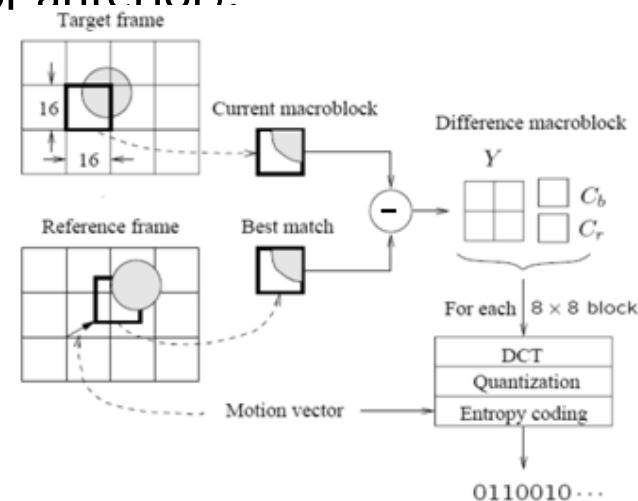
- Macro-bloco (MB): 4 blocos Y, 1 bloco Cb e um bloco Cr  
(Derivado do subsampling 4:2:0)

# H.261

- P-frames (interframes)
  - Remoção de redundância temporal
  - Predição usa uma I-frame ou uma P-frames anterior
  - Codificação da diferença entre MB (16x16 Luminância) e MB predito;
  - Compensação de movimento é opcional no codificador mas obrigatório no descodificador
  - Pesquisa usa-se valores inteiros do pixel numa janela +/-15 pixel;
  - Os vectores de movimento também são codificados (a efectuada a diferença entre o vector e o vector anterior):



Janela de pesquisa

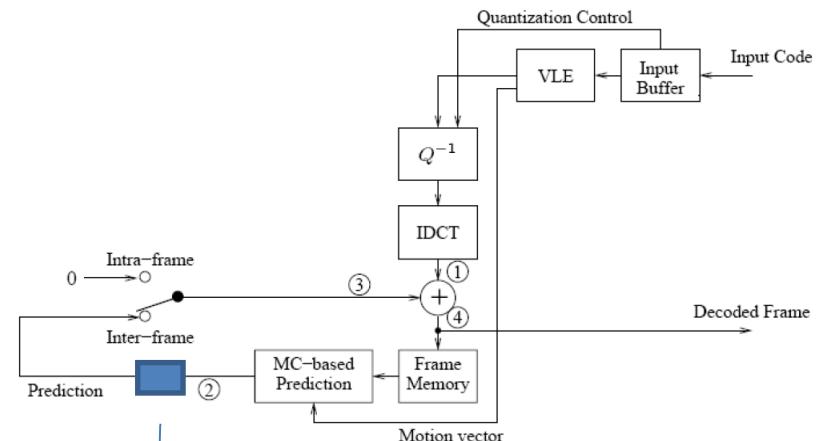
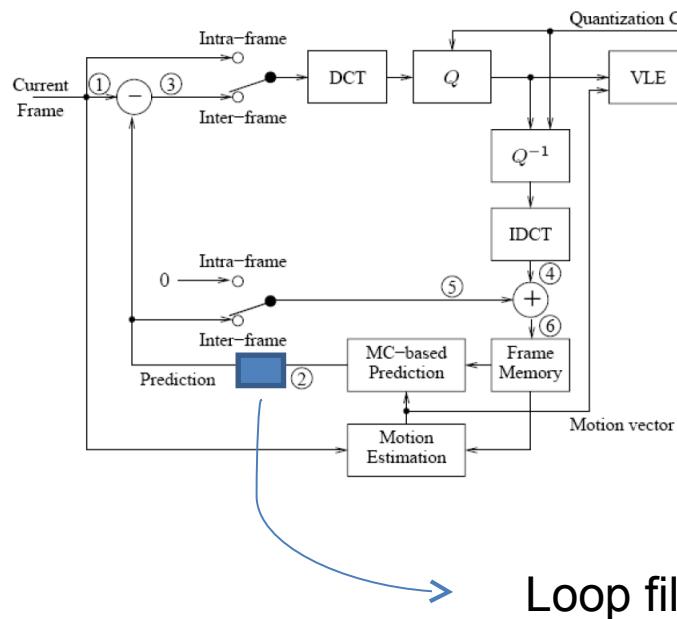


- Se o erro de predição é muito elevado, é codificado o macro-bloco em vez da diferença;

# H.261

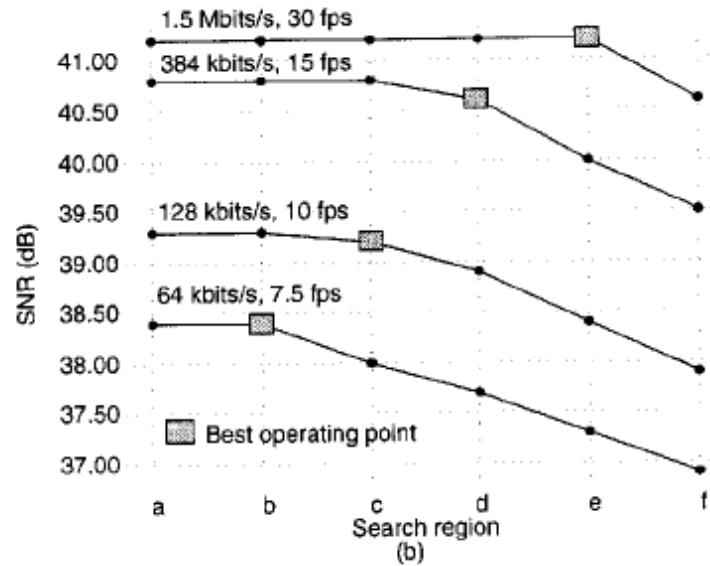
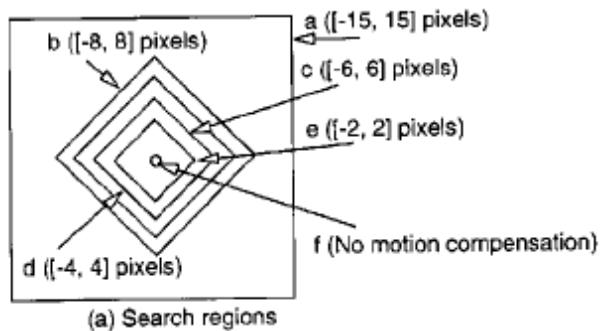
- Quantificação:
  - Usa-se uma constante (em substituição da matriz do JPEG)  
Pode ser qualquer valor par entre 2 e 62  
(uso de quantificador com “dead zone”)
  - Para o valor DC da I-frame, o valor de quantificação é 8.

- Codificador / Descodificador



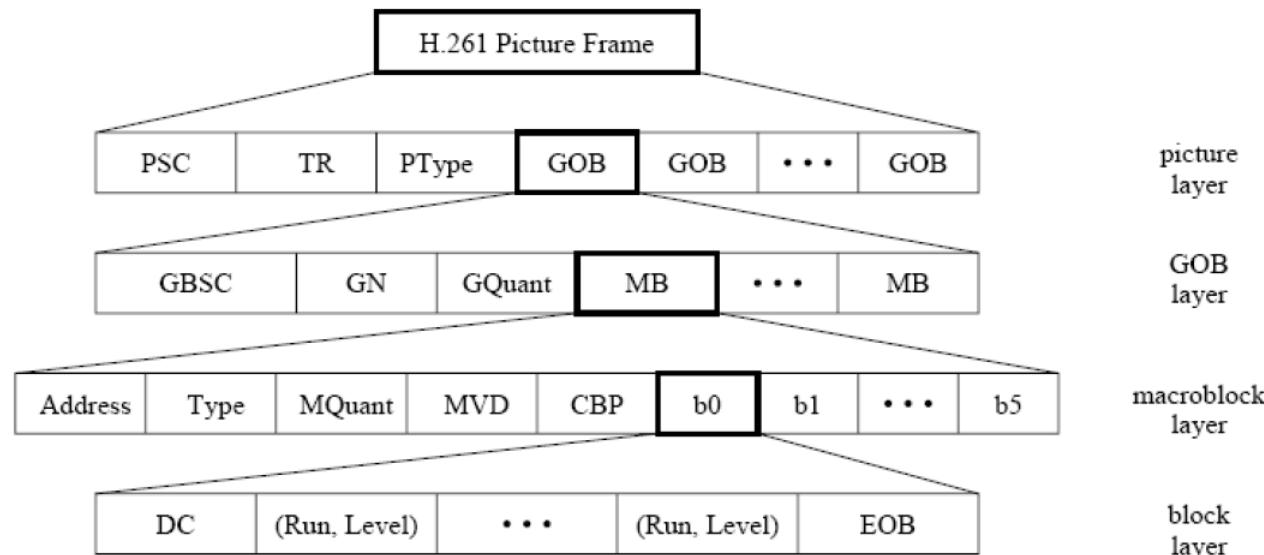
# H.261

- Trata apenas de pequenos movimentos (cabeça e ombros)
- Região de pesquisa é um losângulo em vez de rectângulo



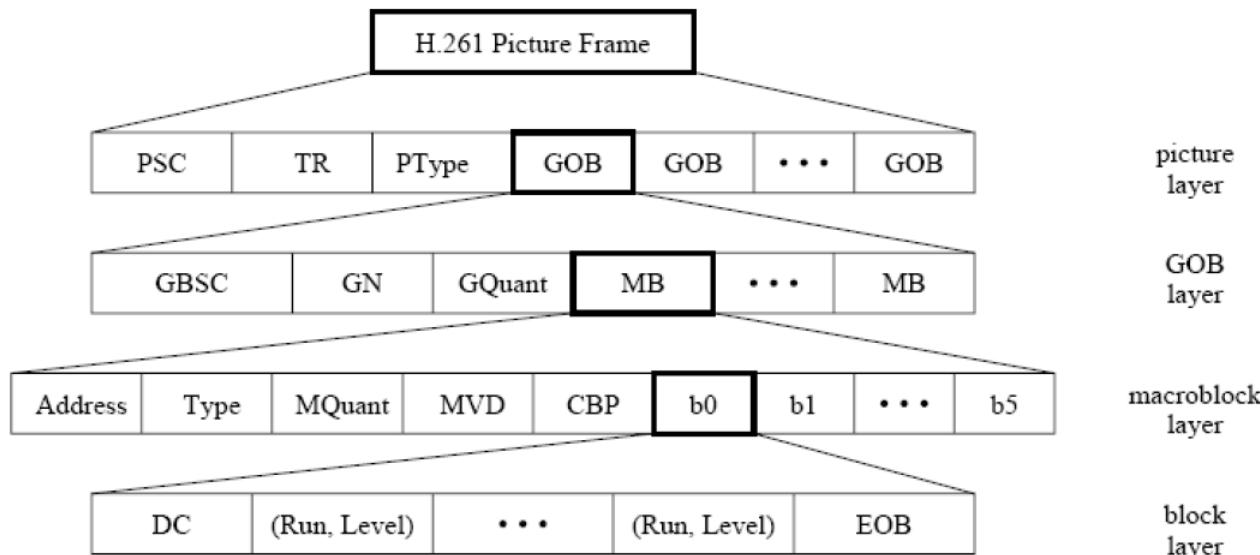
**Figure 7.4** Search regions and video quality for typical H.261 encoding. (a) Diamond-shaped search regions; (b) video quality for the corresponding search regions.

## H.261 – video bitstream



- PSC – Picture Start code
- TR –Temporal Reference  
(importante pois se houver subsampling temporal há frames que não são transmitidas e é necessário manter o sincronismo)
- Ptype – Picture Type (CIF ou QCIF)
- GOB – Group of Blocks (imagem dividida em regiões)  
(12 GOB para CIF e 3GOB para QCIF)
- GBSC – GOBs Start Code
- GN – Group Number
- GQuant – GOB quantizer
- MB – Macro-block (33 MB por cada GOB)

## H.261 – video bitstream



- Address – indica a posição do MB no GOB
- Type – inter ou intra, com/sem compensação de movimento  
(É possível numa imagem ter em simultâneo MB inter e MB intra)
- MQuant – MB quantizer
- MVD – Motion vector Data (diferença do vector de movimento do MB actual e anterior)
- CBP – Coded Block Pattern (indica quais os blocos codificados, os outros são preenchidos a zero)
- b(i) – blocks
- DC
- AC – (run, level); run entre 0 e 63 e level entre -127 e 127 (level sempre diferente de 0)
- EOB – end of Block

## MPEG 1

- Part 1 – Sistema
    - Bitstream, sincronismo audio / video, clock, gestão buffer
  - Part 2 – Video
  - Part 3 – Audio (MP3)
- 
- 1,5 Mbps (1,2Mbps video e 256kbps stereo)
  - Qualidade de video equivalente ao VHS e o audio equivalente ao CD.
  - Compromisso entre qualidade e atraso (entre 150ms e 1s)
- 
- Permite Fast Forward, fast backward, slow motion (forw/back), Pause e Random Access

## MPEG 1

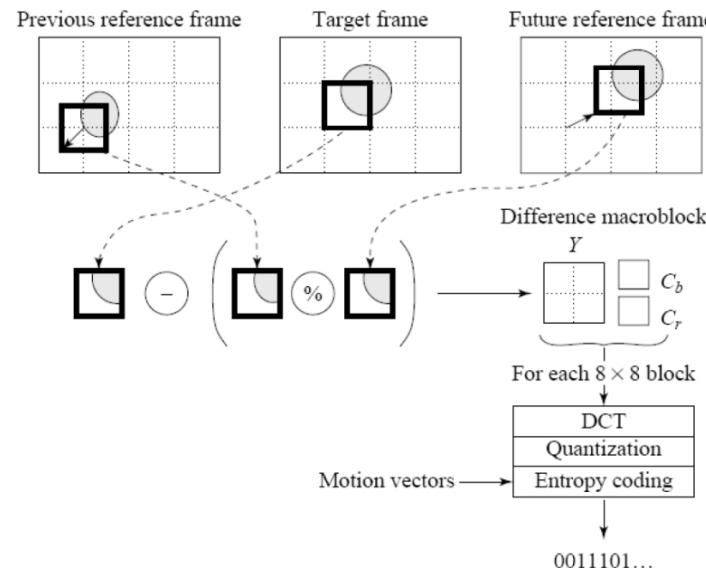
- Adota o formato SIF (source input format)
  - 352x240 (30fps) para NTSC
  - 352x288 (25fps) para PAL
- Só Usa formato Progressivo (não usa formato enterlaçado)
- Usa o chroma subsampling 4:2:0
- Suporta diferentes aspect ratio's (H.261 apenas suporta 4:3)
- Suporta outro tipo de formato, desde que

Número de colunas	$\leq 768$ pixeis
Número de linhas	$\leq 576$ pixeis
Débito de imagens	$\leq 30$ imagens/seg
Número de MB por imagem	$\leq 396$
Débito de MB	$\leq 9900$ MB/seg
Débito binário	$\leq 1.856.000$ bits/seg

- Ao contrário do H.261, cada imagem só tem MB do mesmo tipo

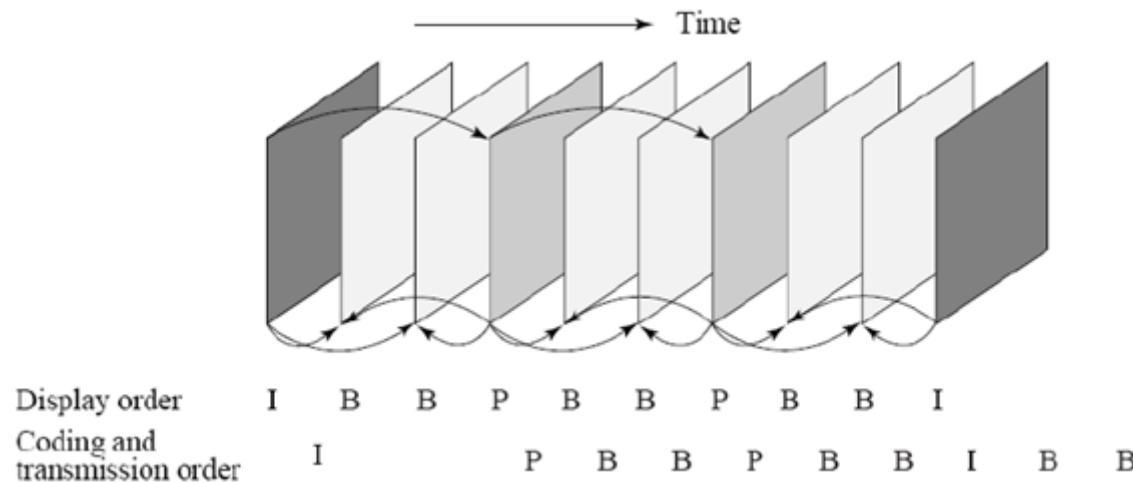
# MPEG 1

- Permite frames do tipo I, P, B ou D
- Compensação de movimento
  - Para cada MB da P-frame actual é determinado qual o MB mais semelhante da I ou P-frame anterior (forward prediction).
  - B-frame Predição bidirecional:  
Para cada MB da frame actual consideram-se 2 MV:  
um da frame anterior (I ou P) e outro  
da frame posterior (I ou P)  
(só se consideram os MV e MB no caso de haver boa correpondência).



## MPEG 1

- Dado que a predição de uma B-frame não é causal, a ordem de transmissão das frames é alterada



- As B-frames são eficientemente codificadas
- Maior bit rate
- Não há propagação de erro
- Necessária mais memória
- Maior atraso

## MPEG 1

- Compensação de movimento:
  - Estimação de movimento baseada em blocos
  - Precisão de  $\frac{1}{2}$  pixel (usando interpolação linear)
  - A janela de pesquisa é maior que na norma H.261



A  $\square$  a      ● b       $\square$  B

● c      ● d

C  $\square$        $\square$  D

# MPEG 1

- A quantificação dos coeficientes da DCT é diferente para as
  - Intra-frames:

$$QDCT[i, j] = \text{round} \left( \frac{8 \times DCT[i, j]}{\text{step\_size}[i, j]} \right) = \text{round} \left( \frac{8 \times DCT[i, j]}{Q_1[i, j] * \text{scale}} \right)$$

- Inter-frames:

$$QDCT[i, j] = \left\lfloor \frac{8 \times DCT[i, j]}{\text{step\_size}[i, j]} \right\rfloor = \left\lfloor \frac{8 \times DCT[i, j]}{Q_2[i, j] * \text{scale}} \right\rfloor$$

- Tabela (intra) –  $Q_1$

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

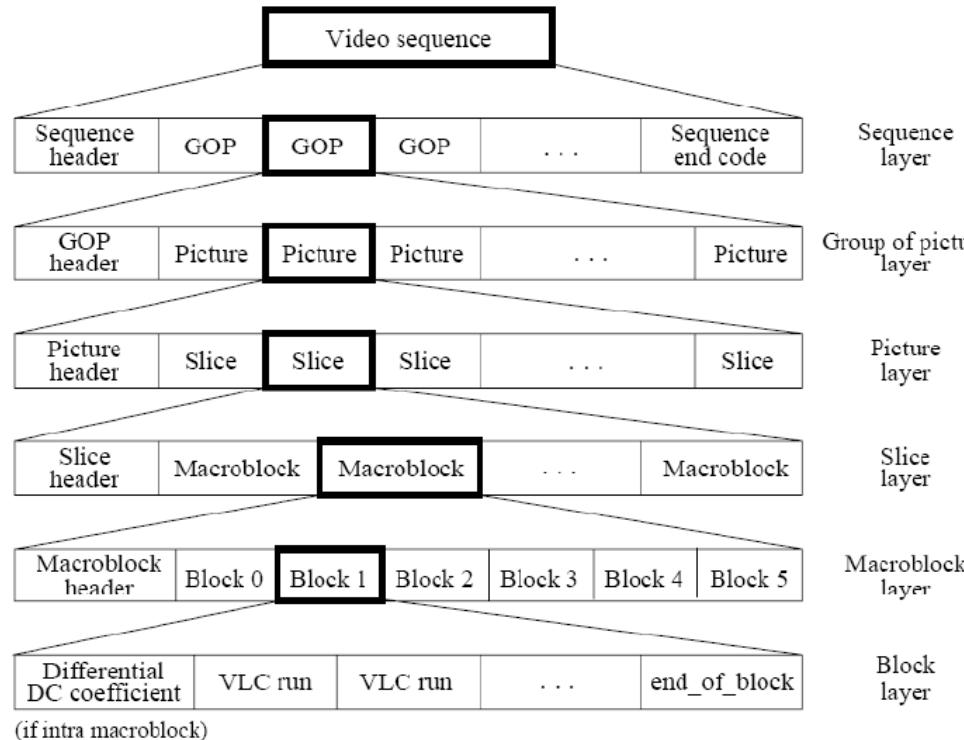
- Tabela (inter) –  $Q_2$

16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	16

- Scale é um inteiro entre 1 e 31

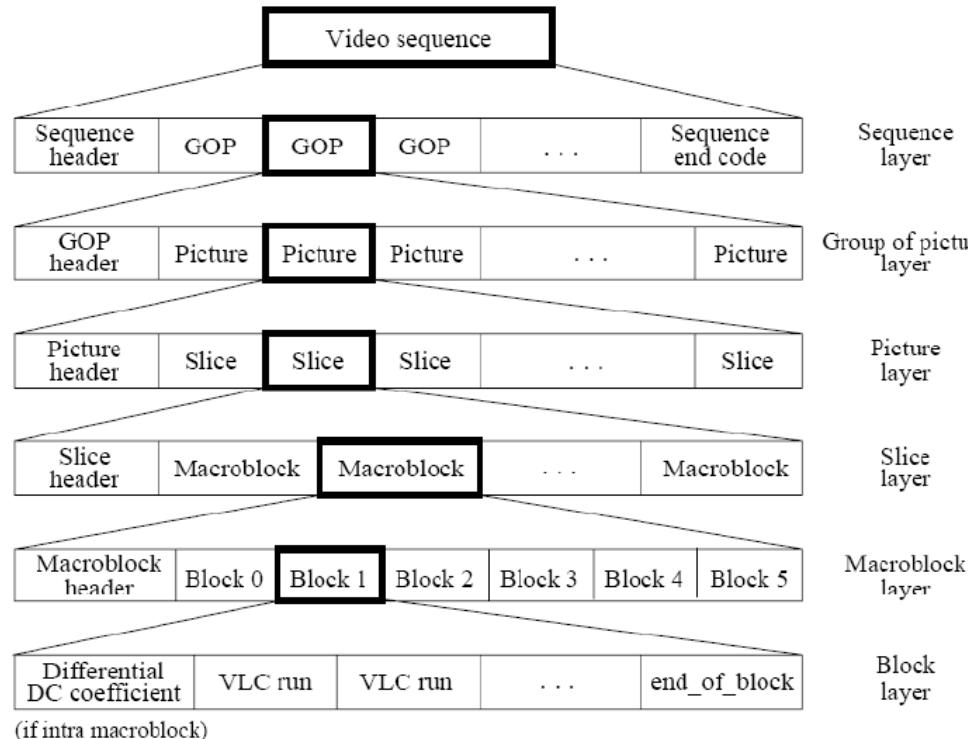
# MPEG 1 – Estrutura de dados

- Sequência de video formada por:
  - Sequence header
    - informação sobre tamanho vertical e horizontal, aspect ratio, frame rate, bit rate, ...
  - vários grupos de frames (GOP) – este conceito não existe no H.261



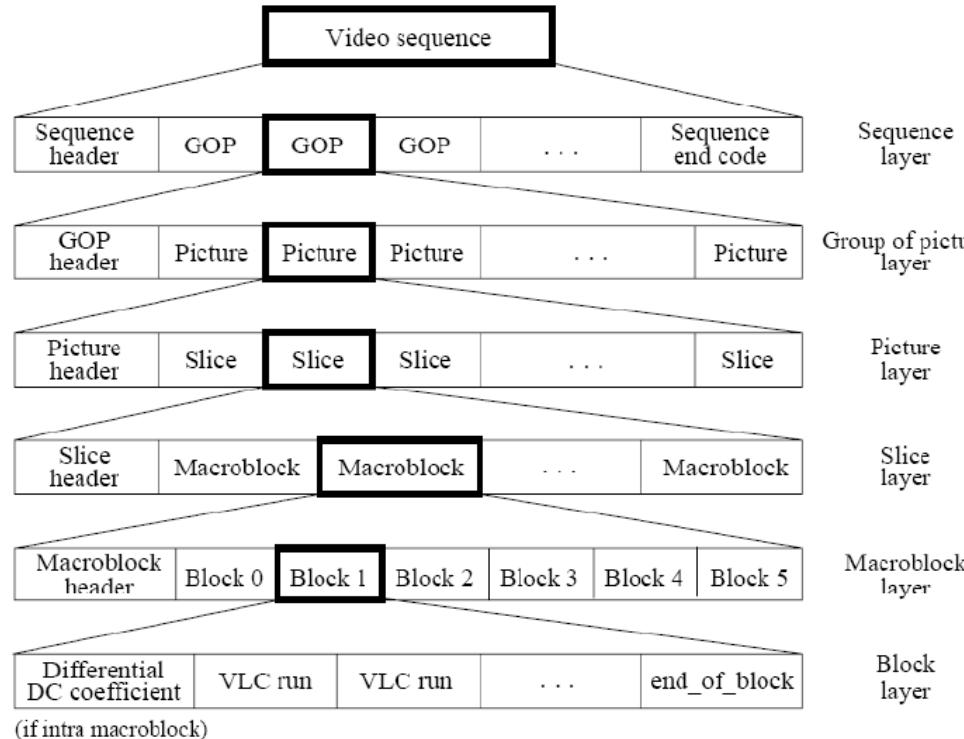
# MPEG 1 – Estrutura de dados

- Cada GOP tem:
  - GOP header: várias informações (ex: time\_code)
  - Número variável de frames (começa com uma I- ou B-frame e acaba com uma I- ou P-frame)
  - Tipicamente tem entre 12 e 15 frames (e com frames P de 3 em 3)



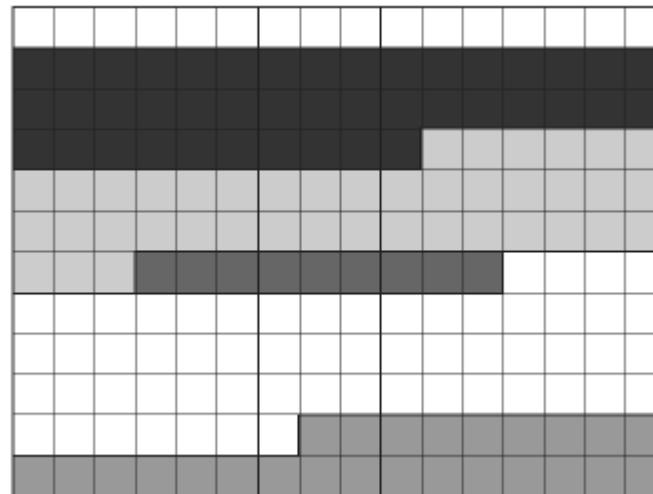
# MPEG 1 – Estrutura de dados

- Cada PICTURE contém :
  - PICTURE header
  - SLICES
- Podem ser I-, P- ou B-frames, ou ainda D-frames
- D-frames: caso particular das I-frames, apenas contém os coeficientes DC's



## MPEG 1

- Uma frame é dividida em vários SLICEs em vez dos GOBs (usados na norma H.261)
- O SLICE começa e acaba numa imagem



- Cada slice contém um número variável de MB.  
(a codificação de cada slice é independente e podem ter tabelas de quantificação diferentes - scale)  
O conceito de slice é importante para a sincronização e a recuperação de erros: é o nível mais baixo na hierarquia que pode ser totalmente recuperado sem descodificar todos os VLC's
- O tamanho e a posição do SLICE é especificado no header

## MPEG 1

- Cada Macrobloco é constituído por 4 blocos (Y) um bloco (U) e um bloco (V)
- Cada bloco tem dimensão 8x8
- 
- Blocos intra:
  - diferença entre coeficientes DC's e VLC dos coeficientes AC's
- outros Blocos:
  - Coeficientes DC e AC codificados com VLC.

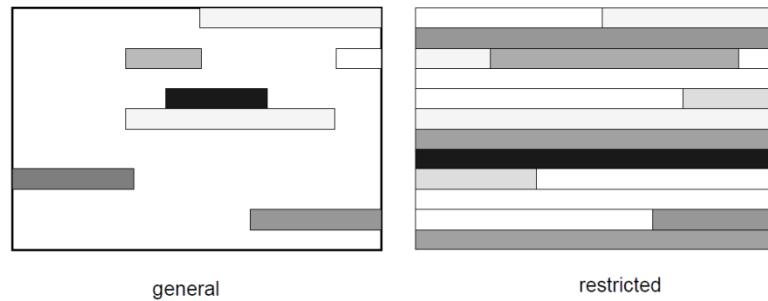
## MPEG 2

- Codificação de alta qualidade (até 20Mbps), inclui:
  - Broadcast Digital TV
  - VOD; direct TV
  - HDTV
  - DVD
- Aumentar a robustez a erros;
- Retro-compatibilidade com MPEG1 (no entanto o GOP é opcional);
  - Resoluções até 16383x16383
- Formato desde CIF até HD (1920x1250 com 60 tramas por seg.)
- Chroma-subsampling (4:2:0, 4:2:2 e 4:4:4)
- Compatibilidade com o video enterlaçado;

## MPEG 2

- Slices em modo normal ou modo restrito:
  - Modo normal
  - Modo restrito

(cada slice começa e termina na mesma linha de MB)



- Escalabilidade na codificação para adaptar a diferentes largura de banda e diferentes receptores;
  - SNR
  - Temporal
  - Espacial
- Modos: perfil (complexidade na codif./ descodific) e nível (restrições bit rate, resolu. Espacial,...)

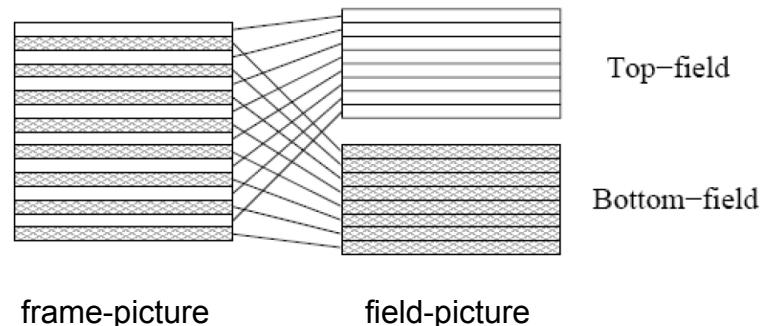
# MPEG 2

		PERFIL (Profile)								
NIVEL (Level)	Resolução	Frames /sec	Simple (SP) 4:2:0	Main (MP) 4:2:0	SNR Escal. SNR 4:2:0 (Main + SNR)	SPATIAL Escal.Espac 4:2:0 (SNR + Espacial)	High (HP) 4:2:0 /4:2:2 (+ espacial)	4:2:2 (extende Main com maior Deb. Bin.)	Aplicações	
NIVEL (Level)	Very High (HL)	1920x1152	60		IPB 62.7Mpix/s 80 Mbps			IPB 100Mbps (3 layers)	IPB 300Mbps	produção de filmes
	High (H14)	1440x1152	60		IPB 47 Mpix/s 60 Mbps		IPB 47Mpix/s 60Mbps (3 layers)	IPB 80Mbps (3 layers)		consumidor HDTV
	Main (ML)	720x576	30/25	IP 10.4Mp/s 15Mbps	IPB 10.4 Mpix/s 15 Mbps	IPB 10.4Mp/s 15Mbps (2 layers)		IPB 20Mbps (3 layers)	IPB 50MBps	estudio de TV
	Low (LL)	356x288	30/25		IPB 3.04 Mpix/s 4 Mbps	IPB 3.04Mp/s 4Mbps (2 layers)				consumidor,qual idade VHS

Na norma MPEG2 também existe no nível Main o perfil MVP – multi view profile

## MPEG 2

- A norma MPEG2 suporta video progressivo e entrelaçado, dado que são modos de HDTV. (MPEG1 só suporta video progressivo)
- Duas estruturas
  - frame-picture (dois campos numa só estrutura)
  - Field-picture (cada campo codificado separadamente)
- Em ambas têm-se frames I, P e B
- No enterlaçado cada campo (field) também são codificados em I, P e B.
- Cada field é dividido em macro-blocos de 16x16.

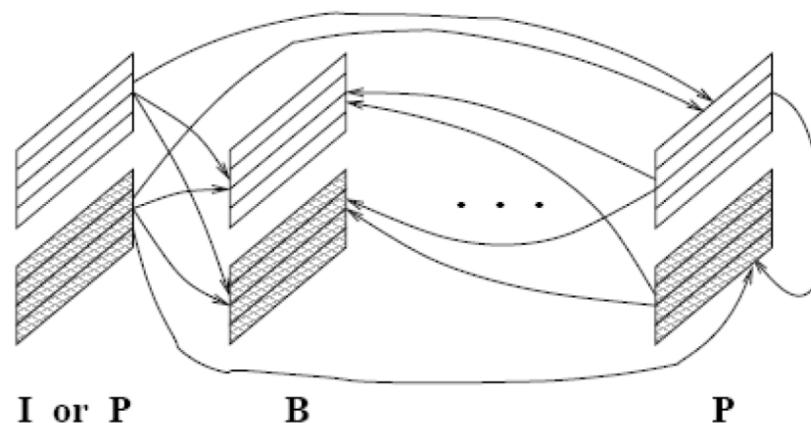


## MPEG 2

- Compensação de movimento
  - Janela de pesquisa aumenta de +/-16 para +/-64pixels ??
- Modos de Predição para compensação de movimento:
  - Frame prediction for frame-pictures
  - Field prediction for field-picture
  - Field prediction for frame-picture
  - 16x8 Macroblock for field-picture
  - Dual-prime for P-picture (para field-picture e frame-picture)

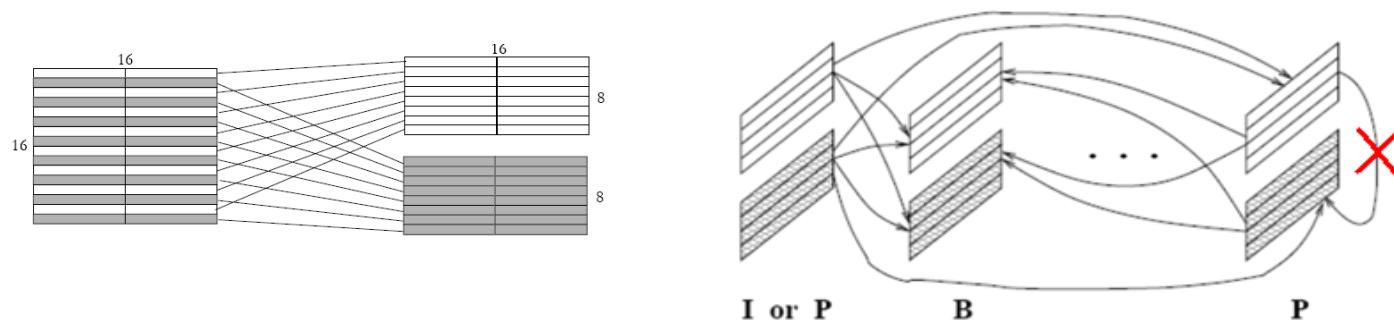
## MPEG 2

- Frame prediction for frame-pictures
  - Idêntico à predição em MPEG1 (P e B-frames).
  - Usa 1 MV para cada MB(16x16) na P-frame e 2 MV para a B-frame.
  - Adoptado em videos com movimentos lentos ou moderados dos objectos ou da câmara.
- Field prediction for field-picture
  - Usa MB de 16x16 das field-picture
  - P-field-picture predição com base nos 2 últimos fields.
  - Predição dos MB do top-field usam o top-field ou bottom-field da frame ant. Predição dos MB do bottom-field usam o top-field da frame actual ou o bottom-field da frame anterior.
  - B-field predita com base nos field-picture das I ou P-frames anterior e post.
  - Usa 1 MV para cada MB(16x16) na P-frame e 2 MV para a B-frame.

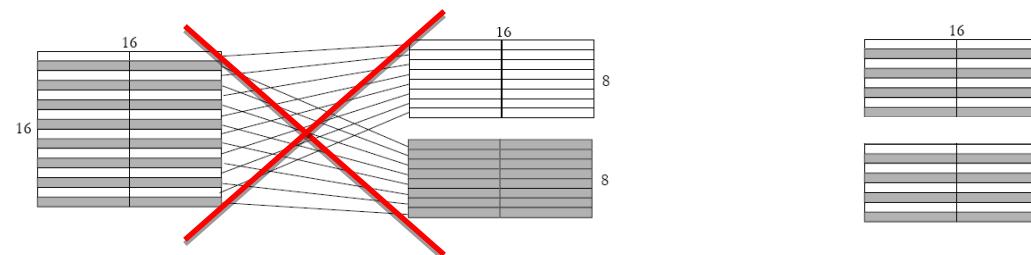


# MPEG 2

- Field prediction for frame-picture
  - Cada field-picture é tratado separadamente.
  - Os MB de 16x16 são divididos em 2 de 16x8 (um de cada field)
  - Usa 2 MV para cada MB(16x16) no P-field e 4 MV para o B-field.
  - Este modo funciona bem com movimentos rápidos



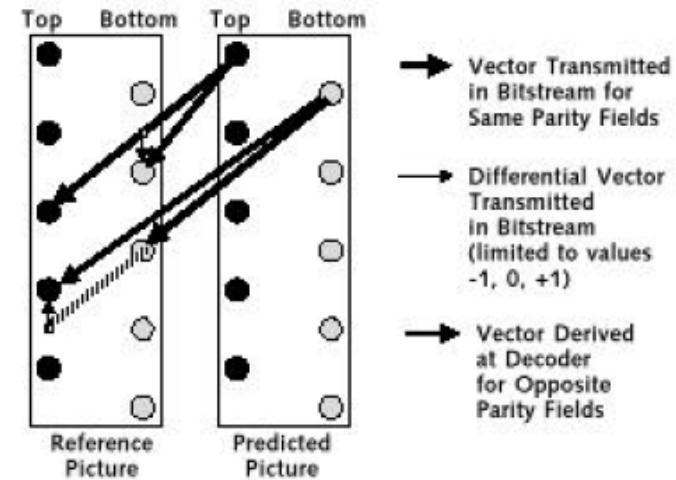
- 16x8 Macroblock for field-picture



- Os MB de 16x16 são divididos ao meio (16x8)
- Usa 2 MV para cada MB(16x8) no P-field e 2 ou 4 MV para o B-field.
- Este modo funciona bem com movimentos irregulares.

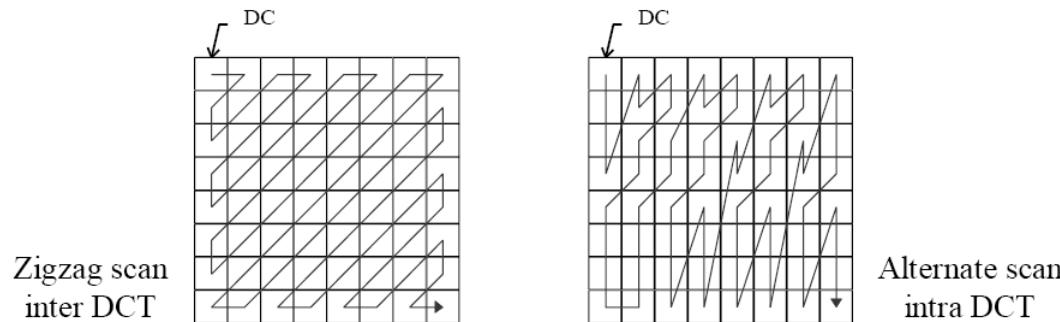
## MPEG 2

- Dual-prime for P-picture
  - Pode ser usado em frame-picture ou em field-picture;
  - Aplicado apenas ao tipo P onde não existam imagens B entre a imagem alvo e a imagefm de referência
  - A predição é feita em relação ao mesmo field da frame de referência.
  - Calcula-se o vector correspondente ao outro field.
  
- Codifica-se um MV e um vector de movimento diferencial
  - Para frame-picture são derivados os 2 MVs (para MB de 16x16) é feita a média para a perdição final
  - Para a field-picture usam-se 2 sub-blocos de 16x8 obtendo-se 4 predições



## MPEG 2

- DCT – aplica-se a frames ou a fields em blocos de 8x8
- Video progressivo usa-se o zig-zag (igual ao JPEG)
- Video entrelaçado:  
Dado que linhas consecutivas do mesmo bloco são de diferentes fields, há menor correlação entre elas do que entre linhas alternadas.



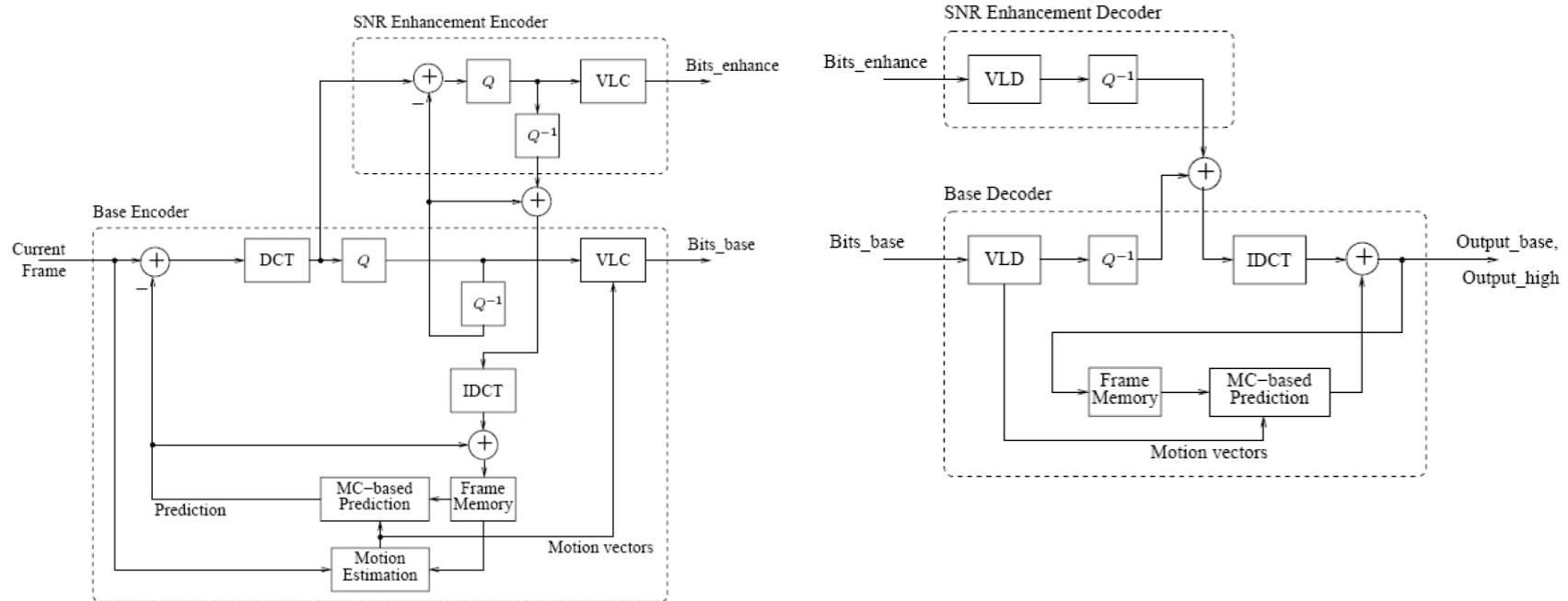
- É melhor em video com movimentos rápidos
- Os ganhos no PSNR são na ordem de 0.3 dB em relação ao zig-zag

## MPEG 2

- Modo de codificação escaláveis:
  - Codificação hierárquica do vídeo;
  - As camadas correspondem a Bitstreams que podem ser transmitidos em canais diferentes
- Modos:
  - SNR : base layer e enhancement layer (melhor SNR)
  - Espacial: base layer e enhancement layer (maior resolução espacial)
  - Temporal: base layer e enhancement layer (maior frame rate)
  - Hibrida: combina dois dos três modos anteriores .
  - Partição de dados: depois da quantificação dos coeficientes da DCT este são separados em diferentes partições.

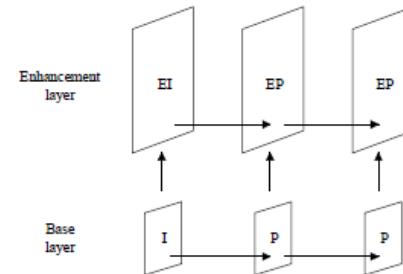
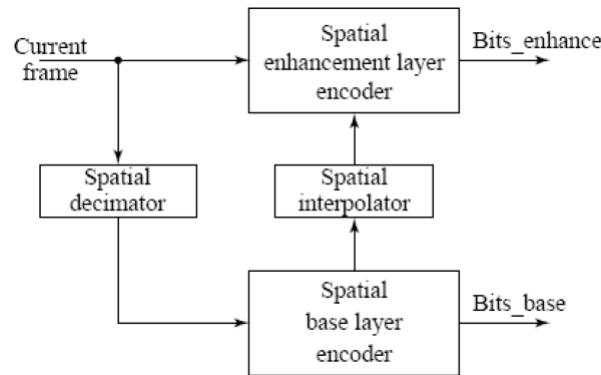
# MPEG 2

- SNR escalável:

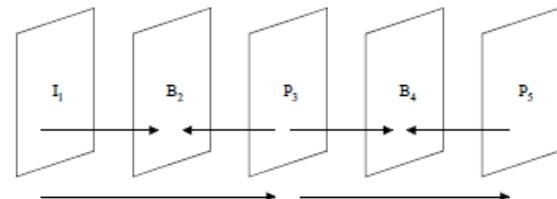
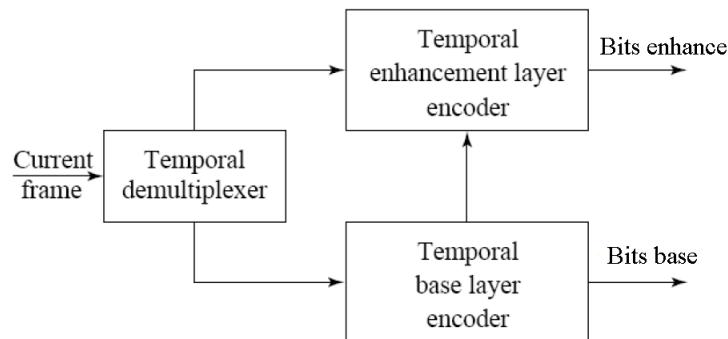


# MPEG 2

## ■ Escalabilidade espacial:

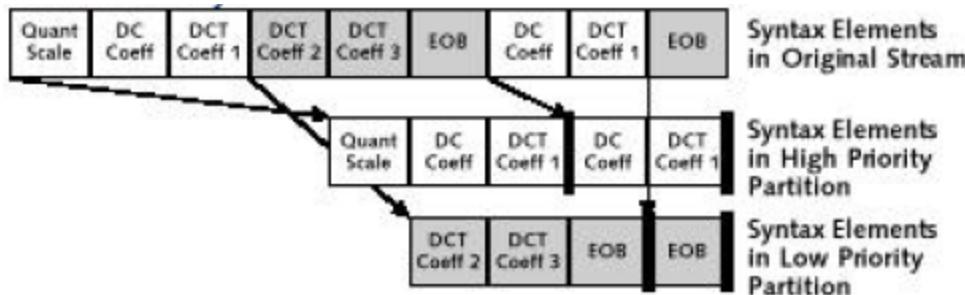


## ■ Escalabilidade temporal:



# MPEG 2

- Partição de dados :



## H.263

- Norma para video conferência sobre a rede telefónica fixa e móvel;
  - Melhoramento da norma H.261 (melhor PSNR)
  - Bit rate baixo (menor que 64kbps)  
mas com qualidade é melhor que H.261
  - Suporta os formatos (CIF e QCIF são obrigatórios):

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	Bit-rate (kbps) BPPmaxKb (compressed)
sub-QCIF	$128 \times 96$	$64 \times 48$	4.4	64
QCIF	$176 \times 144$	$88 \times 72$	9.1	64
CIF	$352 \times 288$	$176 \times 144$	36.5	256
4CIF	$704 \times 576$	$352 \times 288$	146.0	512
16CIF	$1,408 \times 1,152$	$704 \times 576$	583.9	1024

- GOB (são diferentes da H.261) têm dimensão variável, começam e acabam nas margem esquerda e direita da frame.
- É abandonado o “loop filter”
- Imagens do tipo I ou P, mas nas P podem usar-se MB do tipo I ou P
- Usa o formato 4:2:0 como o H.261

## H.263

- Compensação de movimento
  - Calculam-se as componentes vertical e horizontal do vector de movimento

$$u_p = \text{median}(u_1, u_2, u_3),$$

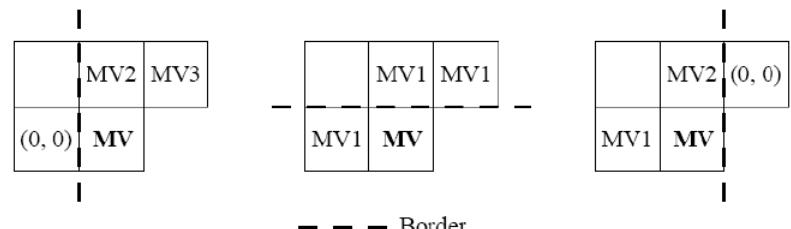
$$v_p = \text{median}(v_1, v_2, v_3).$$

- Usa-se a mediana como predição
- Codifica-se e transmite-se o vector de erro  $(\delta u, \delta v) = (u - u_p, v - v_p)$

	MV2	MV3
MV1	<b>MV</b>	

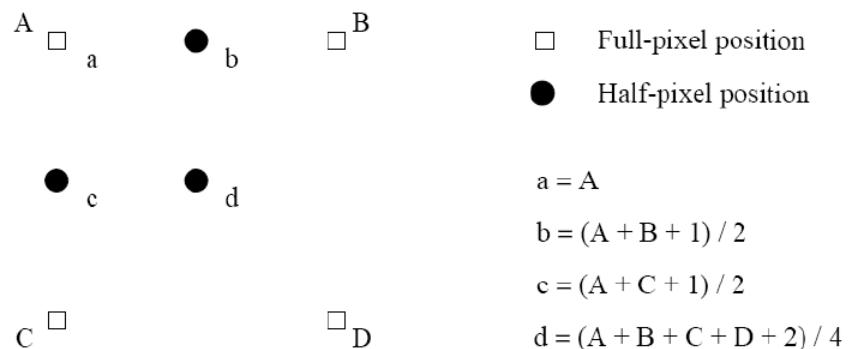
**MV** Current motion vector  
**MV1** Previous motion vector  
**MV2** Above motion vector  
**MV3** Above and right motion vector

(a)



## H.263

- A precisão é de  $\frac{1}{2}$  pixel sendo a janela de -16 a 15.5 (usa interpolação bilinear)

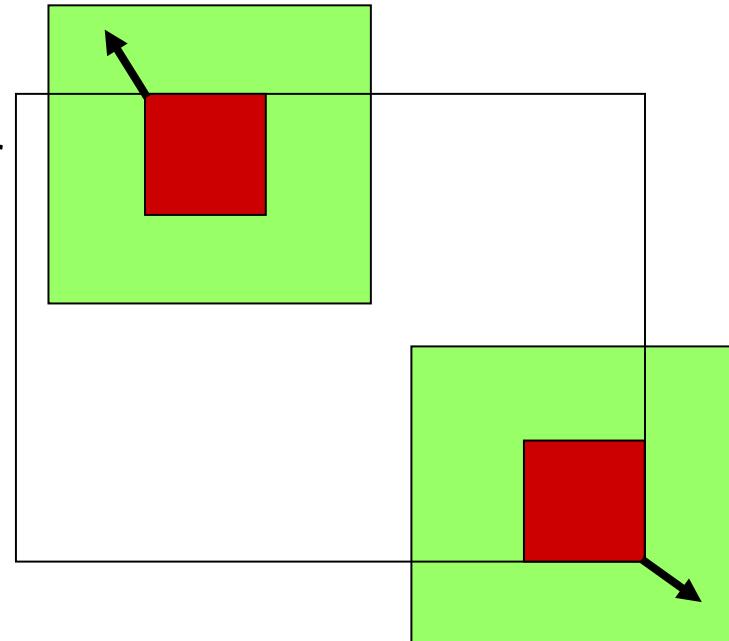


- VLC em vez de usar o par {run, level} usa {last,run,level} (chamado de 3D-VLC)
- Dos 22 bits o last (1bit) indica o último coeficiente (equivalente ao EOB) o par run (6bit) e o level (8 bit) têm as mesmas funções
- A norma tem 18 modos negociáveis para além da codificação base
- A norma teve várias adendas, publicadas em 1998 (H.263+) e 2005 (H.263++) - Anexos

## H.263

- Modos de funcionamento

- Modo não restrito:
  - se o vector de movimento apontar para fora da imagem assume-se o valor do pixel da região fronteira que é mais próximo do pixel para o qual o vector aponta.
  - extende a janela de  $[-16, 15.5]$  para  $[-31.5, 31.5]$
  - (melhora situações com objectos a entrar na cena)



- Modo codificação aritmética baseada na sintaxe:
  - Todos os VLC's das 4 layers usam codificação aritmética.
  - Os ganhos no débito total são na ordem de 4% nas inter-frames e 10% nas intra-frames (mantendo o SNR).

## H.263

- Modos de funcionamento

- Modo PB frame:

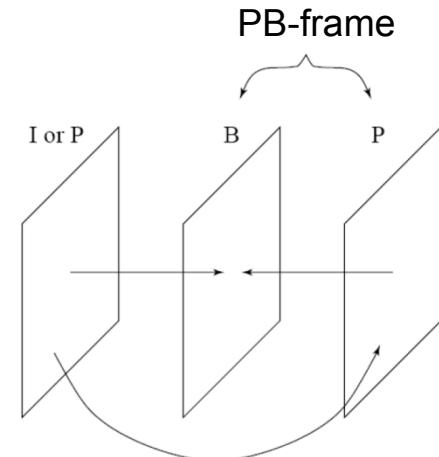
PB-frame são duas frames codificadas como uma só

P-frame é predita com base na P-frame previamente descodificada

B-frame é predita a partir das P-frame já descodificada e na P-frame actual.

São intercalados os MB da frame P e da B

- Anexo M: Melhora estas permitindo optar por apenas MV backward ou forward



# H.263

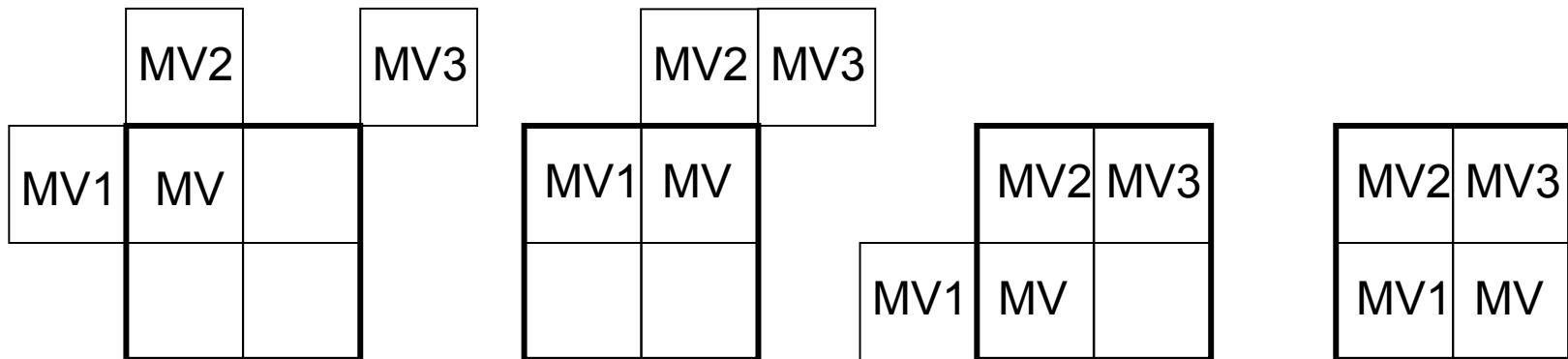
- Modos de funcionamento

- Modo de predição avançado:

método de compensação de movimento com sobreposição de blocos  
(reduz artefactos do efeito de bloco)

Pode-se usar 4 vectores de movimento no MB (um por para cada bloco 8x8) para luminância (MV da crominância é a média destes)  
cada pixel do macrobloco do bloco predito soma ponderada dos vectores do próprio bloco, do bloco da esquerda ou direita e do bloco de cima ou de baixo.

Embora haja mais vectores para transmitir, os ganhos na compressão são muito maiores pois a predição é muito melhor.



## H.263+ e H.263++

- Totalmente compatível com a norma H.263
- Objectivo:
  - Oferecer maior flexibilidade (formatos, aspect ratio, freq. clock);
  - 12 novos modos de funcionamento;
  - Maior eficiência;
  - Redução do efeito de bloco.
- À semelhança das normas MPEG 1 e 2
  - A diferença entre vectores de movimento é codificada com VLC reversível (RVLC) permitindo minimizar erros na transmissão (level passou a ser entre -256 e 256 em vez de -127 e 127 )
  - SLICE substitui o GOB:
    - um SLICE contém um número variável de MB
    - O SLICE pode ter qualquer forma
    - A transmissão pode ser sequêncial ou não
  - Escalabilidade temporal , espacial, SNR.
  - E outras ... ver anexos

## H.263 Anexos

- Anexo C – Multiplexagem de video (4 sub-bitstreams)
- Anexo I – Advanced intra coding  
(neste modo os MB são preditos usando os MB vizinhos)
  - Modo predição apenas DC;
  - Predição Dc e AC vertical;
  - Predição Dc e AC horizontal;
- Anexo J – Filtro de redução de efeito de bloco  
(Filtra pixeis fronteira dos blocos de 8x8)
- Anexo K – Slice structured (Slices substituem o GOB)  
(oferece maior resiliência aos erros e minimiza atrasos)
- Anexo N – Reference picture mode (a imagem de referência deixa de ser apenas a última a ser codificada)
- Anexo O – Escalabilidade Temporal, espacial e SNR (semelhante ao MPEG2)

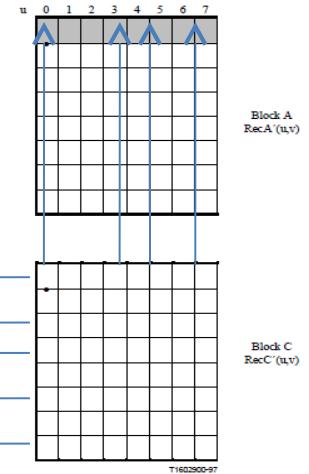


Figure I.3/H.263 – Three neighbouring blocks in the DCT domain

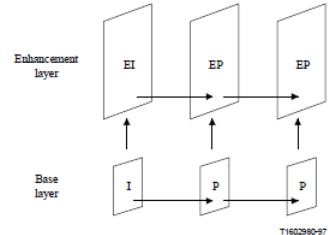
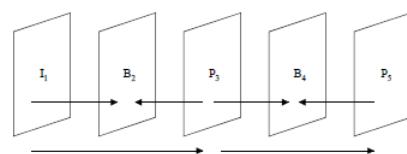


Figure O.3/H.263 – Illustration of spatial scalability

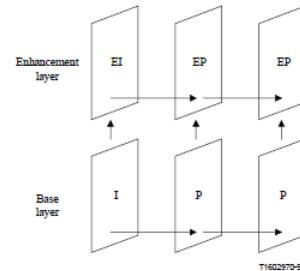


Figure O.2/H.263 – Illustration of SNR scalability

## H.263 Anexos

- Anexo P – Reference picture resampling  
(Imagen de Referência reamostrada ou deformada warping com 4 vectors para a deformar)
- Anexo Q – Reduced-Resolution Update mode  
(codifica apenas as diferenças a resoluções mais baixas)
- Anexo R – Independent segment decoding  
(evita que um erro se propague para outras regiões na imagem)
- Anexo S – Alternative inter VLC  
(Permite aplicar tabelas de intra para blocos inter)
- Anexo T – Modified Quantization  
(maior flexibilidade no controlo do passo de quantificação)
- Anexo U – Enhanced reference picture selection  
(melhora o anexo N, não se podendo usar ambos)
- Anexo V – Data partitioned slice mode  
(submodo de fatias melhora o anexo K)
- Anexo X – Propõe níveis e perfis

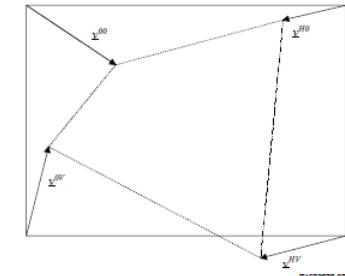


Figure P.1/H.263 – Example of conceptual motion vectors used for warping

## MPEG4

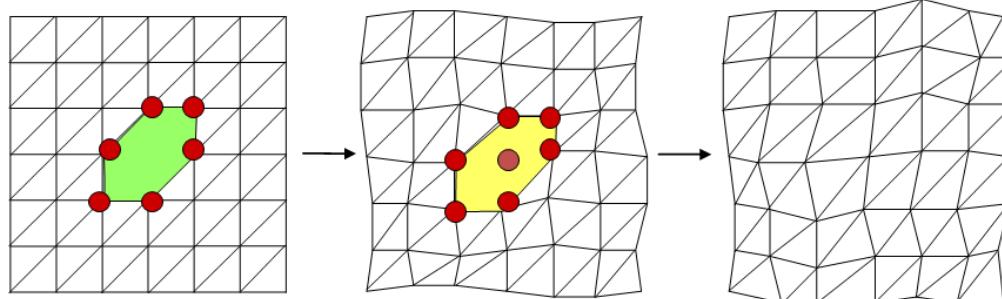
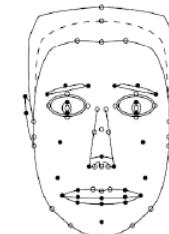
- Desenvolvido com base nas tendências:
  - Comunicação sem fios
  - Aplicações interactivas  
(navegação e retrivial)
  - Múltiplas aplicações  
(múltiplos fluxos de dados em simultâneo –  
vários pontos de vista da mesma cena)
  - Accesso universal  
(robustez a erros em canais ruidosos)
  - Flexibilidade e extensibilidade (Escalabilidade)
- Criada uma toolbox para diferentes tipos de dados audio-visuais
- Inicialmente pretendia-se transmissões com débito Binário baixo  
com o mínimo de distorção e baixa complexidade
- Paradigma da codificação de vídeo convencional ou baseada em  
objectos
- Manipulação e edição do bitstream sem descomprimir-lo

# MPEG4

- Part 1 – sistema
- Part 2 – Video
- Part 3 – Audio
- ...
- Part 10 – Advanced Video Coding (AVC)

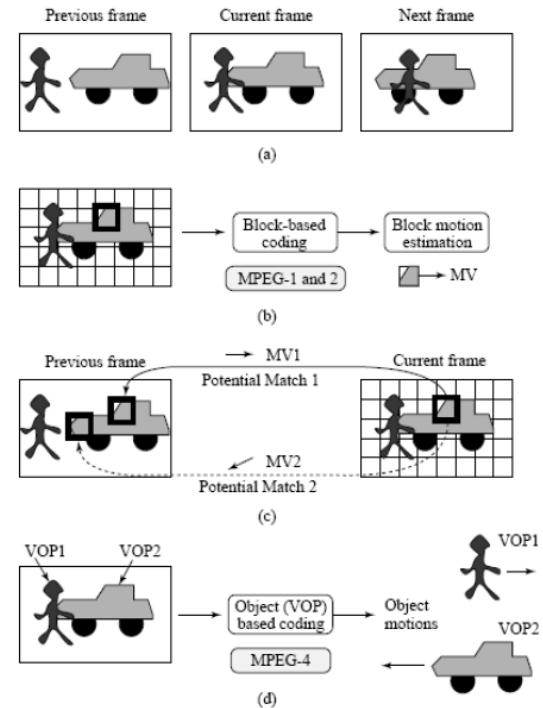
## MPEG4 – Part 2

- Vídeo digital convencional
- Objectos de vídeo (forma arbitrária)
- Objectos naturais e sintéticos
- Animação de faces ou corpos
- Objectos deformáveis (2D ou 3D) tipo malha



## MPEG4 – Part 2

- Aplicações de vídeo convencionais
- Débito binário varia de 5kbps a 10Mbps (vídeo com 12 bit/pixel)
- Vários tamanhos de imagens (Progressivo e enterlaçado)
- Escalabilidade espacial e temporal
  
- Paradigma da codificação de vídeo baseada em pixels, blocos e janelas rectangulares para remover correlações
  
- Codificação de formas e texturas



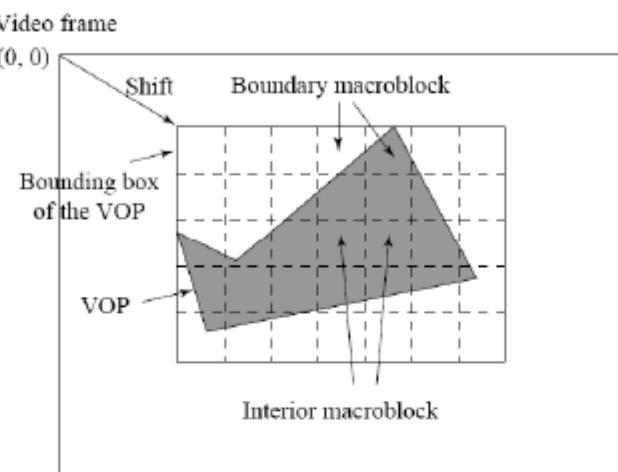
## MPEG4 – Part 2

- Estrutura hierarquica de vídeo orientado por objectos

- VS – sequência de 2D ou 3D de objectos naturais ou sintéticos
- VO – Objecto com forma arbitrária (cada VO pode conter vários VOL)
- VOL – facilita a escalabilidade
- VOP – Snapshot do VO numa imagem 2d num instante

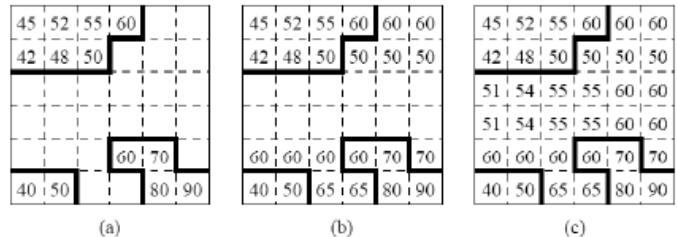
Video-object Sequence (VS)
Video Object (VO)
Video Object Layer (VOL)
Group of VOPs (GOV)
Video Object Plane (VOP)

- Compensação de movimento baseada em objectos:
  - Estimação de movimento
  - Predição e codificação
  - Só se considera os pixeis do VOP
  - VOP dividido em MB



# MPEG4 – Part 2

- Codificação do VOP
  - I-VOP, P-VOP, ou B-VOP
  - Define-se uma bounding box para cada VOP (MB fronteira e MB interiores)
  - Antes da DCT é feito o Padding (apenas nos VOP de referência)
- Codificação de Textura
  - Usa DCT nos MB dos VOP seguida de VLC
  - Para os MB fronteira usa-se SA-DCT (shape adaptive DCT)
  - Também usa transformada Hadamard para os DC dos MB
- Codificação da Forma (2 tipos)
  - Trasparente ou opaco (0 ou 1)
  - Escala de transparência (0-255)



## AVC = H.264 / MPEG4 – Part 10

- Codificação de vídeo baseada em frames
- Aumento da eficiência (50% do débito com a mesma qualidade do H.263, MPEG2 e MPEG4-part 2)
- Resiliência a erros (em meios como internet móvel e wireless)
- Características:
  - Componentes de côn com 8 a 14bits
  - Formatos 4:2:0, 4:2:2, 4:4:4
  - Conversão RGB – YCgCo (evita operações com vírgula flutuante)
  - Transformada inteira de blocos 4x4 (reduz artefactos)
  - Múltiplas tramas de referência – cada MB pode ter até 8 e cada imagem até 16 imagens de referência  
(permite maior compressão e maior resiliência a erros)
  - Cada referência tem peso diferente (weighted prediction)
  - Slice tipo I (predição dentro do mesmo slice) tipo P e tipo B
  - B-slice pode servir como referência
  - B-slice podem usar como referência 2 (ambas posteriores ou anteriores)

## AVC = H.264 / MPEG4 – Part 10

- Características:
  - Compensação de movimento com estrutura em árvore (16x16 , 8x16, 16x8, 8x8, 4x4)
  - Vectores de movimento com resolução de  $\frac{1}{4}$  pixel
  - Deblocking filter in the loop



1) Without filter

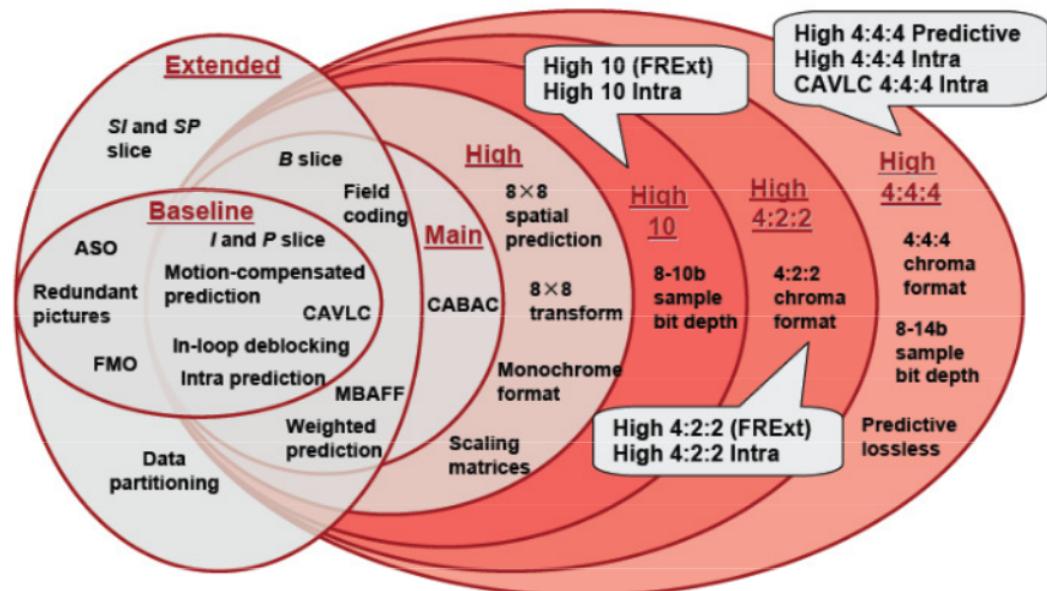
2) With H.264/AVC deblocking

## AVC = H.264 / MPEG4 – Part 10

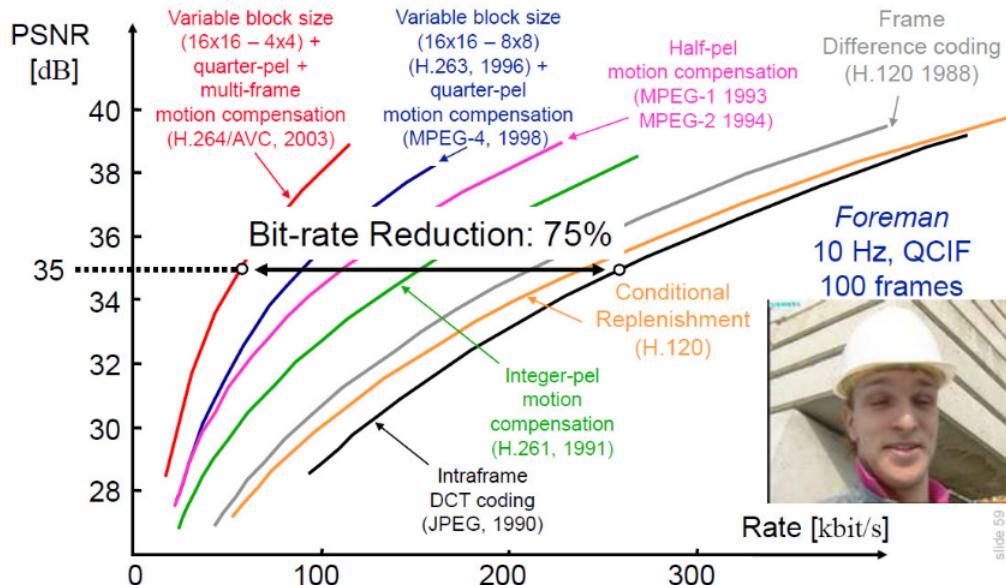
- Características:
  - Codificação entrópica:
    - UVLC – Universal VLC (Usa uma tabela em vez de várias)
    - CAVLC – Context Adaptive VLC (seleciona a tabela com base na informação já processada )
    - CABAC – Context Adaptive Binary Arithmetic Coding (o modelo de probabilidade é actualizado à medida que se processa a informação)
  - Modos flexíveis para evitar erros noadamente em redes IP
    - FMO – Flexible MB ordering
    - ASO – Arbitrary slice ordering
    - RS – Redundant slices

# AVC = H.264 / MPEG4 – Part 10

- Perfis:
  - Main – broadcast and storage (Blue-ray disc)
    - Não inclui FMO, ASO, RS
  - High,
  - High10
  - High 10 intra
  - High 4:2:2
  - High 4:2:2 intra
  - High 4:4:4
  - High 4:4:4 intra



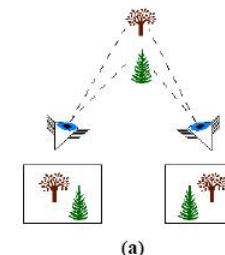
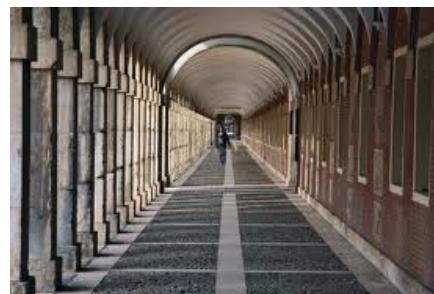
# AVC = H.264 / MPEG4 – Part 10



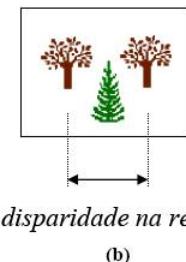
- Extensões
  - SVC – Scalable Video Coding
  - MVC – Multi-view Video Coding
- Próximo Standard – HEVC – High Efficiency Video Coding

# Video 3D

- Crescente interesse:
  - Produção de filmes
  - Jogos
  - Realidade virtual
- Sistema visual humano:
  - Noção de profundidade baseada em duas vistas do mesmo objecto (occlusão)
  - Sequência de imagens de objectos conhecidos (afastam-se da câmara)
  - Perspectiva
  - Focagem selectiva



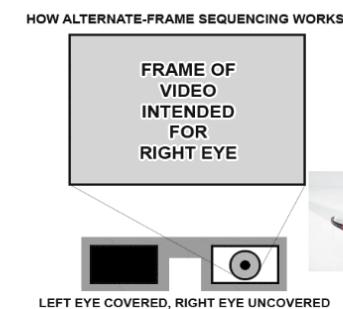
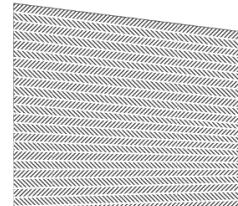
(a)



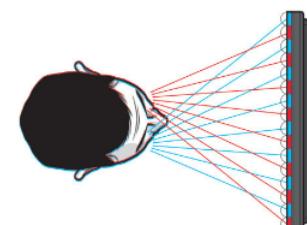
disparidade na retina  
(b)

# Video 3D

- Vídeo esteroescópico
  - Aquisição
    - 2 câmaras à distância inter-ocular
  - Codificação
  - Transmissão/armazenamento
  - Descodificação
  - Visualização
    - Ecrã esteroscópico (com óculos activos ou passivos)
      - Temporal/espatial/polarizado (Micro-polarizer)



- Ecrã autoestereoscópico (não precisa de óculos)
  - Pode depender da posição do utilizador
  - Apenas paralaxe horizontal
- Ecrã autoestereoscópico multivisão
  - Paralaxe ou micro lentes
- Ecrã holográfico/volumétrico



## Vídeo 3D

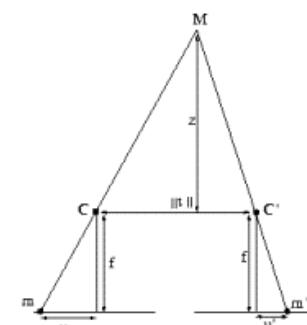
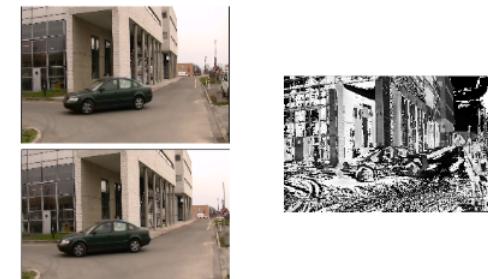
- DIBR – Depth Image Based Rendering
  - Geração de Vídeo 3D com base em conteúdos 2D
  - As imagens da esquerda e direita são geradas com base numa imagem e num mapa de profundidade



- Aquisição (2 câmaras em paralelo)
  - Com base na vista esquerda e direita gera-se o mapa de profundidade mede-se a distância horizontal entre 2 pixels correspondentes
  - Cálculo da profundidade (z):

$$d \equiv v - v' = \frac{b \cdot f}{z}$$

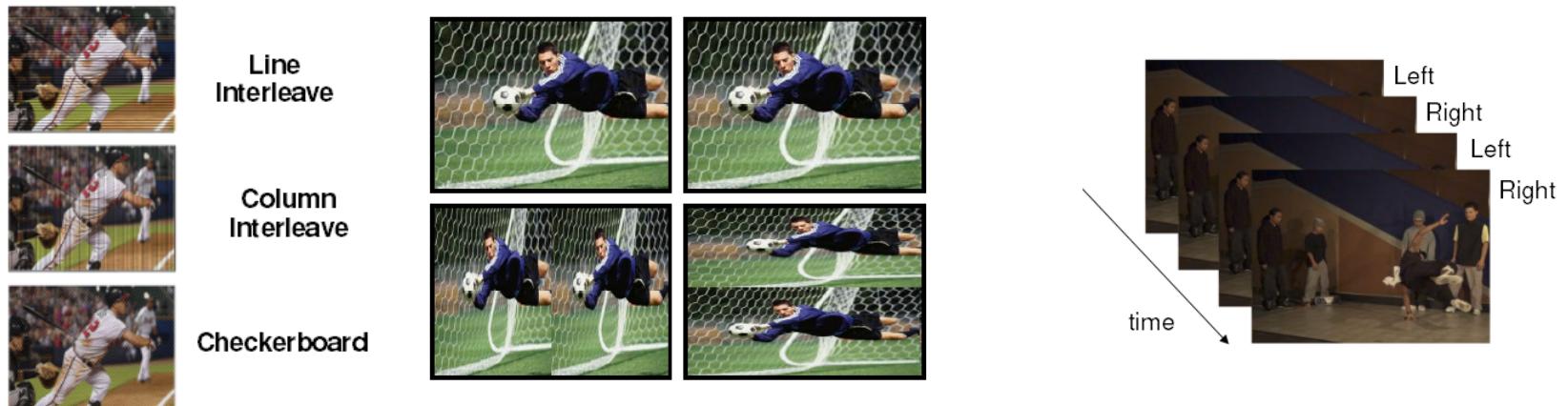
b - distância entre câmaras,  
f - distância focal  
d - disparidade



# 3D video coding

- Standard para 3D?
  - Compressão - Redundância entre as diferentes vistas
  - Compatibilidade
  - Escalabilidade
- Frame Compatibility Stereo Format
  - Multiplexar sinal 3D para o codifica-lo no formato 2D já existente

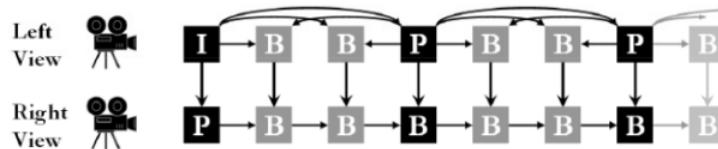
## Multiplexagem espacial versus Multiplexagem temporal



- Indústria optou pelo top/bottom ou pelo side by side

## 3D video coding

- Combinação de predição temporal e inter-visão



- A codificação 2D + profundidade (MPEG-C part 3)
  - Compatível com outros sistemas e formatos
  - Pode ser usado em visão stereo e multiview
  - O aumento de largura de banda é mínimo
  - Problema com oclusões

## 3D video coding

- Free Viewpoint Video (FVV)
  - Aquisição com N câmaras em diferentes ângulos
  - Utilizador pode navegar na cena



- MultiView Video (MVV)
  - N streams de video temporalmente sincronizados
  - O utilizador pode optar por ver uma qualquer perspectiva
  - Pode usar modos baseline e enhancement



- MVC - obriga a um Base View codificado independentemente das outras visões