

H.264: UMA BREVE DISCUSSÃO ACERCA DA CODIFICAÇÃO DE VÍDEO

Eduardo Viana Rezende– eduardo.viana@superig.com.br
Centro Universitário de Belo Horizonte
Av. Prof. Mário Werneck, 1685 - Estoril
CEP: 30455-610 - Belo Horizonte – Minas Gerais

Martineli Priscila Correia – martineli.correia@gmail.com
Centro Universitário de Belo Horizonte
Av. Prof. Mário Werneck, 1685 - Estoril
CEP: 30455-610 - Belo Horizonte – Minas Gerais

Resumo: A recomendação da União Internacional de Telecomunicações H.264, também conhecida como MPEG-4 - Parte 10/Codificação Avançada de Vídeo (AVC) apresenta-se como um padrão de codificação de vídeo promissor. Para tanto, muitos fabricantes têm adotado este padrão aberto, uma vez que suporta as mais eficientes técnicas de compactação de vídeo disponíveis. Nesse sentido, este texto apresenta uma breve discussão acerca do processo de codificação de vídeo na recomendação do padrão H.264.

Palavras-Chave: Codificação de vídeo, H.264, MPEG-4.

1 INTRODUÇÃO

O processo de codificação e compactação do sinal de vídeo digital envolve a aplicação de um algoritmo que cria um arquivo robusto e reduzido, pronto para transmissão ou armazenamento em diferentes redes digitais (PROAXIS, 2001). Para a recuperação do arquivo compactado, aplica-se um algoritmo inverso, de maneira a remontar o sinal de vídeo original e, virtualmente, exibir um conteúdo semelhante ao sinal codificado na fonte. O tempo que um arquivo leva para ser compactado, enviado, descompactado e executado é chamado de *latência* (PROAXIS, 2001). Quanto mais complexo é o algoritmo de compactação, mais alta é a latência, dada a complexidade de processamento. O par de algoritmos que funcionam nesse conjunto é denominado de *codec* (*encoder/decoder*) (RICHARDSON, 2003).

Além disso, a compactação de vídeo também está associada à redução e remoção de dados redundantes do pacote de dados de vídeo, de forma que um arquivo de vídeo digital possa ser efetivamente enviado e armazenado de maneira eficiente (PROAXIS, 2001).

Nesse sentido, de acordo com pesquisas recentes, o *encoder* proposto pela recomendação H.264 pode reduzir o tamanho de um arquivo de vídeo digital em até 80%, se comparado ao formato JPEG, e em mais de 50%, quando comparado ao padrão MPEG-4 - Parte 2 (RICHARDSON, 2003). Isso significa dizer que, dessa maneira, o sinal de vídeo digital ocupará largura de banda e espaço de armazenamento para o arquivo de vídeo bem menor que outros formatos.

2 VISÃO GERAL SOBRE A CODIFICAÇÃO DE VÍDEO

A estrutura básica do processo de codificação implica a existência de um par de sistemas complementares: o codificador e o decodificador. O codificador é o sistema responsável pela redução da quantidade de bits do vídeo para transmissão ou armazenagem. Porém, para que o vídeo seja reproduzido, é necessário que esses bits sejam decodificados e este é o papel do decodificador. O par de sistemas codificador- decodificador é comumente chamado de CODEC (MANOEL, 2007).

Grande parte dos codificadores tem a mesma estrutura básica, sendo diferenciados por alguns recursos e aplicações extras, possíveis devido aos avanços tecnológicos, que permitiu a utilização de mais esforço computacional. A figura 1 ilustra o processo.

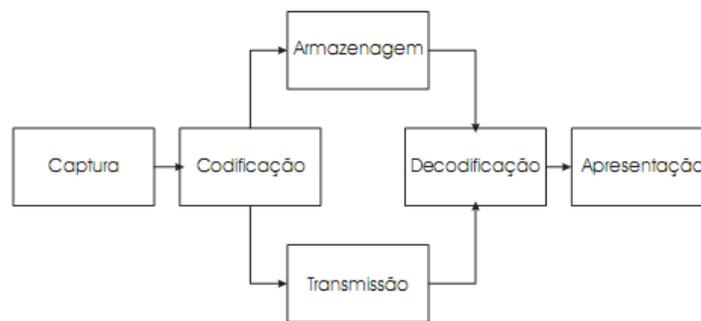


Figura 1 - CODEC contextualizado.

De forma simplificada, a primeira etapa de um codificador começa com a captura e conversão do sinal de RGB para YCbCr. Essa conversão facilita a codificação, pois descorrelaciona as cores, separando a luminância da crominância e, na maioria dos casos, ainda faz uma sub-amostragem da segunda (nos sinais 4:2:2 e 4:2:0). Como a luminância é muito mais importante, por ser mais perceptível ao olho humano, pode ser separado mais espaço para ela e menos para a crominância. Esse sinal é então amostrado e é aplicada uma transformada (a mais comum é a 8x8 DCT – do inglês, *Discrete Cosine Transform*). Depois, o sinal é compactado por entropia, tendo como resultado uma seqüência binária que representa o arquivo de vídeo.

Para fazer a decodificação deste sinal é aplicado o inverso de cada etapa na ordem inversa do processo de codificação. Exceto pela quantização que, por não poder ser um processo inversível, é feita uma aproximação desta. Existem muitos blocos em cada codificador, mas, geralmente, são utilizados os mesmos encontrados em todos os codificadores, sendo algumas características alteradas, adicionadas e outras removidas. Com o passar do tempo e o avanço tecnológico, se tornou possível utilizar técnicas muito mais sofisticadas e que exigem muito menos recursos computacionais.

A Transformada Cosseno Discreta (DCT) é importante pois muda a maneira de ver o sinal, alterando o seu domínio espacial. A DCT é uma das transformadas mais utilizada por mais se aproximar da transformada ótima, a Transformada de Karhunen-Loève (KLT) (ALCAIM, 1999). Estas transformadas modificam o domínio do sinal do tempo para frequência, concentrando a energia deste em poucos coeficientes e removendo a redundância espacial.

No caso de imagens, é aplicada uma transformada bi-direcional, já que o sinal tem duas dimensões. Qualquer que seja a transformada, ela não é aplicada diretamente sobre a imagem inteira. Primeiramente, a imagem é dividida em blocos menores. É então aplicada uma DCT uni-direcional nas colunas destes blocos e posteriormente nas linhas.

O resultado deste processo é um conjunto de $N \times N$ coeficientes que representam a imagem original no domínio da frequência. Esses coeficientes são chamados de peso das funções base da DCT. O primeiro coeficiente de cada bloco, de coordenadas horizontal e vertical nulas, é chamado de DC e representa o brilho médio deste. Os outros coeficientes são chamados de AC e muitos deles são nulos (SCHAFER, 2003).

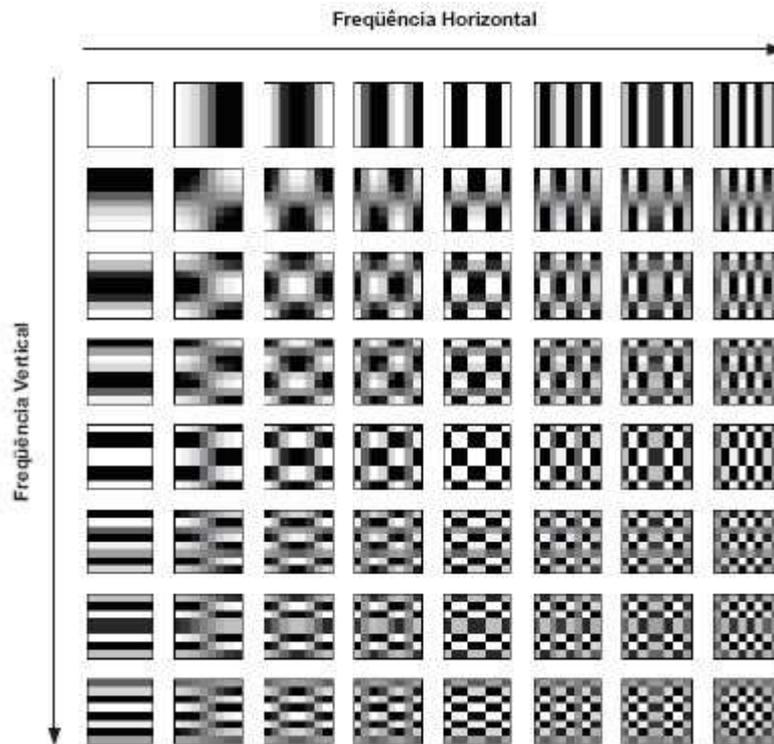


Figura 2 - Funções Base da DCT 8x8.

Posteriormente, pode-se reconstruir o sinal combinando-se as funções base da DCT da figura 2, com cada função da DCT sendo multiplicada pelo seu peso, ou coeficiente, correspondente. A DCT não reduz a quantidade de bits necessária para representar cada pixel. Na verdade ela até usa mais bits para representar cada coeficiente. A redução da quantidade de bits necessários para representar cada bloco vem do fato que a transformada concentra energia de baixa frequência e os coeficientes de alta frequência tendem a valores próximos de zero.

Assim, não são transmitidos os coeficientes com valores próximos a zero e os outros coeficientes são quantizados e codificados. Isso é fruto da correlação espacial existente entre os pixels próximos num mesmo quadro.

A figura 3 apresenta a transformada de um *macrobloco* da imagem LENA, monocromática, de 256x256 pixels.

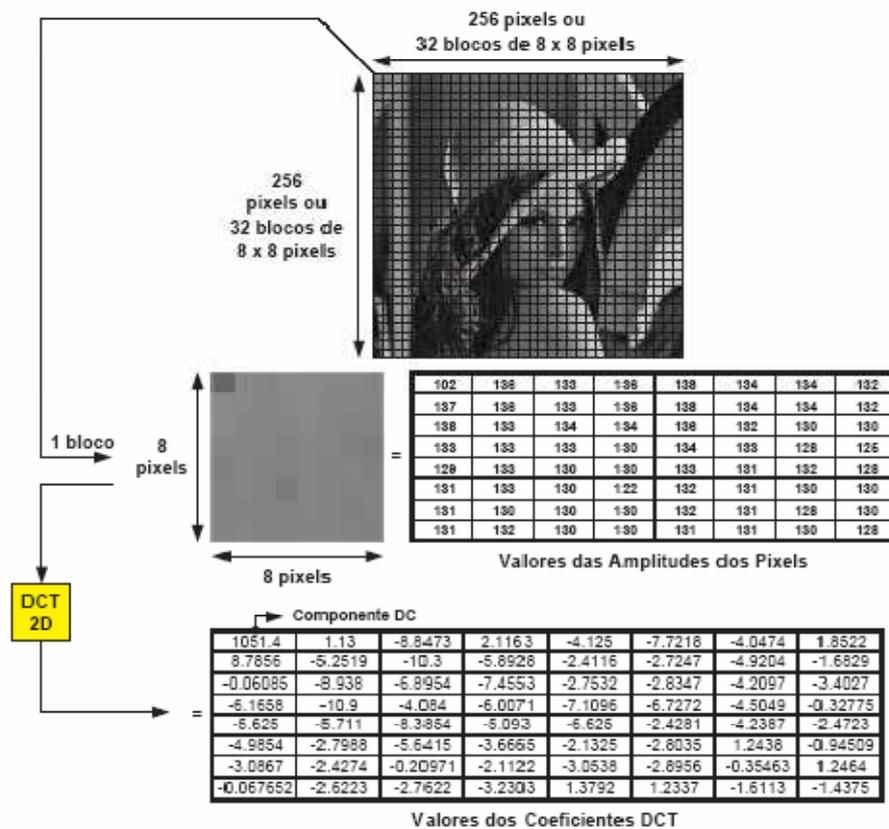


Figura 3 – Aplicação da DCT

Depois de aplicada a transformada é então feito o processo de quantização. Mais uma vez existem diversos métodos, uns mais eficientes e outros com uma qualidade do sinal de saída melhor. O objetivo da quantização é enviar menos informações (bits) para ser codificada e posteriormente decodificada. Este processo reduz a qualidade da imagem e é o único que é irreversível, então deve-se tomar muito cuidado para não alterar demasiadamente as características originais da imagem (ALCAIM, 1999).

O erro de quantização é definido como a diferença entre o sinal que entra e o sinal que sai do quantizador. Uma quantização menos precisa é mais rápida, e reduz o número de bits a ser enviado, porem aumentam o erro de quantização, aumentando também a degradação da imagem. Considerando que um vídeo é uma seqüência de imagens no tempo, estas organizadas em três tipos de quadro intra-quadro (tipo I), preditivo (tipo P) e bidirecional (tipo B). Um quadro I, ou intra-quadro, é um quadro auto-contido que pode ser independentemente decodificado, sem qualquer referência a outras imagens. Este tipo de quadro só tem codificação espacial e não depende de outros quadros para ser codificado, não existindo compensação e predição de movimentos. A primeira imagem numa seqüência de vídeo é sempre um quadro I.

O primeiro quadro de uma GOP, do tipo I, é intra codificado. Um pixel de um quadro intracodificado utiliza outros pixels da sua vizinhança para ser representado de forma mais “econômica” (gastando menos bits) e depois são codificados (RICHARDSON, 2003).

Os quadros I são necessários como pontos de partida para novas exibições ou pontos de re-sincronização, caso o fluxo de bit transmitido esteja danificado. Podem ser utilizados os quadros I para

implementar funções como avanço rápido, “rebobinar” e outras de acessos aleatórios para monitoramento de áreas de segurança (HARWOOD, 2008).

O *encoder* de vídeo comum pode inserir automaticamente quadros I a intervalos regulares, ou sob demanda, se houver a previsão de que novos clientes que queiram compartilhar o fluxo (DAMJANOVSKI, 2005). Desse modo, os quadros I devem consumir muitos bits e aumentar a largura de banda necessária para transmissão dos dados.

O quadro P, isto é quadro preditivo, faz referências às partes dos quadros I e/ou P anteriores, necessários para codificação de quadro. O quadro B, bidirecional, faz referências às partes dos quadros I e/ou P anteriores ou posteriores. A Figura 4, abaixo apresenta uma seqüência dos quadros:

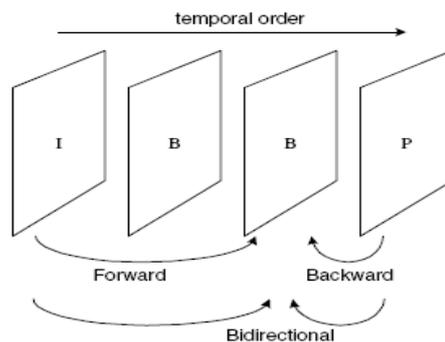


Figura 4 – Quadros de Vídeo I, B, P

3 H.264: PARÂMETROS RELEVANTES

A recomendação da União Internacional de Telecomunicações H.264 foi desenvolvida pelo *Moving Picture Experts Group and the Video Coding Experts Group* (MPEG/VCEG), em 2003, com título *Advanced Video Coding for Generic Audiovisual Services (AVC)* (RICHARDSON, 2003). Essa recomendação também foi conhecida com a parte 10 do MPEG-4.

O formato H.264 possui sete perfis, cada qual tendo como meta uma classe específica de aplicações. Cada perfil define que conjunto de recursos que o *encoder* pode utilizar e limita a complexidade de implementação do *decoder*. A Figura 5, a seguir, apresenta o diagrama em blocos do *encoder* H.264.

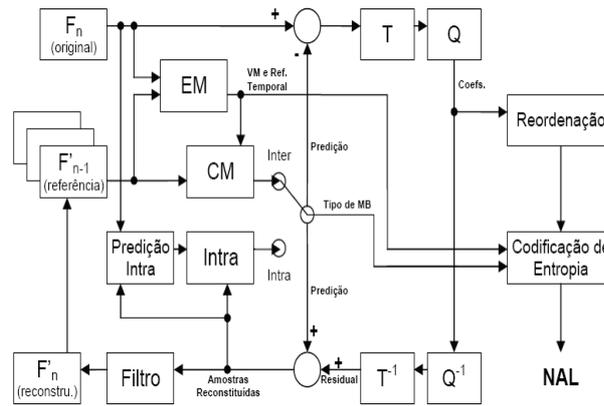


Figura 5 – Encoder H.264

Ainda assim, o H.264 possui 11 níveis ou graus de capacidade para limitar os requisitos de desempenho, largura de banda e memória. Cada nível define a taxa de bit, taxa de codificação e macro-blocos por segundo para a resolução desejada. Quanto mais alta for a resolução, maior será o nível necessário (RICHARDSON, 2003).

Nesse caso, dependendo de um perfil, o *decoder* poderá usar diferentes tipos de quadros, tais como quadros I, quadros P e quadros B. O *decoder* pode ser avaliado na Figura 6:

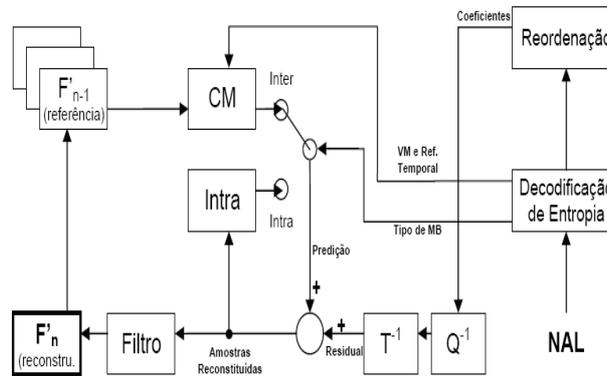


Figura 6 – Decoder H.264

A estrutura do H.264 é muito semelhante a dos seus antecessores, como o MPEG-2, como pode ser visto na figura 7, que representa os principais blocos de um codificador H.264.

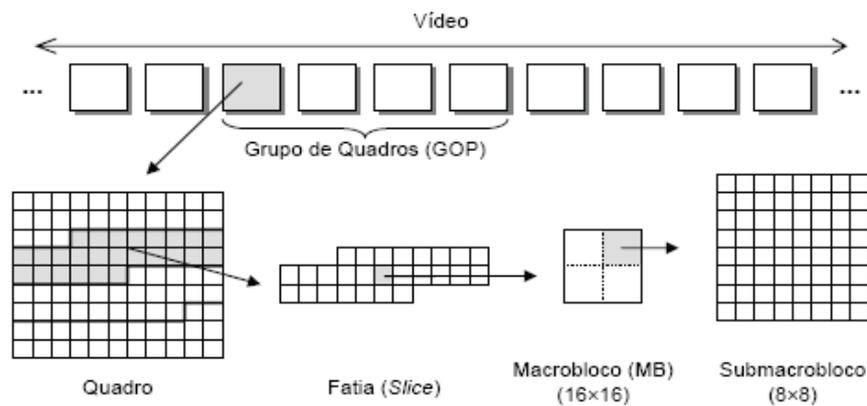


Figura 8 – Transformada Hadamard no padrão H.264.

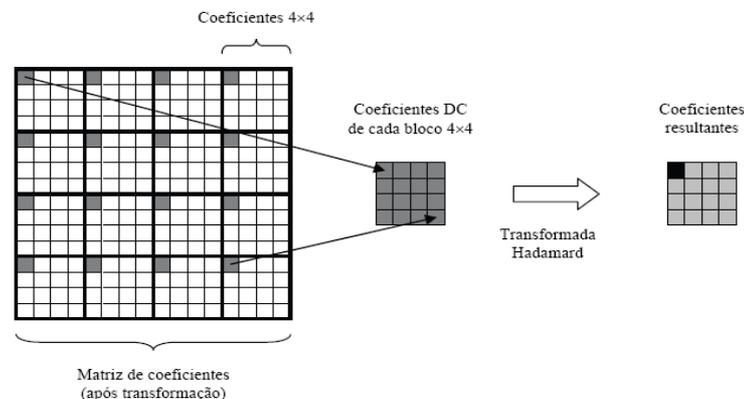


Figura 9 – Transformada Hadamard no padrão H.264.

A compensação de movimento é outro aspecto relevante e, no caso do H.264, a compensação de movimento é feita está baseada nos blocos, sendo utilizada a codificação de quadros P e B (RICHARDSON, 2003).

O *encoder* H.264 pode selecionar blocos com pixels associados, variando o tamanho e a forma do bloco. Nas áreas em que não se encontram blocos associados no quadro de referência, utilizam-se os macro-blocos intra-codificados (RICHARDSON, 2003).

No H.264 ainda foram introduzidos dois novos tipos de quadros: os quadros do tipo SI e do tipo SP. Os quadros do segundo tipo tem a quantização da predição temporal realizada no domínio da frequência.

Quadros SP em casos onde a predição temporal não pode ser utilizada para evitar erros. Depois de feitas as predições, todos os quadros sem exceção são quantizados e compactados. Diversas técnicas de compactação são aplicadas, podendo, inclusive, ser utilizada mais de uma e mais de uma vez (SCHAFER, *et al.*2003).

3 CONCLUSÕES

A recomendação H.264 apresenta técnicas que permitem uma eficiente e segura compactação de sinais de vídeo digitais. Esse método possui uma capacidade de previsão de quadros rápida, assim

como uma robusta blindagem aos erros de transmissão. Nesse sentido, o padrão oferece novas possibilidades de criação *encoders* de vídeo mais eficientes, capazes de permitir um fluxo de dados de melhor qualidade de imagem, elevadas taxas de quadros e melhor resolução.

O H.264 reincorporou uma série de recursos que haviam sido retirados do H.261 por necessitarem de muito recurso computacional e ainda inclui outras ferramentas, novas, capazes de aumentar muito o nível de compressão de uma seqüência de vídeo, se tornando muito superior aos codificadores que o antecederam e atualmente o mais eficiente. Dentre as mudanças e principais recursos destaca-se a compensação de movimento muito mais eficiente em comparação com os codificadores que o antecederam, utilizando até 32 quadros de referência. Este recurso oferece um ganho na compressão do vídeo, embora não seja adequado o seu uso em seqüências com movimentos muito rápidos e em seqüências que tenham pouca relação entre os seus quadros.

O tamanho dos blocos utilizados para compensação de movimento pode ser escolhido, variando desde 4 x 4 até 16 x 16, permitindo uma acuridade surpreendente para a predição de movimentos. A precisão desta ainda foi melhorada para ¼ de pixel, tornando-se ainda mais preciso.

O filtro anti-blocos é outro artifício retirado dos codificadores que agora foi reincorporado, oferecendo uma qualidade muito superior aos vídeos codificados em baixas taxas de bits.

Outro aspecto relevante é uma maior proteção contra erros com recursos como reordenação de blocos, o envio de alguns quadros extra (normalmente em taxas mais baixas) entre outros.

Estes e outros recursos dão ao H.264 a possibilidade de ser utilizado para muitos fins, sendo robusto a erros e oferecendo a mesma qualidade de vídeo utilizando taxas de bits muito menores.

REFERÊNCIAS

DAMJANOVSKI, Vlado. **CCTV Networking and Digital Technology**. London: Elsevier Butterworth–Heinemann, 2005.

HARWOOD, Emily. **Digital CCTV: A Security Professional's Guide Elsevier Butterworth–Heinemann**, 2008.

MANOEL, Edson T. M. UNIVERSIDADE FEDERAL DE SANTA CATARINA, **Codificação de Vídeo H.264**, 2007. Tese (Mestrado)

PROAXIS, John G. **Digital Communications**. McGraw-Hill, 2001.

RICHARDSON, Iain E. G. **H.264 and MPEG-4 Video Compression for Next Generation**. London: Wiley & Sons Ltd. 2003.

SCHAFER, Ralf; WIEGAND, Thomas; SCHWARZ, Heiko. **The Emerging H.264/AVC Standard**. 2003.