

### 3

## Técnicas Utilizadas no Padrão H.264/AVC

### 3.1

#### Introdução

O H.264/AVC é o mais novo padrão internacional para compressão de vídeo, tendo sido aprovado em outubro de 2003. Ele proporciona taxas de compressão significativamente maiores que os padrões anteriores e também é conhecido como H.26L (L: *Long term*), MPEG-4 Parte 10 ou apenas H.264. O seu desenvolvimento foi motivado pelo número crescente de serviços usando vídeo, pela popularidade da TV de alta definição bem como pelo aparecimento de novos meios de transmissão, que oferecem canais com capacidade bem menor do que os usados para rádio-difusão. Estes serviços demandam um padrão com eficiência de codificação bem maior do que os existentes.

Entre as características principais deste padrão estão a de proporcionar em geral 50% de economia em termos de taxa com relação ao H.263 ou MPEG-4 Parte 2, com vídeo de alta qualidade em todas as taxas; a adaptação tanto a aplicações com baixo retardo (vídeo-conferência), quanto alto retardo (armazenamento, *streaming*); robustez a erros - ele tanto lida bem com erros de bit em transmissão sem fio ou com perdas de pacotes em redes de pacotes; além disso, possui uma separação clara entre a camada de compressão de vídeo (*video coding layer* - VCL) e a camada de redes (*network adaptation layer* - NAL).

As suas principais aplicações são: a rádio-difusão em cabo, satélite, *cable modem*, terrestre; o armazenamento serial ou interativo em dispositivos óticos ou magnéticos, DVD; os serviços conversacionais sobre redes móveis e *ethernet*; o vídeo sob demanda e etc.

### 3.2 Camada de Abstração de Rede

A VCL, a qual será descrita posteriormente, é especificada para representar, eficientemente, o conteúdo dos dados de vídeo. A NAL é especificada para formatar estes dados e fornecer informação de cabeçalho de uma maneira apropriada para as camadas de transporte ou para o armazenamento de mídia. Todos os dados ficam contidos em unidades NAL, cada qual contendo um número inteiro de bytes. A unidade NAL especifica um formato genérico para o uso em sistemas orientados a pacotes ou *bitstreams*. O formato das unidades NAL, para ambos os sistemas, são idênticos - exceto que cada unidade NAL pode ser precedida por um prefixo de começo de código em camadas de transporte orientadas a *bitstream*.

A Figura 3.1 [16] ilustra o conceito acima:

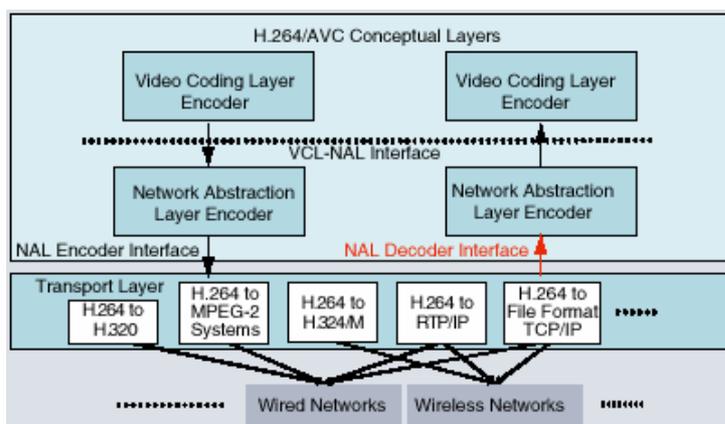


Figura 3.1: Camadas do padrão H.264

### 3.3 Camada de Codificação de Vídeo

A camada de codificação de vídeo do padrão H.264 é similar à dos outros padrões, como o MPEG-2. Ela consiste de um modelo híbrido, conforme pode ser observado na Figura 3.2 [9].

Em suma, a imagem é dividida em blocos. A primeira imagem de uma seqüência (aquela que fornece acesso aleatório) é tipicamente intra-codificada, ou seja, não utiliza informação de imagens passadas ou futuras. Cada amostra de um bloco em um quadro Intra é predita usando amostras espacialmente vizinhas de blocos codificados anteriormente. O processo de codificação escolhe a forma como as amostras são usadas para predição



### 3.3.1 Estimação de Movimento

A estimação de movimento no padrão H.264 pode ser realizada com blocos de tamanhos variáveis. Este padrão fornece maior flexibilidade na seleção do tamanho e da forma dos blocos para a estimação de movimento do que qualquer outro padrão anterior.

A Figura 5.3 [9] apresenta os sete tipos de blocos utilizados para a estimação de movimento.

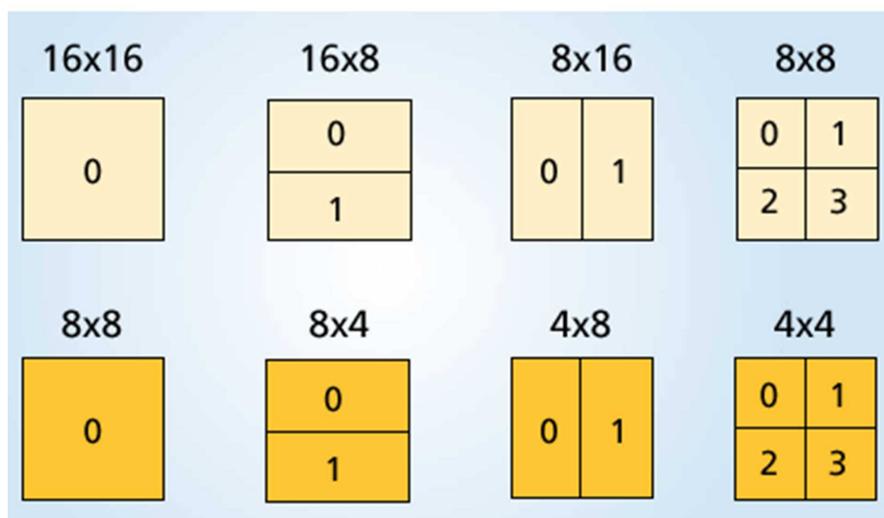


Figura 3.3: Tipos de blocos para estimação de movimento no padrão H.264

Com relação à acuracidade dos vetores de movimento, a maioria dos padrões anteriores forneciam a possibilidade de trabalhar com vetores de movimento de  $\frac{1}{2}$  pixel. Este novo desenho melhora o esquema de estimação de movimento adicionando precisão de  $\frac{1}{4}$  de pixel para o vetor de movimento, o qual também é utilizado no Perfil Avançado do padrão MPEG-4 Parte 2. Porém, o padrão H.264 conseguiu reduzir a complexidade do processo de interpolação comparado com MPEG-4 Parte 2.

Enquanto os vetores de movimento no padrão MPEG-2 são requeridos para apontar apenas as áreas dentro das imagens de referência previamente decodificadas, uma técnica de extrapolação de borda da imagem (utilizada primeiramente no padrão H.263) foi também incluída.

Uma outra inovação do padrão em questão é a utilização de múltiplos quadros para referência. Quadros codificados utilizando predição em apenas uma direção (chamados quadros P) usavam apenas um quadro anterior para prever os valores do quadro subsequente. Uma nova arquitetura estendeu, para o padrão H.264, uma técnica utilizada no padrão H.263++ para aumentar a eficiência de codificação, permitindo assim que o codificador

seleciona dentro de um grande número de quadros, que foram decodificados e armazenados no decodificador, o que melhor representará o quadro que está sendo predito. A Figura 3.4 [9] mostra o conceito acima.

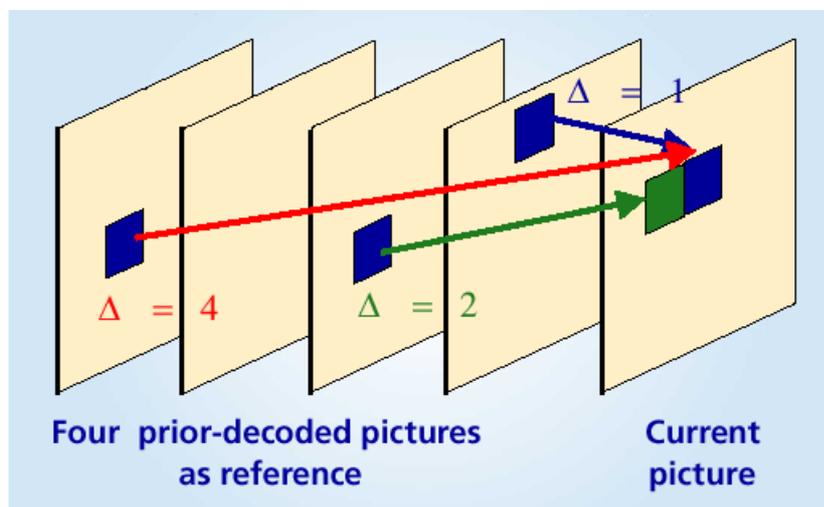


Figura 3.4: Múltiplos quadros para referência no padrão H.264

### 3.3.2 Slices e Grupos de Slice

Uma imagem pode ser dividida em uma ou mais fatias, como mostra a Figura 3.5 [16]. Dessa forma, podemos considerar que uma imagem é uma coleção de um ou mais *slices*, que representam subconjuntos que podem ser decodificados independentemente.

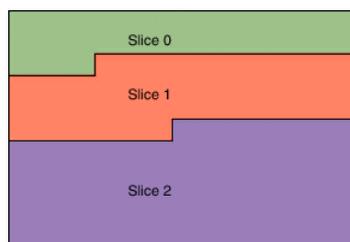


Figura 3.5: Subdivisão da imagem em *slices*

O padrão H.264/AVC também suporta uma característica conhecida como *Flexible Macroblock Ordering* (FMO), que modifica a forma como as imagens são particionadas em *slices* e macroblocos, através da utilização do conceito de grupos de *slices*. Cada grupo de *slices* é um conjunto de macroblocos definidos por um mapeamento, que incorpora um número de identificação do grupo para cada macrobloco na imagem. Cada grupo de *slices* é transmitido separadamente e pode ser particionado em uma ou mais

fatias. Usando FMO, uma imagem pode ser dividida em muitos padrões diferentes (vide Figura 3.6[16]).

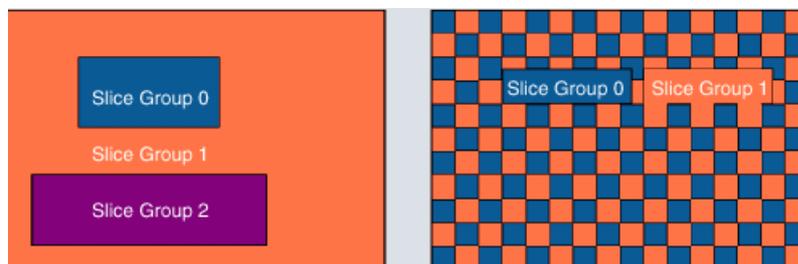


Figura 3.6: Subdivisão de uma imagem em *slices*, utilizando FMO

Quando utilizado efetivamente, FMO pode melhorar a robustez do processo à perda de dados, através do gerenciamento da relação espacial entre as regiões que são codificadas em cada *slice*.

O padrão suporta 5 diferentes tipos de codificação de *slices*: o tipo mais simples é o *slice* tipo I (Intra), onde todos os macroblocos são codificados sem se referir a outras imagens da seqüência de vídeo. Por outro lado, *slices* tipo P ou B utilizam imagens já codificadas para formar a predição para macroblocos pertencentes à fatia. Os outros dois tipos são *slices* tipos SP (*switching P*) e SI (*switching I*), e são especificados para permitir uma comutação eficiente entre *bitstreams* codificados em várias taxas de bits.

### 3.3.3 Transformada e Quantização

No H.264, a transformação é aplicada a blocos de 4x4 amostras, e ao invés da DCT, é utilizada uma transformação inteira 4x4 com as mesmas propriedades de uma 4x4 DCT. Já que a transformada inversa é definida por uma operação inteira exata, imprecisões provenientes de operações desse tipo são evitadas. Uma transformada de Hadamard adicional 2x2 é aplicada aos quatro coeficientes DC de cada componente de crominância. Se um macrobloco é codificado no modo Intra-16x16, uma transformada 4x4 similar é realizada para os 4x4 coeficientes DC do sinal de luminância.

Um parâmetro de quantização (QP - *Quantization Parameter*) é utilizado para determinar a quantização dos coeficientes transformados no H.264. O parâmetro pode tomar 52 valores. Os quantizadores são arranjados de tal forma que existe um aumento de aproximadamente 12,5% no passo de quantização, quando incrementamos o QP de 1. Os coeficientes transformados e quantizados de um bloco sofrem varredura na forma zig-zag e são transmitidos usando métodos de codificação por entropia.

### 3.3.4 Codificação por Entropia

Dois métodos de codificação por entropia são suportados. O método padrão utiliza uma simples tabela de palavras-código para todos os símbolos a serem codificados, exceto para os coeficientes transformados e quantizados. Para a transmissão dos coeficientes transformados e quantizados, um método mais eficiente chamado *Context Adaptive Variable Length Coding* (CAVLC) é empregado. Nesse esquema, tabelas VLC para vários símbolos da sintaxe são comutadas dependendo dos que já foram transmitidos. Considerando que tabelas VLC são projetadas para explorar a estatística condicional correspondente, o desempenho da codificação é melhorado em comparação com esquemas que utilizam uma simples tabela VLC. No método CAVLC, o número de coeficientes quantizados diferentes de zero e o tamanho real e posição dos coeficientes são codificados separadamente.

A eficiência da codificação pode ser aumentada se o *Context-Adaptive Binary Arithmetic Coding* (CABAC) for utilizado. Uma propriedade importante do CABAC é sua modelagem contextual. As estatísticas referentes aos elementos sintáticos já codificados são utilizadas para estimar as probabilidades condicionais. Essas probabilidades condicionais são utilizadas para comutar entre diversos modelos probabilísticos estimados. No H.264 o operador de codificação aritmética e sua associada estimação de probabilidade são especificados como métodos de baixa complexidade já que usam apenas deslocamentos e consultas as tabelas. Comparado com o CAVLC, a codificação CABAC tipicamente provê uma redução na taxa de bit entre 5 e 15%. Os maiores ganhos são obtidos quando se codifica sinais de TV entrelaçados.

### 3.3.5 Filtro *In-Loop Deblocking*

Uma característica peculiar da codificação baseada em blocos é a produção acidental do efeito de blocagem, do inglês *blocking*, que são blocos visíveis e com bordas bem marcadas. Esse efeito é considerado um dos mais visíveis problemas que se apresentam nos métodos de compressão, que usam transformadas, e pode ser compreendido facilmente, já que as bordas dos blocos são tipicamente reconstruídas com menos precisão do que regiões no interior do bloco. Por essa razão, o padrão H.264/AVC define um filtro adaptativo, onde o nível de filtragem é controlado pelos valores de vários elementos da sintaxe. O efeito de blocagem é reduzido sem afetar

muito o contraste do conteúdo; conseqüentemente, a qualidade subjetiva é aumentada. Ao mesmo tempo, o filtro reduz taxas de bits tipicamente entre 5 e 10% enquanto reproduz a mesma qualidade objetiva dos vídeos não filtrados.

A idéia básica por trás do filtro baseia-se na diferença entre as amostras localizadas perto dos limites dos blocos. Quando são mensuradas diferenças relativamente grandes entre as amostras, é bastante provável a ocorrência do efeito de blocagem; dessa forma, deve-se filtrar a imagem. Entretanto, se a magnitude dessa diferença é tão grande que não pode ser explicada pela quantização utilizada na codificação, é bastante provável que essa borda reflita o comportamento real da imagem de origem e não deve ser suavizada.

Na Figura 3.7 [9] apresentamos um exemplo da utilização deste filtro:



Figura 3.7: (a) Sem filtro *deblocking* (b) Com filtro *deblocking*

### 3.4 Perfis e Níveis

O padrão H.264/AVC possui os seguintes conjuntos, ou perfis, os quais especificam classes de aplicações:

- *Baseline*: A princípio este perfil foi desenhado para aplicações que não necessitam de alta carga computacional, como nas aplicações de vídeo conferência.
- *Main*: Desenhado para ser o principal perfil das aplicações de transmissão e armazenamento, porém o perfil *High*, elaborado posteriormente, assumiu este papel.
- *Extended*: Possui como objetivo as aplicações de *streaming* que exijam alta capacidade de compressão.

- *High*: Tornou-se o perfil para transmissão e aplicações de armazenamento, incluindo sinais de alta definição (HDTV).
- *High 10*: Perfil que adiciona a possibilidade de codificação de amostras de vídeo com mais de 10 bits.
- *High 4:2:2*: A princípio foi desenhado com o objetivo de utilização para as aplicações profissionais que utilizam vídeo entrelaçado. Este perfil contrói uma camada em cima do perfil *High 10* adicionando o suporte ao formato de croma 4:2:2.
- *High 4:4:4 Predictive*: Este perfil constrói uma camada em cima do perfil anterior (*High 4:2:2*) adicionando o suporte de amostragem de croma no formato 4:4:4, codificação de amostras com 14 bits.

Os níveis do padrão H.264/AVC representam os parâmetros que podem ser utilizados em termos de:

- Número máximo de macroblocos por segundo.
- Tamanho máximo do quadro, em macroblocos.
- Taxa de bits máxima para codificação de vídeo.
- Resoluções permitidas.
- Taxa de quadros por segundo.

Esses níveis para o padrão H.264/AVC são nomeados como: 1, 1b, 1.1, 1.2, 1.3, 2, 2.1, 2.2, 3, 3.1, 3.2, 4, 4.1, 4.2, 5 e 5.1.