

4 H.264

Foi desenvolvido pelos grupos MPEG (Moving Picture Expert Group) e VCEG (Video Coding Expert Group), através de um esforço conjunto denominado JVT (Joint Video Team), um novo padrão de codificação de vídeo denominado AVC (Advanced Vídeo Coding). Este padrão foi lançado como a Parte 10 do MPEG-4 e como Recomendação ITU-T H.264 e fornece uma capacidade de compressão maior de vídeo se comparado com os padrões anteriores (MPEG-4 e H.263).

Neste capítulo, será abordado alguns conceitos sobre H.264 relevantes para este trabalho como, por exemplo, tipos de predição e slices. Além disso, será caracterizado codificadores e decodificadores H.264, perfis deste padrão de codificação e suas principais áreas de utilização e parâmetros para a medida de desempenho de codificadores.

4.1 Conceitos importantes

Campos são linhas de ordem par ou ímpar presentes em um quadro de uma seqüência de vídeo entrelaçada.

Um quadro codificado apresenta um número de quadro (sinalizado no bitstream) que não é necessariamente relacionado com a ordem de decodificação e cada campo do quadro tem uma ordem de contagem associada, que define a ordem de decodificação dos campos. Quadros previamente codificados, chamados de figuras ou quadros de referência, podem ser usados para inter predição, cujo conceito será apresentado adiante, de quadros que serão posteriormente codificados. Figuras de referência são organizadas em uma ou duas listas, descritas como listas 0 e 1. A lista 0 é constituída por quadros de referência anteriores na ordem temporal ao quadro atual. A lista 1 é constituída por quadros de referência posteriores na ordem temporal ao quadro atual.

Uma figura codificada é composta por macroblocos, cada uma contendo amostras de luminância 16 x 16 e amostras de crominância associadas (amostras 8x8 Cb e 8x8 Cr no padrão atual). Dentro de cada figura, macroblocos são organizados em slices (pedaços), onde um slice é um conjunto de macroblocos que estão na ordem de varredura (raster scan), mas não necessariamente contínuo. Um slice I pode conter somente macroblocos do tipo I, um slice P pode conter macroblocos do tipo P e I e um slice B pode conter macroblocos do tipo B e I. Existem dois outros tipos de slices, SI e SP, que não serão apresentados aqui pois são utilizados em aplicações de streaming de vídeo e que, portanto, fogem do escopo deste trabalho.

Macroblocos I são aqueles cuja predição é do tipo intra. Esta ocorre a partir de amostras decodificadas no slice atual. Uma predição é realizada por um completo macrobloco ou por cada bloco 4x4 de amostras de luminância (e amostras de crominância associadas) no macrobloco.

Macroblocos P são aqueles cuja predição é do tipo inter. Esta ocorre a partir de figuras de referência. Um macrobloco intercodificado pode ser dividido em partições de macrobloco, isto é, blocos de luminância de tamanhos 16x16, 16x8, 8x16 ou 8x8 (e amostras de crominância associados). Se o tamanho da partição escolhida é 8x8, cada sub-macrobloco 8x8 pode ser dividido em partições de sub-macrobloco de luminância de tamanhos 8x8, 8x4, 4x8 ou 4x4 (e amostras de crominância associadas). A predição de cada partição de macrobloco pode ocorrer a partir de uma figura na lista 0. Se presente, a predição de toda partição de sub-macrobloco em um sub-macrobloco é feita a partir de uma mesma figura na lista 0.

Macroblocos B são aqueles cuja predição é do tipo inter. Esta ocorre a partir de figuras de referência. Neste caso, a predição de cada partição de macrobloco pode ocorrer a partir de uma figura na lista 0 e/ou uma figura na lista 1. Se presente, a predição de toda partição de sub-macrobloco em um sub-macrobloco é feita a partir de uma ou duas figuras de referência, uma figura na lista 0 e/ou uma figura na lista 1.

4.2 Codificadores e Decodificadores H.264:

Um codificador e um decodificador H.264 apresenta os elementos mostrados nas figuras 27 e 28. Com exceção do filtro deblocking, a maioria dos elementos funcionais básicos (predição, transformação, quantização, codificação entrópica) estão presentes nos padrões anteriores (MPEG-1, MPEG-2, MPEG-4, H.261, H.262). A forma de implementação de cada bloco funcional é que muda no H.264.

4.2.1 Codificador H264:

O codificador apresentado na Figura 25 possui um fluxo direto de dados e um fluxo de dados de reconstrução ou realimentação. A seguir será apresentado os passos usados no fluxo direto de dados da codificação de um quadro de vídeo.

Um quadro de entrada é processado em unidades de macrobloco. Cada macrobloco é codificado em modo intra ou inter e, para cada bloco no macrobloco, uma predição, denotada por P na figura, é formada baseada nas amostras de reconstrução das figuras. No modo intra, P é formado a partir de amostras no slice corrente que tem codificação, decodificação e reconstrução prévia. No modo inter, P é formado por predição por compensação de movimento a partir de uma ou duas figuras de referência, selecionadas a partir do conjunto de listas 0 e 1 de figuras de referência. Os detalhes das etapas de estimação e compensação de movimento do padrão H.264 serão abordados no item 5.2.

A predição P é subtraída do bloco corrente para produzir um bloco residual (diferença) D_n e vetores de movimento correspondentes. D_n é transformado (usando uma transformada de bloco) e quantizado para originar X, um conjunto de coeficientes transformados e quantizados que são reordenados e codificados entropicamente. Os vetores de movimento correspondentes a cada bloco também são codificados entropicamente. As etapas de transformação, quantização e codificação entrópica foram apresentadas no item 2.4. Os coeficientes codificados, junto com a informação requerida para decodificar cada bloco dentro do macrobloco (modo de predição, parâmetro de quantização, informação de

vetores de movimento) formam o bitstream comprimido, que é transmitido ou armazenado.

Além de codificar e transmitir cada macrobloco, o codificador decodifica (reconstrói) este macrobloco para fornecer uma referência para futuras previsões. Os coeficientes X são escalonados (operação inversa da quantização) e transformados inversamente para produzir o bloco diferença D_n . O bloco de previsão P é adicionado a D_n para criar um bloco reconstruído uF_n (uma versão decodificada do bloco original; u indica sem filtro ou unfiltered). Um filtro é aplicado para reduzir os efeitos da distorção do bloco e a figura de referência reconstruída é criada a partir de uma série de blocos F_n .

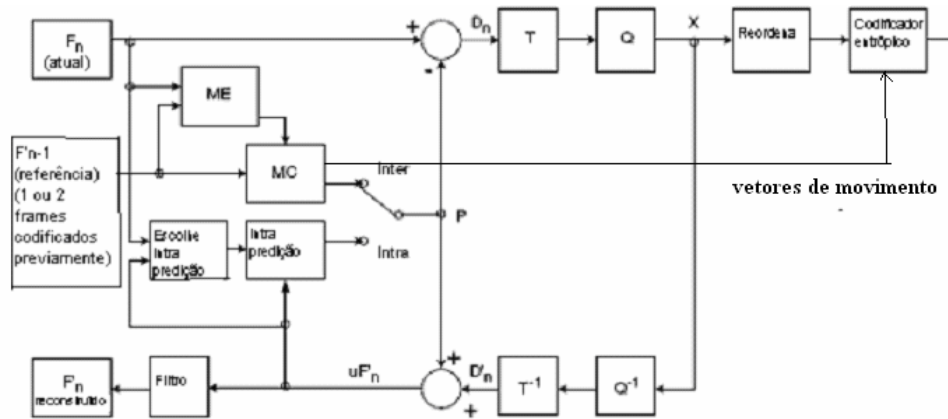


Figura 27- Codificador H264

Fonte: Ref [1]

4.2.2 Decodificador

Com o objetivo de apresentar as semelhanças entre a codificação e a decodificação, o fluxo de dados no decodificador é mostrado da direita para esquerda.

O decodificador recebe um bitstream comprimido e decodifica entropicamente os elementos de dados para produzir um conjunto de coeficientes quantizados X . Estes são escalonados e transformados inversamente para produzir D_n (idêntico a D_n mostrado no codificador). Usando o cabeçalho de informação decodificado a partir do bitstream, o decodificador cria um bloco de previsão P ,

idêntico à predição original P formada no codificador. P é adicionado à Dn' para produzir uFn' que é filtrado para criar cada bloco decodificado F_n' .

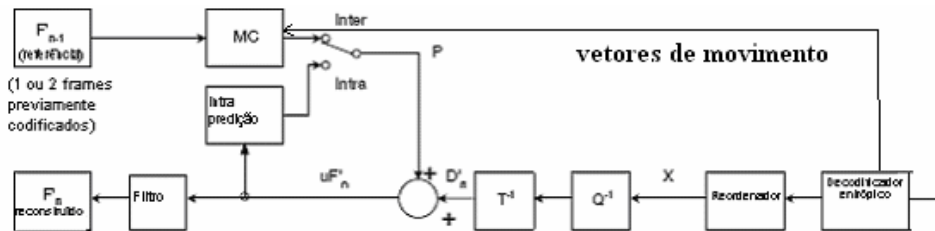


Figura 28- Decodificador H.264

Fonte: Ref [1]

4.3 Perfis do H.264

O padrão H.264 apresenta 6 perfis (profiles), diferenciando-se uns dos outros pelos tipos de aplicações e requisitos específicos de cada codificador/decodificador.

O perfil Baseline usa codificação intra e inter (usando slices I e P) e codificação entrópica com CAVLC [1] (context-adaptive variable-length codes). O perfil Main inclui suporte para vídeo entrelaçado, inter codificação usando slices B e codificação entrópica usando CABAC [1] (context-based arithmetic coding). O perfil Extended não suporta vídeo entrelaçado ou CABAC, mas adiciona modos para realizar de forma eficiente um "chaveamento" entre bitstreams codificados, usando slices do tipo SP e SI. O perfil High (HiP) apresenta controle dos parâmetros de quantização (QP) Cb e Cr separados e transformada adaptativa 8x8 e 4x4. O perfil High 10 (Hi10P) tem precisão de 10 bits por amostra. O perfil High 4:2:2 (Hi422P) tem precisão de 10 bits por amostra e suporte para formatos de amostras de croma 4:2:2. Aplicações potenciais do perfil Baseline incluem videotelefonia, videoconferência e comunicações wireless. Aplicações potenciais do perfil Main incluem televisão broadcasting e vídeo armazenado. O perfil Extended pode ser particularmente útil para aplicações de mídia streaming (por exemplo, rádio, TV ou evento ao vivo ou pré-gravado na internet.). O perfil High é usado para aplicações de armazenamento de vídeo e broadcast, particularmente aplicações de televisão de alta definição. O perfil High 10 vai além da capacidade dos equipamentos dos consumidores atualmente. O

perfil High 4:2:2 tem por objetivo atender a aplicações profissionais que usam vídeo entrelaçado.

Contudo, cada perfil tem flexibilidade suficiente para suportar uma ampla variedade de aplicações e, portanto, os exemplos de aplicações apresentados não devem ser considerados definitivos.

4.4

Modos de inter predição no H.264

Como descrito no item 3.6, o padrão H.264 usa um tamanho de bloco adaptativo na compensação de movimento. De acordo com o tamanho do bloco (em pixel), são definidos 8 modos de predição como mostrado na Figura seguinte:

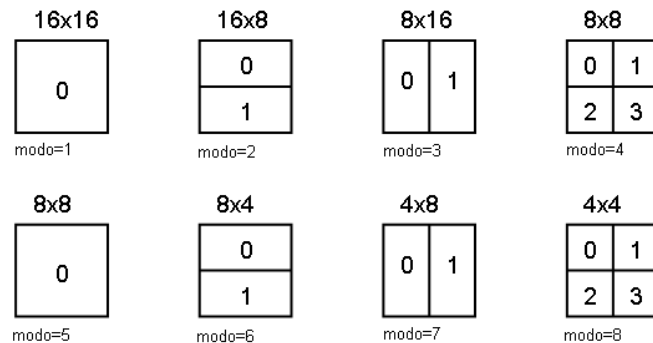


Figura 29- Diferentes tamanhos de partição em um macrobloco e modos de predição no H.264

Fonte: Ref [1]

4.5

Medidas de desempenho

A seguir iremos apresentar duas medidas objetivas de desempenho de CODEC H264/AVC que foram usadas em nossa pesquisa: desempenho quanto a taxa-distorção e desempenho computacional.

4.5.1

Desempenho quanto à taxa-distorção

Medidas da taxa de bits e PSNR em uma variedade de parâmetros de quantização fornecem uma estimativa do desempenho de compressão, sendo este mais preciso numericamente do que comparações subjetivas.

Idealmente, o codificador deve representar a sequência de vídeo usando o mínimo possível de bits com a maior fidelidade possível. Esses dois objetivos (eficiência de compressão e alta qualidade) são geralmente conflitantes, pois uma baixa taxa de bits implica em redução na qualidade da imagem no decodificador, medida por Y-PSNR. A relação entre esta taxa de bits e a qualidade é usualmente denominada teoria R-D, em função do nome na língua inglesa para taxa-distorção ser rate-distortion.

Com o objetivo de exemplificar o conceito de desempenho quanto a taxa-distorção, algumas comparações do desempenho do MPEG-4 Visual e do H.264 para duas sequências de vídeo estão presentes aqui e foram extraídas de [1].

Figuras 30 e 31 comparam o desempenho do MPEG-4 Visual e do H.264 (perfil Baseline, 1 quadro de referência) para as sequências ‘Office’ e ‘Grasses’. (Figuras 32 e 33, respectivamente). Note que a sequência ‘Office’ é mais fácil de comprimir que a ‘Grasses’ pois, a uma dada taxa de bits, o valor de PSNR de ‘Office’ é significativamente mais alto do que ‘Grasses’. O H.264 supera a compressão MPEG-4 em todas as taxas de bits testadas em termos de taxa-distorção, sendo que a sequência ‘Office’ apresenta esta diferença mais nítida.

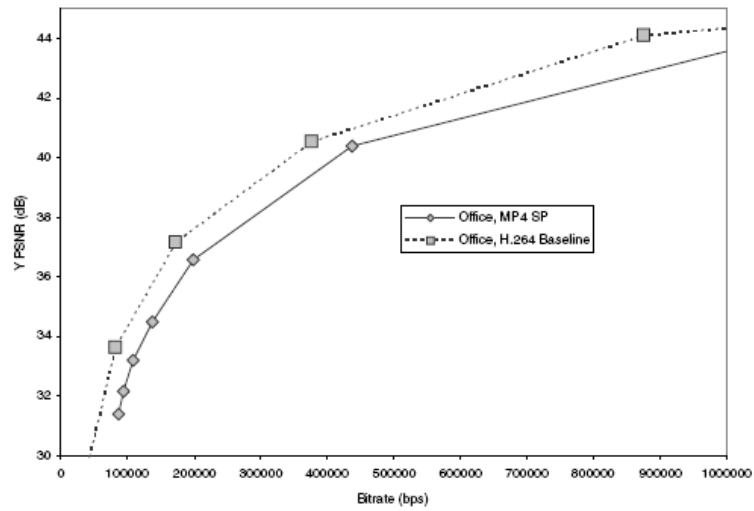


Figura 30- Comparação do desempenho quanto à taxa-distorção para a seqüência 'Office', CIF

Fonte: Ref [1]

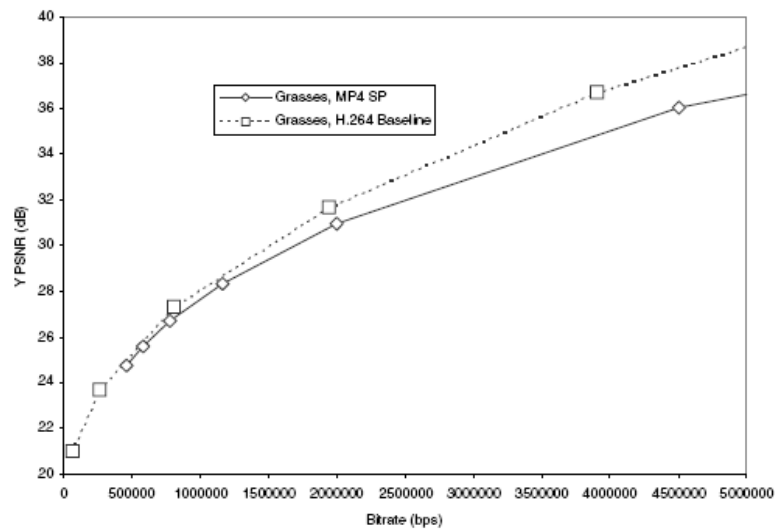


Figura 31- Comparação do desempenho quanto à taxa-distorção para a seqüência 'Grasses', CIF

Fonte: Ref [1]



Figura 32- Frame original da seqüência 'Office'
Fonte: Ref [1]



Figura 33- Frame original da seqüência 'Grasses'

Fonte: Ref [1]

4.5.2 Desempenho computacional

H.264 apresenta vários modos de codificação que detêm a potencialidade de melhorar o desempenho de compressão. Por exemplo, o perfil Main do H.264 é capaz de fornecer melhor eficiência de compressão do que o perfil Baseline. Além disso, se for considerado um perfil específico, um CODEC pode ter opcionalmente alguns de seus parâmetros de codificação habilitados ou não e, portanto, terá uma variação no desempenho computacional.

O aumento da eficiência de codificação está diretamente relacionado ao aumento do custo computacional. Além disso, este mesmo custo computacional e a eficiência de codificação depende muito da fonte de imagem. Em uma aplicação prática, a escolha dos possíveis modos de codificação depende das limitações da plataforma de processamento e, portanto, seria necessário escolher parâmetros de codificação para ajustar a fonte de imagem aos recursos de processamento disponíveis.

O exemplo dado a seguir, extraído de [1], mostra o desempenho computacional, em termos de tempo de processamento, e o desempenho em termos de taxa-distorção, em termos Y-PSNR e taxa de bits, de acordo com os parâmetros do CODEC H264/AVC.

Exemplo:

Os primeiros 25 quadros da sequência ‘Violin’ (Figura 34), que encontra-se no formato QCIF, com uma taxa de 25 quadros por segundo, foram codificados usando o software de referência H.264, em sua versão JM4.0, com um parâmetro de quantização (QP) fixo de 36. A sequência foi codificada com uma variedade de parâmetros de codificação com o objetivo de investigar os efeitos de cada um destes parâmetros no desempenho de compressão e no tempo de codificação. Duas configurações de referência foram usadas como segue:

Configuração básica: codificação entrópica CAVLC, sem quadros B (somente quadros I e P), loop filter [1] habilitado, rate-distortion optimisation (indicado pelo parâmetro RDOptimization, no qual o codificador repetidamente

codifica cada macrobloco de diferentes maneiras a fim de encontrar os melhores parâmetros de casamento) desabilitado, um quadro de referência para compensação de movimento, todos os tamanhos de blocos disponíveis.

Configuração Avançada: codificação entrópica CABAC, todo segundo quadro codificado como quadro do tipo B, loop filter [1] habilitado, rate-distortion optimisation (parâmetro RDOptimization) habilitado, cinco quadros de referência para compensação de movimento, todos os tamanhos de blocos disponíveis.

A configuração básica é apropriada para um sistema de baixa complexidade e CODEC de tempo real, enquanto que a configuração avançada é apropriada para um CODEC de alta complexidade e alta eficiência. Tabela 3 mostra os resultados obtidos. O PSNR da luminância (qualidade objetiva) de cada sequência é quase idêntica e as diferenças no desempenho ocorrem na taxa de bits e no tempo de codificação.

Configuração	Y-PSNR médio(dB)	Taxa de bits codificados (P/B slices)	Tempo de Codificação (segundos)
Básico	29,06	45,9	40,4
Básico + tamanho de bloco mínimo 8x8	29,0	46,6	33,9
Básico + 5 quadros de referência	29,12	46,2	57,2
Básico + taxa-distorção ótima	29,18	44,6	60,5
Básico + 2º quadro codificado como B	29,19	42,2	55,7
Básico + CABAC	29,06	44,0	40,5
Avançado	29,57	38,2	180
Avançado (somente 1 quadro de referência)	29,42	38,8	77

Tabela 3- Desempenho computacional do H.264 (Violin, QCF, 25 quadros)

Fonte: Ref [1]

A análise dos dados da Tabela 3 são apresentados a seguir. A configuração básica leva 40 segundos para codificar a sequência e produz uma taxa de bits de 46 Kbps (excluindo os bits produzidos pelo primeiro slice I). Usando somente blocos de tamanho 8x8 pixels ou maior para compensação de movimento reduz-se o tempo de codificação por 6 segundos, mas aumenta a taxa de bits, como seria esperado. Usando múltiplos quadros de referência (cinco neste caso) aumenta-se o tempo de codificação por quase quatro vezes, mas resulta em um aumento na taxa de bits codificada. Adicionando rate-distortion optimisation, reduz-se a taxa de bits mas aumenta-se em 50% o tempo de codificação. Quadros B fornecem um ganho de compressão ao custo do aumento do tempo de codificação de aproximadamente 50%. O CABAC fornece uma melhoria na taxa de compressão e não aumenta o tempo de codificação.

A configuração avançada gasta quatro vezes mais tempo do que a configuração básica para codificação, mas produz uma taxa de bits 17% menor do que a configuração básica. Quando é usado somente um quadro de referência, o tempo de codificação é significativamente reduzido ao custo de um aumento na eficiência de compressão.



Figura 34- Imagem mostrando ruído em uma câmera (metade inferior)

Fonte: Ref [1]

Esses resultados mostram que, para essa sequência e para este codificador, o melhor desempenho (em termos de melhoria da eficiência de codificação e complexidade computacional) ocorre com CABAC e quadros B. Esta configuração fornece um respeitável aumento na compressão sem uma alta penalidade computacional. De modo oposto, quadros de referência múltiplos fornecem somente uma pequena melhoria e são computacionalmente dispendiosos.

Deve-se ressaltar, contudo, que os diferentes resultados obtidos nestes testes seriam esperados com outras fontes de imagem e que o modelo de codificação de referência não está otimizado para eficiência computacional.

4.6

Resumo e conclusão do capítulo

Neste capítulo, foram abordados alguns conceitos sobre H.264 relevantes para este trabalho como, por exemplo, tipos de predição e slices. Além disso, foi caracterizado codificadores e decodificadores H.264, perfis deste padrão de codificação e suas principais áreas de utilização e parâmetros para a medida de desempenho de codificadores.

Uma figura codificada é composta por macroblocos, cada uma contendo amostras de luminância 16 x 16 e amostras de crominância associadas (amostras 8x8 Cb e 8x8 Cr no padrão atual). Dentro de cada figura, macroblocos são organizados em slices (pedaços), onde um slice é um conjunto de macroblocos que estão na ordem de varredura (raster scan), mas não necessariamente contínuo. Um slice I pode conter somente macroblocos do tipo I, um slice P pode conter macroblocos do tipo P e I e um slice B pode conter macroblocos do tipo B e I. Existem dois outros tipos de slices, SI e SP. Macroblocos I são aqueles cuja predição é do tipo intra. Esta ocorre a partir de amostras decodificadas no slice atual. Macroblocos P são aqueles cuja predição é do tipo inter. Esta ocorre a partir de figuras de referência.

Um codificador e um decodificador H.264 apresenta os elementos mostrados nas figuras 27 e 28. Com exceção do filtro deblocking, a maioria dos

elementos funcionais básicos (predição, transformação, quantização, codificação entrópica) estão presentes nos padrões anteriores (MPEG-1, MPEG-2, MPEG-4, H.261, H262).

O padrão H.264 apresenta 6 perfis (Baseline, Main, Extended, High, High 10 e High 4:2:2), diferenciando-se uns dos outros pelos tipos de aplicações e requisitos específicos de cada codificador/decodificador.

H.264 usa um tamanho de bloco adaptativo na compensação de movimento. De acordo com o tamanho do bloco (em pixel), são definidos 8 modos de predição como mostrado na Figura 29.

Duas medidas objetivas de desempenho de CODEC H264/AVC foram usadas em nossa pesquisa: desempenho quanto a taxa-distorção e desempenho computacional. O primeiro é obtido através da medida da taxa de bits e PSNR em uma variedade de parâmetros de quantização. O segundo está associado ao custo computacional.