Capítulo V - Introdução ao Sistema MPEG de Codificação de Vídeo

5.1 Introdução

O grupo MPEG - *Moving Pictures Expert Group* (http://www.mpeg.org) iniciou seus trabalhos em 1988 como um grupo de trabalho da *International Standards Organization* (ISO) com o objetivo de definir padrões para compressão digital de sinais de áudio e vídeo. O grupo tomou como base o padrão para vídeo-conferência e vídeo-telefonia JPEG (*Joint Photographic Experts Group*) - hoje conhecido como H261 - o qual foi inicialmente desenvolvido para comprimir imagens estáticas, tais como em fotografia eletrônica.

O primeiro objetivo do grupo MPEG foi definir um algoritmo de codificação de vídeo para armazenamento em meio digital; em particular, CD-ROM. O padrão resultante foi publicado em 1993 e compreendia três partes:

1. Aspectos de sistema (incluindo multiplexação e sincronização)
2. Codificação de Vídeo
3. Codificação de áudio

O padrão, chamado MPEG-1, foi aplicado no sistema *Interactive CD* (CDi) para possibilitar a execução de *playback* em CDs. O sistema CDi foi largamente utilizado em aplicações PC para as quais existe uma grande gama de codificadores de *hardware* e *software*. O padrão é restrito a formatos de vídeo não entrelaçados e primeiramente objetivava a codificação de vídeo a taxas de bits não maiores do que 1.5 Mbits/s.

Nota: Em um sinal de vídeo entrelaçado a varredura das linhas da tela é tal que as linhas pares são percorridas primeiramente (*field* par) seguido da varredura das linhas ímpares (*field* ímpar). O sinal de vídeo entrelaçado visa reduzir o efeito de "cintilação" no sistema visual humano.

Em 1990 iniciaram os estudos para um segundo padrão de vídeo que objetivava ser capaz de codificar imagens entrelaçadas diretamente e, originalmente, suportar aplicações de alta qualidade a taxas de transmissão de 5 a 10 Mbit/s. Este segundo padrão, que foi denominado MPEG-2, também suporta formatos de alta definição a taxas de bits no intervalo de 15 a 30 Mbits/s. Assim como no padrão MPEG-1, o padrão MPEG-2 (publicado em 1994) compreende três partes: sistemas, vídeo e áudio.

É importante notar que os padrões MPEG especificam somente a sintaxe, a semântica das seqüências de bits e o processo de decodificação. Não especificam, no entanto, o processo de codificação, ficando livre a proposta de novas técnicas de codificação que visem melhorar o desempenho do sistema.

5.2 Princípios de Codificação de Vídeo

Uma imagem com qualidade de estúdio de 625 linhas, quando digitalizada de acordo com a ITU *Recommendation* 601/656 (isto é, amostragem 4:2:2 - ver nota abaixo)

requer uma taxa de transmissão de 216 Mbit/s para representar a luminância e duas amostras de crominância (ver Figura 5.1). Para canais de transmissão com largura de banda restrita (tais como canais de satélite) é necessário reduzir esta alta taxa para que a transmissão seja comercialmente viável.

Nota: A norma CCIR 601 [9] estabelece que uma fonte de sinal de vídeo colorido deve ter 3 componentes: uma componente de luminância Y (intensidade luminosa) e duas componentes de crominância C_b e C_r (que combinadas, definem a tonalidade da cor). Existem duas opções para resoluções espaciais desta norma. A primeira opção (adequada para o sistema de televisão NTSC) usa 525 linhas por *frame* (quadro) e 60 *frames* por segundo. O *frame* de luminância tem 720 × 480 pixels e cada *frame* de crominância tem 360 × 480 pixels. A segunda opção (adequada para o sistema de televisão PAL-M) usa 625 linhas por *frame* e 50 *frames* por segundo. O *frame* de luminância tem 720 × 576 pixels e cada *frame* de crominância tem 360 × 576 pixels.

Esta forma de amostragem do sinal de vídeo é denominada **formato de sub-amostragem 4:2:2**.

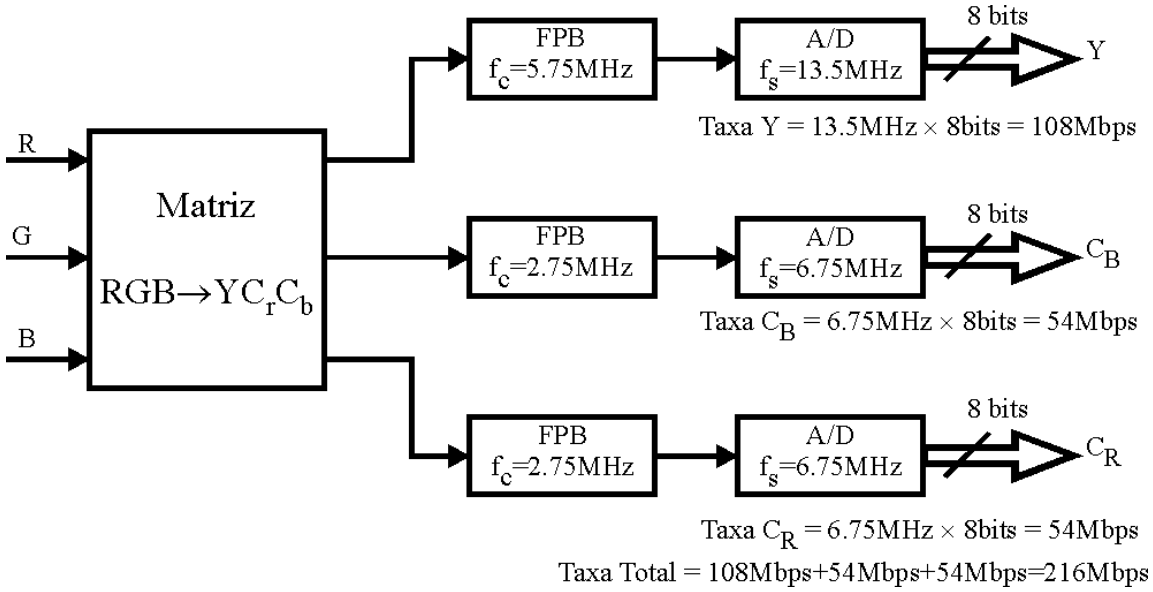

Figura 5.1: Amostragem 4:2:2. A conversão $RGB \rightarrow YC_rC_b$ obedece o mapeamento $Y = 0.25R + 0.5G + 0.5B$, $C_b = 0.25(B - Y) + 0.5$ e $C_r = 0.3125(R - Y) + 0.5$, sendo R,G e B respectivamente a intensidade das componentes *Red*, *Green* e *Blue* do sinal de vídeo da câmera. f_s é a freqüência de amostragem do A/D e f_c é a freqüência de corte do filtro passa baixa (FPB).

Um sistema para compressão de vídeo objetiva reduzir a taxa de transmissão e opera removendo a redundância e/ou informações de menor importância do sinal antes da transmissão. Tal sistema é implementado pelo Codificador de Fonte de um transmissor de vídeo digital. No receptor, o Decodificador de Fonte reconstrói uma aproximação da

imagem a partir da informação remanescente após o processo de compressão. Em sinais de vídeo, três tipos distintos de redundância podem ser identificados:

- Redundância temporal e espacial: valores de pixels não são independentes, mas são correlacionados com seus vizinhos, tanto dentro do mesmo *frame* (redundância espacial) quanto entre *frame*s consecutivos (redundância temporal). Assim, dentro de alguns limites, o valor de um pixel pode ser predito a partir dos valores dos pixels vizinhos assim como regiões de um *frame* futuro podem ser preditas a partir do *frame* atual.

- Redundância em entropia: para qualquer sinal digitalizado não-aleatório alguns valores codificados ocorrem mais freqüentemente que outros. Esta característica pode ser explorada através da codificação dos valores que ocorrem mais freqüentemente com códigos menores, enquanto que códigos maiores podem ser usados para valores mais raros em ocorrência (Codificação por Entropia - Código de Huffman).

- Redundância psico-visual: Esta forma de remoção de redundância resulta do princípio de funcionamento do olho e do cérebro humanos (sistema visual humano). O limite de definição fina de detalhes que o olho pode resolver (limites de resolução espacial) quanto os limites na habilidade de acompanhar imagens que se movem rapidamente (limites de resolução temporal) são utilizados como limiares para que seja descartado aquele sub-conjunto do fluxo de informação de vídeo que ultrapassa estes limites. Visto que o sistema visual humano não é capaz de perceber este tipo de informação, não há razão para que ela seja transmitida, resultando, assim, em compressão.

5.3 Algoritmos Utilizados no Sistema MPEG

O primeiro processo de compressão efetuado em um sistema MPEG é a sub-amostragem de cor. Esta sub-amostragem baseia-se na característica do sistema visual humano em que a definição fina de detalhes que o olho+cérebro pode resolver é menor para diferenciações de cores do que para diferenciações de intensidade luminosa. Os formatos de sub-amostragem aceitos pelo sistema MPEG para uma taxa de 60 *frames*/s (NTSC) são

Componente de Cor	Pixels/linha × linha (resolução espacial da imagem)		
	formato 4:2:0	formato 4:2:2	formato 4:4:4
Y	720 × 480	720 × 480	720 × 480
C_b	360 × 240	360 × 480	720 × 480
C_r	360 × 240	360 × 480	720 × 480

e para uma taxa de 50 *frames*/s (PAL-M) os formatos de sub-amostragem aceitos pelo sistema MPEG são

Componente de Cor	Pixels/linha × linha		
	formato 4:2:0	formato 4:2:2	formato 4:4:4
Y	720 × 576	720 × 576	720 × 576
C_b	360 × 288	360 × 576	720 × 576
C_r	360 × 288	360 × 576	720 × 576

Além desta redução da taxa de amostragem, o padrão MPEG inclui dois tipos diferentes de processos para explorar redundância em imagens:

- **Transformada Coseno Discreta (DCT - *Discrete Cosine Transform*):** É similar à Transformada de Fourier Discreta (DFT). O propósito de usar esta transformação ortogonal é ajudar o processamento a remover redundância espacial através da concentração de energia do sinal em relativamente poucos coeficientes.

- **Predição por Compensação de Movimento Inter-*Frames*:** é usada para remover redundância temporal. É baseada em técnicas similares à bem conhecida *Differencial Pulse-Code Modulation* (DPCM).

5.3.1 A Transformada Coseno Discreta

Por simplicidade, consideremos apenas o sinal de luminância Y de uma imagem PAL-M gerada sob amostragem 4:2:0 (Ver Figura 5.2). No sistema de codificação MPEG , a redundância espacial é removida processando os sinais digitalizados em blocos bidimensionais de 8 pixels/linha × 8 linhas.

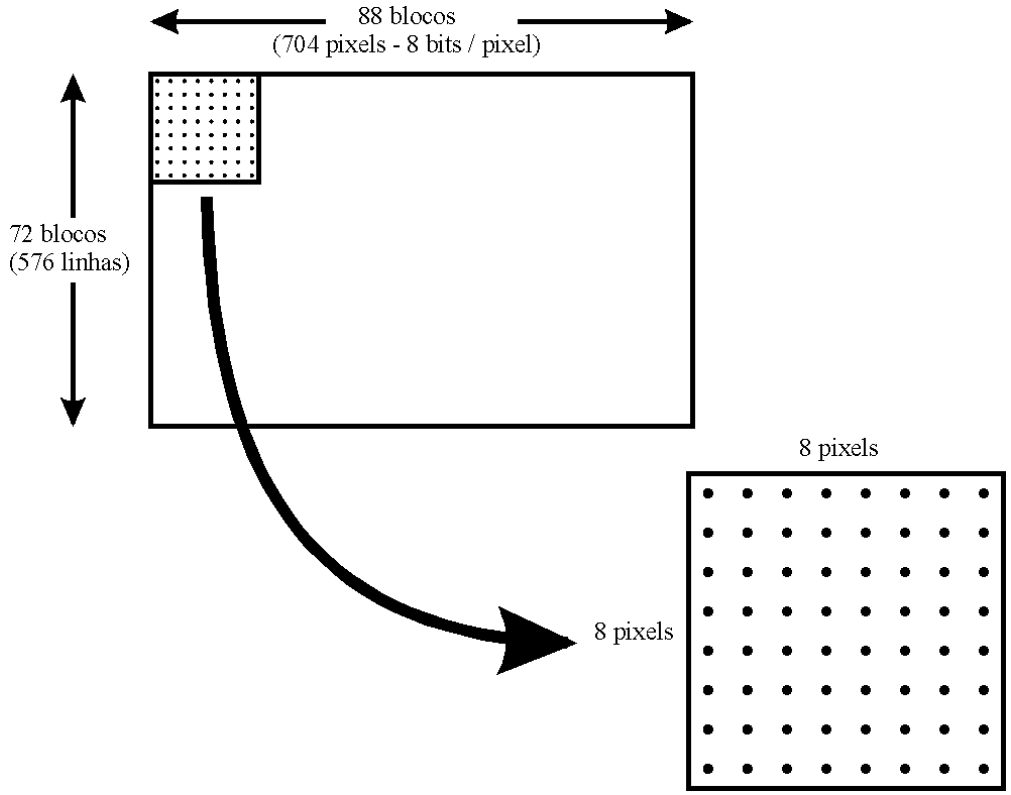

Figura 5.2: Obtenção de um bloco de 8×8 pixels a ser transformado para o domínio freqüência espacial pela DCT. Note que existe uma margem de 8 pixels à esquerda e à direita de cada *frame* de modo que cada linha possui 704 pixels.

A Transformada DCT é um processo reversível (IDCT - *Inverse Discrete Cosine Transform*) que efetua o mapeamento entre a representação de uma imagem bidimensional e a sua representação no domínio freqüência espacial. Especificamente, a imagem bidimensional é particionada em blocos de 8×8 pixels x_{ij} e a DCT é aplicada sobre cada

um dos blocos. O bloco resultante da transformação é também um bloco de tamanho 8×8, composto de coeficientes y_{kl}, conforme mostra a Figura 5.3.

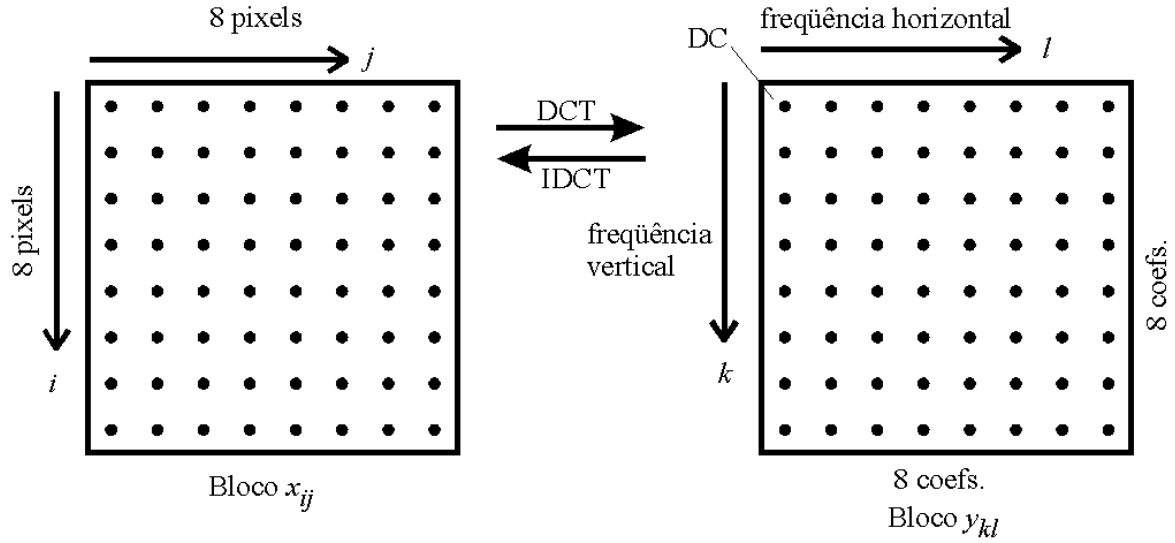

Figura 5.3: Pares de Transformadas DCT 8×8 (direta e inversa).

Cada coeficiente indica a contribuição de uma diferente função de base DCT. O coeficiente y_{00} ao alto e à esquerda do bloco, localizado na coordenada $(k,l)=(0,0)$, é chamado de coeficiente DC e representa a luminosidade média do bloco. Os coeficientes que distribuem-se ao longo da direção vertical do bloco representam as freqüências espaciais verticais. A freqüência espacial vertical aumenta com o valor de *k*. Os coeficientes que distribuem-se ao longo da direção horizontal do bloco representam freqüências espaciais horizontais. A freqüência espacial horizontal aumenta com o valor de *l*. O valor numérico de um coeficiente y_{kl} indica a intensidade conjunta da freqüência espacial vertical *k* e da freqüência espacial horizontal *l*. Um bloco resultante da DCT cujos coeficientes y_{kl} apresentam valor significativo em regiões de alta freqüência espacial horizontal *l* indica que o bloco de pixels originais x_{ij} apresenta alta diferenciação de luminosidade na direção horizontal. Um bloco resultante da DCT cujos coeficientes y_{kl} apresentam valor significativo em regiões de alta freqüência espacial vertical *k* indica que o bloco de pixels originais x_{ij} apresenta alta diferenciação de luminosidade na direção vertical.

A transformação de um bloco 8×8 x_{ij} em um bloco 8×8 y_{kl} através da DCT é denotada por $X \xrightarrow{8\times 8 DCT} Y$ e é matematicamente definida por

$$y_{kl} = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x_{ij}\cos\left(\frac{(2i+1)k\pi}{2N}\right)\cos\left(\frac{(2j+1)l\pi}{2N}\right) \tag{5.1}$$

onde $N=8$ e $k,l = 0,1,\cdots,N-1$.

A transformação de um bloco 8×8 y_{kl} em um bloco 8×8 x_{ij} através da IDCT (DCT inversa) é denotada por $Y \xrightarrow{8\times 8\,IDCT} X$ e é matematicamente definida por

$$x_{ij} = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} y_{kl}\,c(k)c(l)\cos\left(\frac{(2i+1)k\pi}{2N}\right)\cos\left(\frac{(2j+1)l\pi}{2N}\right) \tag{5.2}$$

onde $i, j = 0,1,\cdots,N-1$ e

$$c(a) = \begin{cases} 1, & a = 0 \\ 2, & a \neq 0 \end{cases} \tag{5.3}$$

A DCT não reduz diretamente o número de bits requerido para representar o bloco. Na verdade, para um bloco x_{ij} de 8×8 pixels, com 8 bits/pixel, a DCT produz um bloco y_{kl} de 8×8 coeficientes, com no mínimo 11 bits/coeficiente para permitir reversibilidade sem erros de truncamento excessivos. A redução no número de bits deriva do fato de que, para blocos típicos de imagens naturais, a distribuição dos coeficientes y_{kl} não é uniforme – a transformação tende a concentrar a energia nos coeficientes de baixa freqüência próximo ao coeficiente y_{00}. Os demais coeficientes y_{kl} possuem valor próximo a zero, podendo ser descartados. Ou seja, a redução da taxa de bits é obtida através da não transmissão dos coeficientes y_{kl} de valor próximo a zero (os quais serão zerados pelo processo de quantização do sistema MPEG) e codificação dos coeficientes remanescentes conforme será descrito abaixo. A distribuição não uniforme dos coeficientes y_{kl} é um resultado da redundância espacial presente no bloco original x_{ij}.

Muitas formas diferentes de transformações têm sido investigadas para redução de taxa de bits. As melhoras transformações são aquelas que tendem a concentrar a energia de um bloco da imagem em poucos coeficientes. A transformação com maior capacidade de concentração de energia é a transformação KLT (*Karhunen-Loève Transform*). A DCT somente perde para a KLT quando o coeficiente de correlação entre os pixels do bloco x_{ij} é menor que 0.7. Portanto, a DCT pode ser considerada como uma das melhores transformações neste sentido e tem a vantagem de que, tanto a DCT quanto seu inverso (IDCT) apresentam um custo computacional baixo. A escolha de blocos de tamanho 8×8 resulta do fato de que um bloco maior do que 8×8 não aumenta significativamente a concentração da energia nos coeficientes de baixa freqüência.

5.3.2 Quantização dos Coeficientes

Após a obtenção de um bloco y_{kl} através de transformação $X \xrightarrow{8\times 8\,DCT} Y$, os coeficientes da DCT são quantizados. A cada coeficiente é aplicada uma particular quantização, em função da freqüência espacial dentro do bloco que ele representa. O objetivo é minimizar o número de bits que devem ser transmitidos para o decodificador de tal forma que ele possa fazer a transformada inversa e reconstruir a imagem: uma quantização menos acurada irá reduzir o número de bits que precisam ser transmitidos para representar um dado coeficiente da DCT, mas aumentam o erro de quantização (ruído de

quantização) para aquele coeficiente. Note que o ruído de quantização introduzido pelo codificador não é reversível no decodificador, de tal forma que o processo de codificação e decodificação é "com perdas".

Mais erro de quantização pode ser tolerado nos coeficientes de alta freqüência porque o ruído de alta freqüência é menos visível. Ainda, o ruído de quantização é menos perceptível nos componentes de crominância do que nos componentes de luminância. Em decorrência destas particularidades do sistema visual humano, o padrão MPEG usa matrizes de peso para definir a relativa adequação da quantização dos diferentes coeficientes. Diferentes matrizes de pesos podem ser usadas para diferentes *frames* dependendo do modo de predição usado.

Os coeficientes ponderados pelas matrizes de peso são então passados através de um critério fixo de quantização, o qual é, usualmente, um critério linear. Entretanto, para alguns modos de predição há um limiar (isto é, uma zona morta) ao redor do zero. O efeito deste limiar é maximizar o número de coeficientes que são quantizados para zero. Na prática, pequenas variações ao redor do zero são usualmente causadas por ruído no sinal de vídeo, de forma que suprimindo estes valores resulta em uma aparente melhora na qualidade subjetiva da imagem e um aumento da compressão global.

O ruído de quantização é mais visível em alguns blocos do que em outros, como, por exemplo, em blocos que contêm uma borda de alto contraste entre duas áreas planas. Nestes blocos, os parâmetros de quantização podem ser modificados para limitar o máximo erro de quantização, particularmente nos coeficientes de alta freqüência.

5.3.3 Varredura zig-zag, codificação RLE e codificação VLC

Após a quantização, os blocos y_{kl} de 8×8 coeficientes quantizados são varridos em um padrão zig-zag para transformar os 64 coeficientes do bloco em uma seqüência serial de coeficientes quantizados. A varredura é feita diagonalmente do topo esquerdo superior à base direita, conforme ilustrado na Figura 5.4.

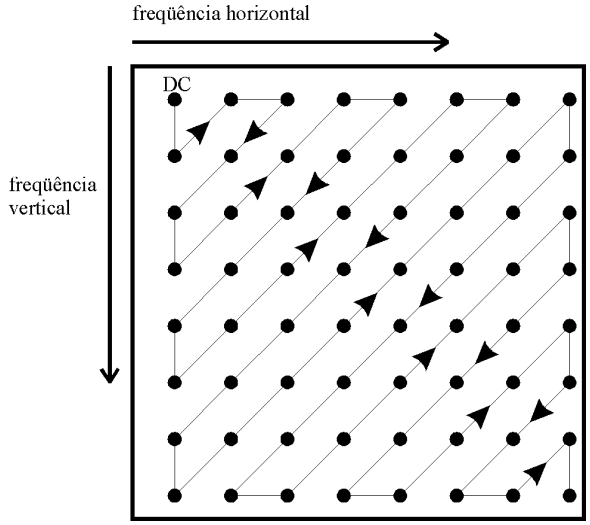

Figura 5.4: Varredura de um bloco y_{kl} e geração da seqüência de amostras para codificação RLE e codificação VLC.

Note que, resultante da quantização, ocorrerão longas seqüências de coeficientes nulos. Denomina-se de *run* cada sequência de coeficientes nulos (zeros) que precedem um coeficiente não nulo.

As seqüências de coeficientes produzidas pela varredura zig-zag são codificados pela contagem do número r_0 de coeficientes zero que precedem um coeficiente c^* não-zero, processo denominado de codificação RLE (*Run-Length Encoding*). A cada valor r_0 há um respectivo valor c^* associado. Cada par $\{r_0,c^*\}$ é então codificado usando um código de comprimento variável (VLC – *Variable Length Coding*) basicamente similar ao Código de Huffman. A codificação VLC explora o fato de que *run*s curtas são mais prováveis do que *runs* longas e que coeficientes c^* de valor pequeno são mais prováveis do que coeficientes c^* de maior valor. O VLC gera um código que tem diferentes comprimentos dependendo da freqüência esperada de ocorrência de cada par $\{r_0,c^*\}$. Combinações comuns usam palavras-códigos curtas, combinações menos comuns usam palavras-código longas, de modo idêntico ao Código de Huffman.

Todas as outras combinações que não encontram-se previstas na tabela de codificação VLC são codificadas pela combinação de um código de escape (para indicar que a tabela VLC não reconhece o par $\{r_0,c^*\}$) e dois códigos de comprimento fixo. Os dois códigos de comprimento fixo são: uma palavra de 6 bits para representar r_0 e uma palavra de 12 bits para representar c^*.

5.3.4 Controle da Taxa de Transmissão para o Canal

A quantização dos coeficientes da DCT, a codificação RLE e a codificação VLC são processos que produzem uma taxa de bits variável que depende da complexidade da imagem e da quantidade e tipo do movimento na imagem. Para produzir uma taxa de bits constante, a qual é desejável para limitar a banda ocupada no canal de transmissão, é necessário um *buffer* para suavizar as variações na taxa de bits. Para prevenir *overflow* e *underflow* deste *buffer*, a sua ocupação é controlada pelo Controle de Ocupação do Buffer, conforme mostra a Figura 5.5.

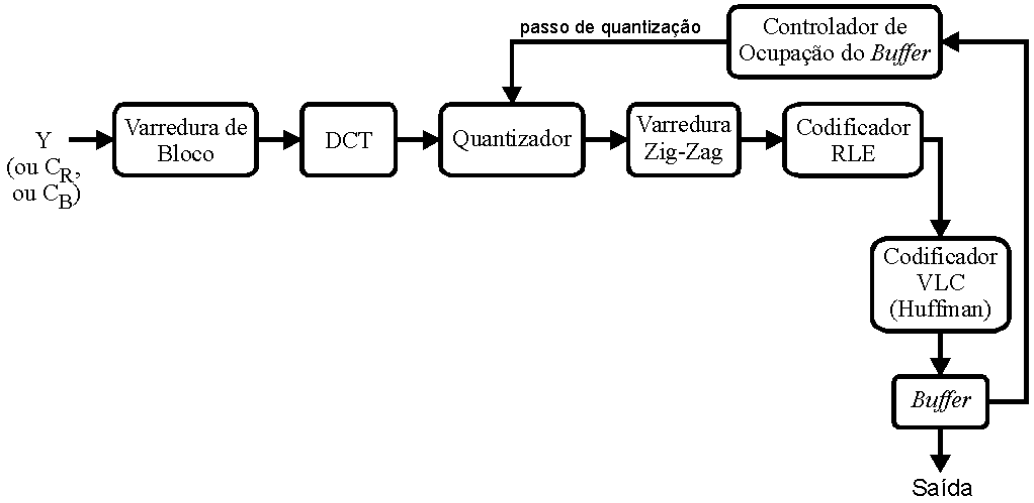

Figura 5.5: Codificador por DCT básico. O Controle de Ocupação do *Buffer* é usualmente a heurística TM5 - *Test Model 5*, definida em http://www.mpeg.org .

O processo de quantização dos coeficientes da DCT é usado para prover controle direto da entrada do *buffer*. À medida que o *buffer* enche, o passo de quantização do quantizador é aumentado de forma a reduzir o número de bits usado para codificar cada coeficiente DCT. Por outro lado, à medida que o *buffer* esvazia, o passo de quantização é diminuído de forma a aumentar o número de bits atribuído a cada coeficiente. O codificador informa ao decodificador o valor do passo de quantização instantaneamente adotado.

É importante notar que quanto menor a taxa de transmissão (em bps) na saída de um codificador de vídeo MPEG mais vagarosamente o *buffer* irá esvaziar. Nesta situação o codificador tentará compensar este efeito para evitar *overflow* do *buffer* através do aumento do passo de quantização utilizado para quantizar os coeficientes da DCT. No entanto, quanto maior o passo de quantização maior será o ruído de quantização, e, portanto, a qualidade da imagem será tanto pior quanto mais baixa for a desejada taxa de transmissão para o canal. Uma menor taxa de transmissão para o canal é sempre desejável sob o ponto de vista do custo da largura de banda contratada, visto que quanto menor a taxa de transmissão menor a banda ocupada no canal.

5.3.5 Redução de Redundância Temporal: Predição *Interframe*

Para explorar o fato de que imagens freqüentemente mudam pouco de um *frame* para o próximo, o padrão MPEG utiliza predição temporal para estimar o próximo *frame* a ser codificado a partir de um *frame* de referência prévio.

A Figura 5.6 ilustra uma possível implementação de um codificador baseado em predição temporal, denominado codificador PCM Diferencial (DPCM – *Differential Pulse Code Modulation*). Em um codificador DPCM apenas as diferenças entre a imagem de entrada e uma predição baseada na saída prévia localmente decodificada são quantizadas e transmitidas. Note que a predição não pode ser baseada em imagens fonte prévias, porque a predição tem que poder ser repetida no decodificador, onde as imagens fonte não estão disponíveis. Conseqüentemente, o codificador no transmissor contém um decodificador local, constituído pelo preditor que recebe em sua entrada a sua própria saída adicionada do erro de predição quantizado. O decodificador local reconstrói as imagens exatamente como elas seriam no decodificador no receptor. Esta forma de predição, em que amostras de um *frame* de referência são usadas na predição de amostras de outro *frame*, denomina-se de predição *Interframe*. O objetivo da predição *Interframe* é reduzir a redundância temporal intrínseca à seqüência de imagens representada pela seqüência de *frames*.

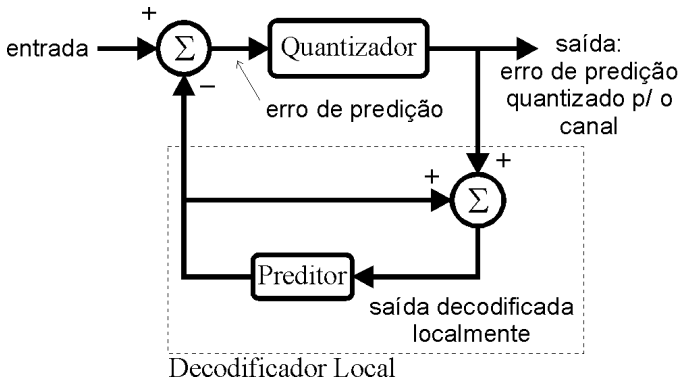

Figura 5.6: Codificador DPCM básico.

Na codificação MPEG, a predição *Interframe* é combinada com a DCT e codificação RLE/VLC já discutidas, como mostra a Figura 5.7. O codificador subtrai a predição da entrada estimada pelo *frame* anterior para formar uma "imagem-erro-de-predicão". A DCT é aplicada ao erro de predição, os coeficientes resultantes são quantizados, e estes valores quantizados são codificados usando RLE/VLC.

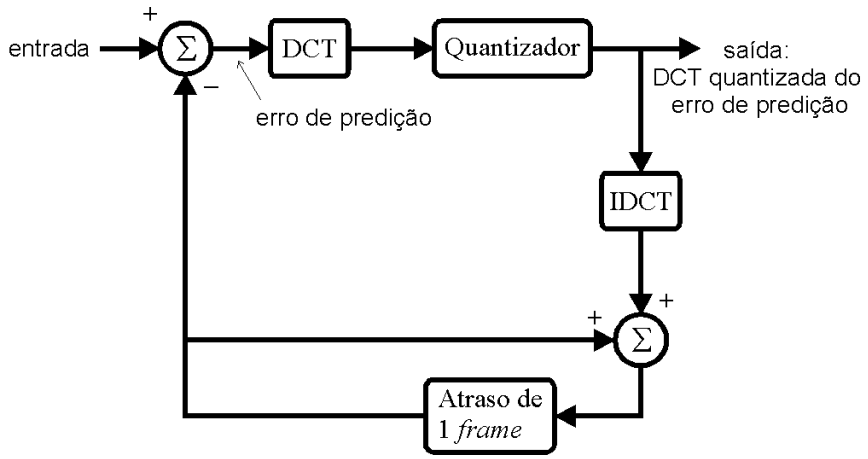

Figura 5.7: Codificador por DCT com predição *interframe*.

A predição *Interframe* mais simples consiste em estimar blocos (em geral retangulares) de amostras no *frame* a ser predito a partir de blocos respectivamente localizados na mesma posição no *frame* de referência. Neste caso, a estimativa consiste no bloco localizado na mesma posição no *frame* de referência, ou seja, exatamente um *frame* atrás, conforme mostrado na Figura 5.7. Esta predição é precisa para regiões da imagem com comportamento estacionário, mas pobre em regiões que apresentam movimento. Quando os blocos que compõe a imagem não podem ser caracterizados como estacionários, o codificador utiliza Predição *Interframe* com Compensação de Movimento.

5.3.6 Predição *Interframe* com Compensação de Movimento

Quando a seqüência de cenas transmitidas é tal que ocorre movimento entre um determinado bloco do *frame* a ser predito e o respectivo bloco no *frame* de referência, o método mais adequado para eliminar redundância temporal é a Predição *Interframe* com Compensação de Movimento. Este tipo de predição procura compensar qualquer movimento de translação que tenha ocorrido entre o bloco que está sendo codificado e o respectivo bloco no *frame* de referência que é usado como ponto de partida para predição (ver Figura 5.8).

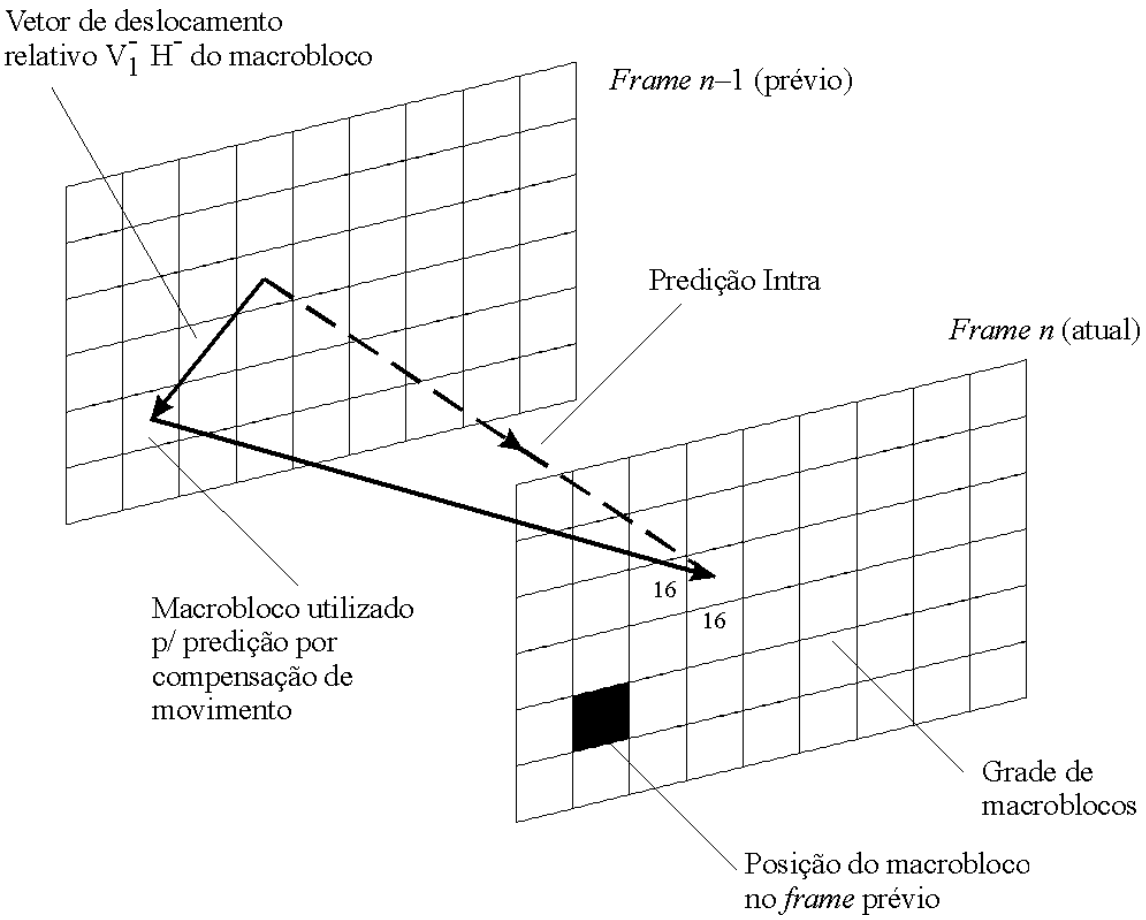

Figura 5.8: Predição *interframe* com compensação de movimento.

Um método para determinar o movimento que ocorreu entre o bloco que está sendo codificado e o bloco no *frame* de referência, é a busca *block-matching*, na qual um grande número de tentativas de deslocamento são testados no codificador (ver Figura 5.9). O melhor deslocamento é selecionado com base na medida do mínimo erro entre o bloco que está sendo codificado e a estimativa de predição. Já que o padrão MPEG define apenas o processo de decodificação, e não o processo de codificação, a escolha do algoritmo de medida do movimento é deixado a critério do projetista do decodificador. Portanto, esta é uma área em que ocorrem diferenças consideráveis de desempenho entre diferentes algoritmos e diferentes implementações. Um requerimento principal é ter uma área de busca grande o suficiente para cobrir qualquer movimento que ocorra de *frame* para *frame*. Entretanto, o aumento do tamanho da área de busca aumenta o custo computacional necessário para encontrar o *best match* (*best match*: situação em que ocorre maior semelhança entre o bloco referência e o bloco sendo codificado). Várias técnicas, tais como a *hierarchical block matching* (em que o *best match* é encontrado com base em sucessivos níveis de resolução) podem ser utilizadas para superar este dilema.

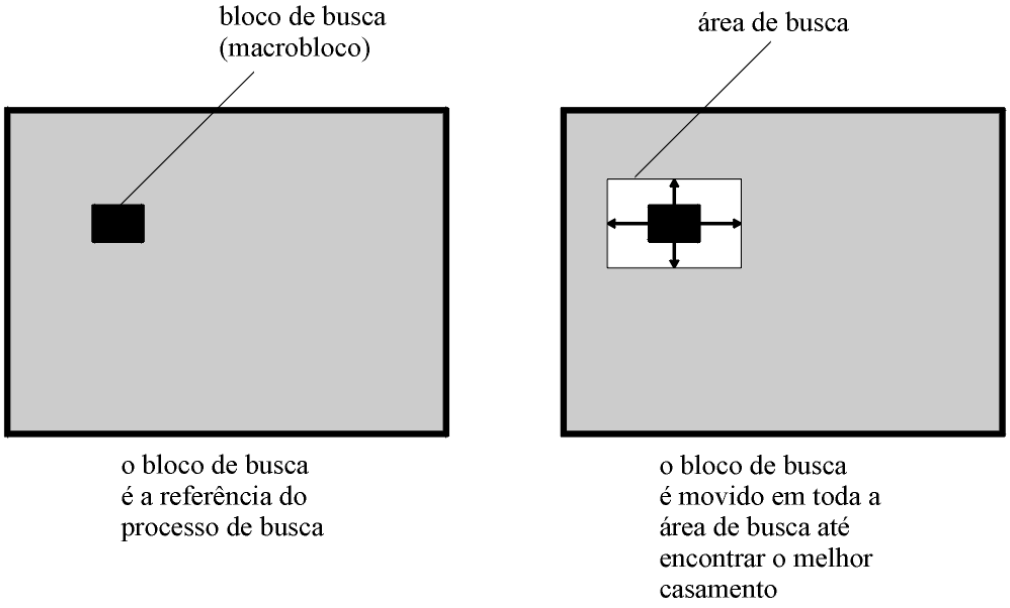

Figura 5.9: Processo de busca *block matching*.

Uma outra forma de predição utilizada no padrão MPEG é a denominada predição bidirecional , em que a predição é feita nas direções *backward* (a partir de um *frame* subsequente) e *forward* (a partir de um *frame* prévio). A predição bidirecional estabelece uma combinação linear destes dois *frames*, interpolando os dois deslocamentos (ver Figura 5.10).

A predição bidirecional é particularmente útil quando o movimento da cena torna visível áreas de detalhe, situação em que a predição em uma única direção não é muito eficiente. Note que, para permitir predição *backward* a partir de um *frame* futuro, o codificador reordena as imagens de tal forma que sejam transmitidas em uma ordem diferente daquela em que são mostradas. Este processo e a reordenação para corrigir a ordem de apresentação no decodificador introduzem um atraso considerável no processamento (*end-to-end delay*), atraso este que pode ser um problema em algumas aplicações. Para superar este problema, o padrão MPEG define um perfil operacional que não usa a predição bidirecional, conforme será visto adiante.

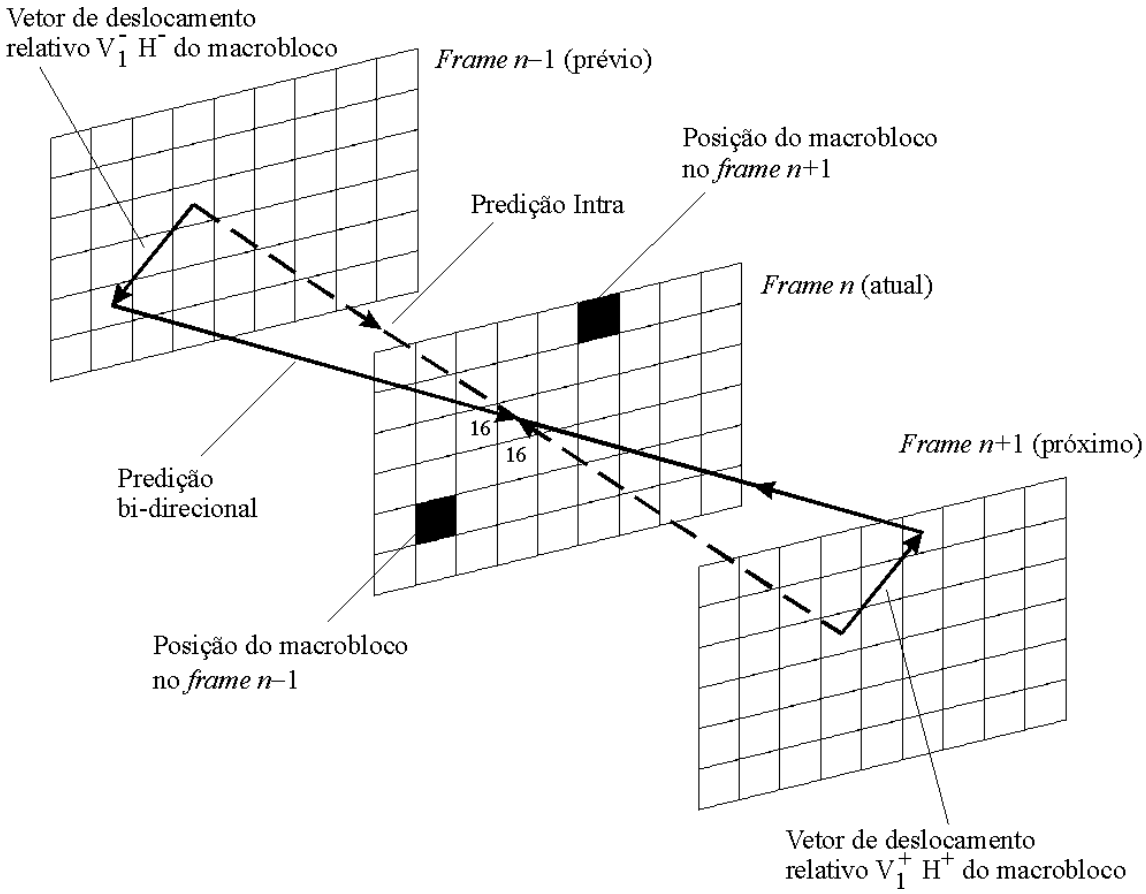

Figura 5.10: Predição bidirecional com compensação de movimento.

Enquanto que a unidade básica de codificação para redundância espacial no padrão MPEG é baseada em blocos de 8×8 pixels, a compensação de movimento é usualmente baseada em **macroblocos** de 16 pixels por 16 linhas. O tamanho do macrobloco é um compromisso entre

1- a necessidade de minimizar a taxa de bits necessária para transmitir a representação do movimento ocorrido entre os blocos (conhecida como **vetores de movimento**), a qual aponta para a utilização de um tamanho de macrobloco maior e

2- a necessidade de variar o processo de predição localmente dentro do conteúdo da imagem (e do movimento associado), o que sugere a necessidade de um tamanho menor de macrobloco.

Para minimizar a taxa de bits necessária para transmitir os vetores de movimento, os vetores de movimento são codificados diferencialmente com referência aos vetores de movimento prévios. O erro de predição do vetor de movimento é codificado através de codificação por comprimento variável, utilizando uma outra tabela VLC.

A Figura 5.11 mostra um codificador por DCT simplificado com compensação de movimento *interframe* no qual, para fins didáticos, a implementação do processo de predição por compensação de movimento é ilustrado pelo uso de um atraso variável. O atraso variável representa as tentativas de deslocamento na busca do *best match*, quando,

então, o erro de predição é mínimo. Uma vez ocorrendo o *best match*, o atraso undidimesional é convertido em vetores de movimento bidimensionais.

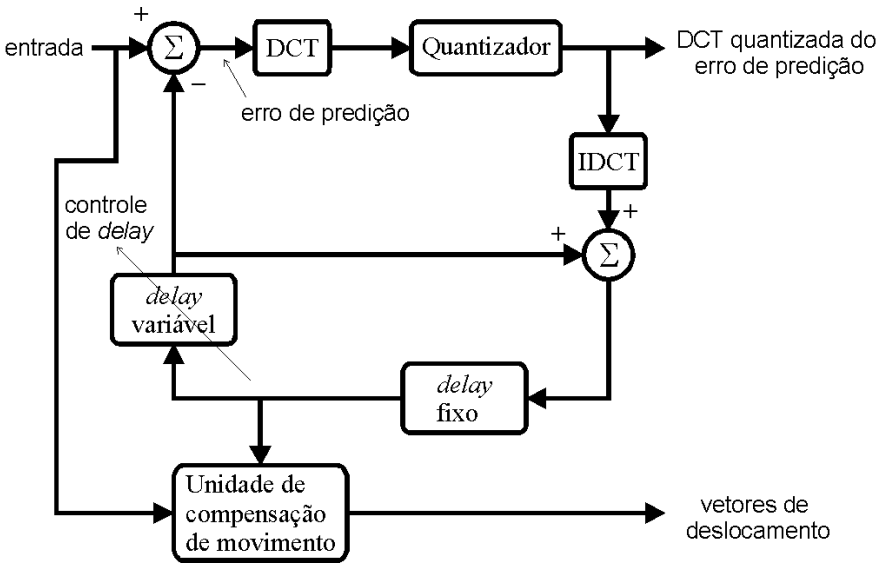

Figura 5.11: Codificador por DCT com predição *interframe* e compensação de movimento.

5.3.7 Modos de Predição

Em um codificador MPEG-2, o preditor por compensação de movimento permite a utilização de muitos métodos para obter a predição. Por exemplo, um macrobloco pode ser predito por predição *forward*, a partir de uma imagem passada, por predição *backward* a partir de uma imagem futura (codificação a partir de *frames* do tipo P, conforme veremos), ou interpolada através da média entre uma predição *forward* e uma predição *backward* (codificação de *frames* do tipo B, a ser visto adiante). Outra opção é fazer uma predição de valor zero, de tal forma que o próprio bloco na imagem fonte seja codificado pela DCT, ao invés de codificar com base no erro de predição. A codificação de tais macroblocos é conhecida como Intra ou do tipo I. Macroblocos do tipo Intra podem transportar informação de vetores de movimento, embora nenhuma informação de predição seja necessária. A informação de vetores de movimento para um macrobloco do tipo I não é usada em circunstâncias normais, mas a sua função é prover um meio de contornar erros de decodificação quando erros (devido à ruído/interferência no canal de transmissão) na seqüência de bits recebidos no receptor tornam impossível decodificar os dados para aquele macrobloco.

Para cada macrobloco a ser codificado, o codificador escolhe entre os possíveis modos de predição, tentando minimizar as distorções na imagem decodificada, dentro das restrições da disponibilidade da taxa de bits do canal. A informação sobre o modo de predição escolhido é transmitida ao decodificador, juntamente com o erro de predição, de tal forma que o decodificador possa regenerar a predição correta.

A Figura 5.12 ilustra como um macrobloco codificado bidirecionalmente no transmissor (um macrobloco do tipo B) é decodificado no receptor. As chaves ilustram os vários modos de predição disponíveis para decodificar um macrobloco.

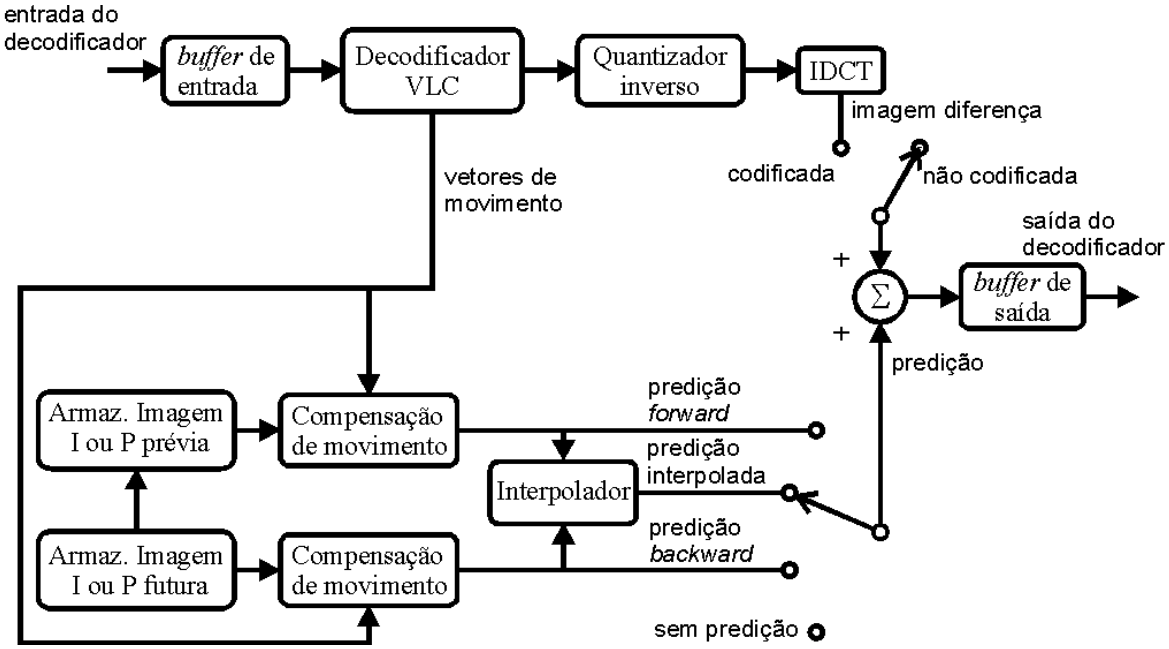

Figura 5.12: Decodificação de um macrobloco "B".

Note que o codificador no transmissor tem a opção de não codificar alguns macroblocos. Neste caso nenhuma informação sobre os coeficientes da DCT é transmitida para o receptor e o contador de endereço dos macroblocos simplesmente avança para o endereço do próximo macrobloco. A saída do decodificador para os macroblocos não-codificados consiste simplesmente da saída do preditor.

5.3.8 Tipos de Imagens

Nos referiremos aqui a imagens como sendo indistintamente *frames* ou *fields* (*field* é uma das duas subdivisões de *frame* quando a imagem é entrelaçada). No padrão MPEG-2, três tipos de imagens são definidos (ver Figura 5.13). O tipo de imagem define qual modo de predição pode ser usado para codificar cada macrobloco.

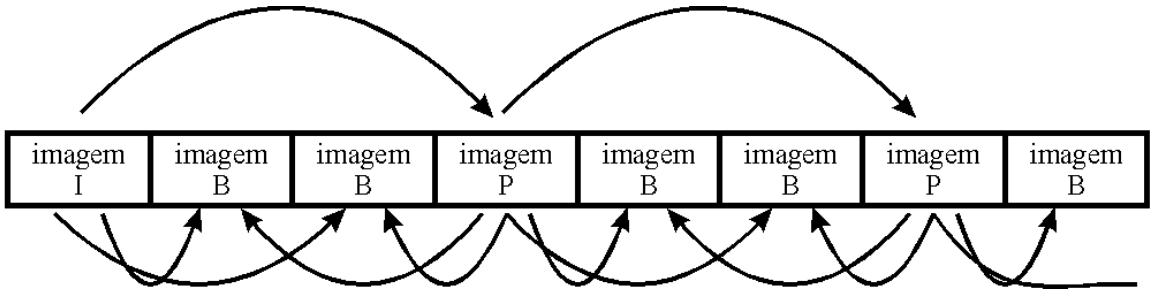

Figura 5.13: Tipos de imagem no sistema MPEG. A ponta de cada seta indica a imagem resultante do processo de predição e a origem da seta indica a imagem de referência usada no processo de predição.

5.3.8.1 Imagem Intra (Imagem do Tipo I): São codificadas sem referência a outras imagens. Uma compressão moderada é obtida através da redução da redundância espacial, mas não ocorre compressão por redução da redundância temporal. São importantes no

sentido de que possibilitam pontos de acesso na seqüência de bits onde a decodificação pode começar, sem referência a imagens prévias. São úteis quando, no receptor, deseja-se iniciar a decodificação em um ponto intermediário na seqüência total de bits.

5.3.8.2 Imagem Preditiva (Imagem do Tipo P): São codificadas utilizando predição por compensação de movimento a partir de uma imagem I ou P passada e podem ser usadas como referência para predição *forward*. Através da redução das redundâncias espacial e temporal, as imagens do tipo P oferecem uma maior compressão, comparativamente às imagens do tipo I.

5.3.8.3 Imagem do Bidirecionalmente Preditiva (Imagem do Tipo B): Usam tanto imagens I ou P passadas quanto futuras para compensação de movimento, e oferecem o mais alto grau de compressão. Como notado acima, para possibilitar predição *backward* a partir de um *frame* futuro, o codificador reordena as imagens a partir da ordem natural em que são mostradas para a ordem de transmissão (ou seqüência de bits transmitidos) de tal forma que a imagem B é transmitida após as imagens passada e futura, às quais a imagem B se referencia (ver Figura 5.14). Esta operação introduz um atraso que depende do número de imagens B consecutivas.

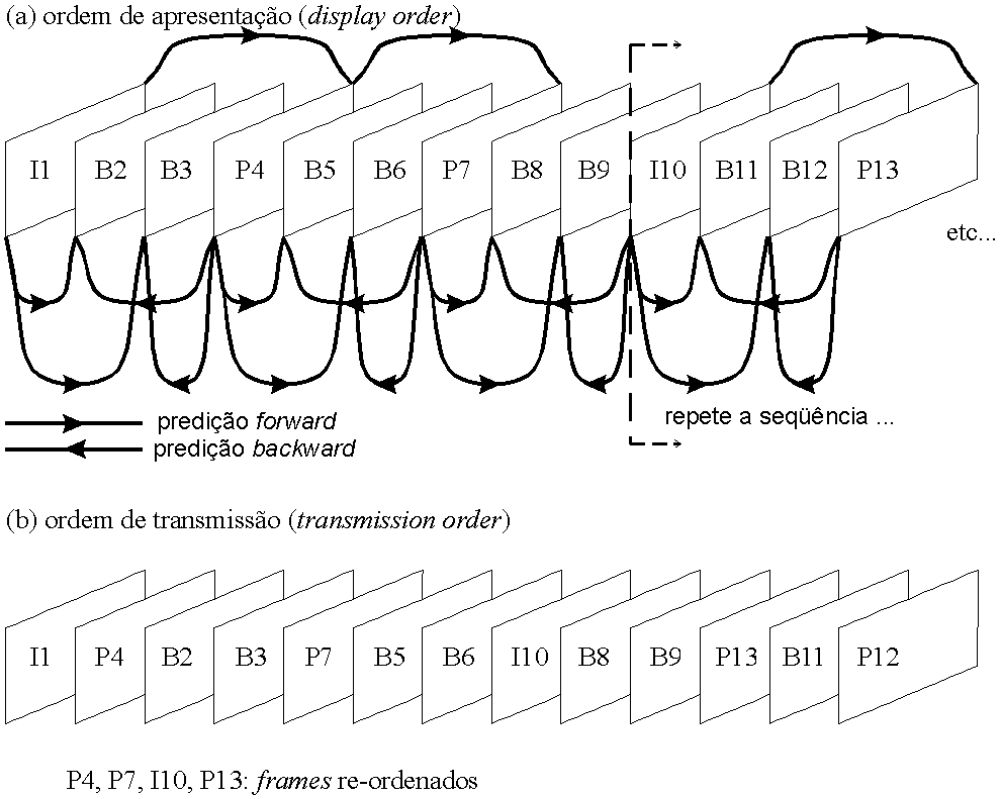

Figura 5.14: Exemplo de GOP com **N** = 9 e **M** = 3.

5.3.8.4 Grupo de Imagens (GOP - *Group of Pictures*): Os diferentes tipos de imagens ocorrem em uma seqüência que se repete chamada Grupo de Imagens ou GOP. Um típico GOP é ilustrado em sua ordem de apresentação (*display order*) na Figura 5.14(a). O mesmo GOP é mostrado na ordem de transmissão (*transmission order*) na Figura 5.14(b).

A estrutura de um GOP pode ser descrita por dois parâmetros: o **N** número de imagens no GOP e o número **M** de imagens entre duas imagens P incluindo uma delas. A estrutura do GOP ilustrado na Figura 5.14 é descrita como **N** = 9 e **M** = 3.

Para uma dada qualidade de imagem decodificada, a codificação usando cada tipo de imagem produz um número diferente de bits. Em uma seqüência típica, uma imagem codificada do tipo I gera três vezes mais bits do que uma imagem codificada do tipo P, a qual por sua vez gera 50 % mais bits do que uma imagem codificada do tipo B.

5.4 Perfis Operacionais e Níveis do Padrão MPEG

O Padrão MPEG-2 pretende ser genérico, permitindo uma gama variada de aplicações. Diferentes algoritmos de compressão, desenvolvidos para variadas situações, foram integrados em uma única sintaxe da seqüência de bits codificada.

A implementação de toda a sintaxe em todos os decodificadores é uma operação desnecessariamente complexa. Desta forma, um pequeno número de subconjuntos ou **perfis operacionais (*profiles*)** da sintaxe completa têm sido definidos. Ainda, dentro de um dado perfil, um **nível** é definido, o qual especifica um conjunto de restrições aos parâmetros do perfil, tal como a resolução da imagem. Em geral, os perfis definidos se ajustam de tal forma que um perfil mais alto é um superconjunto de um inferior. Um decodificador que suporta um determinado perfil e nível particular deve suportar o correspondente subconjunto da sintaxe completa e um conjunto de restrições de parâmetros. Para restringir o número de opções que deve ser suportado, apenas combinações selecionadas de perfis e níveis são definidas como **pontos de conformidade** (ver tabela 1).

	Denominação do Perfil e taxa de transmissão (Mbit/s)					
	Resolução (H×V pixels) da Imagem / Taxa [Hz]	**Simples**	**Principal**	**Ajustável em Qualidade de Vídeo (SNR)**	**Ajustável em Resolução Espacial**	**Alto**
N í v e l	High 1920×1152/60	-	MP @HL (80Mbit/s)	-	-	HP @HL (100Mbit/s)
	High-1140 1440×1152/60	-	MP @H-14 (60Mbit/s)	-	SPT@H-14 (60Mbit/s)	HP @H-14 (80Mbit/s)
	Main 720×576/30	SP@ML 15Mbit/s	MP @ML (15Mbit/s)	SNR @ML (15Mbit/s)	-	HP @ML (20Mbit/s)
	Low 352×280/30	-	MP @LL (4Mbit/s)	SNR @LL (4Mbit/s)	-	-
	ISO 11172 (MPEG-1) 1.856 Mbit/s	-	-	-	-	-
Notas:	Todos os decodificadores devem poder decodificar seqüências de bits ISO/IEC 11172.					

Tabela 1: Perfis e Níveis MPEG

Segue uma lista dos principais perfis:

- **Perfil Simples:** não utiliza *frames* B, e, portanto, não utiliza predição *backward* ou interpolada. Conseqüentemente, nenhuma reordenação de imagens é requerida, o que faz com que este perfil seja apropriado para aplicações que toleram pequeno atraso, tais como vídeo conferência.

- **Perfil Principal:** inclui suporte para imagens B, o que melhora a qualidade de imagem para uma dada taxa de transmissão de bits, mas aumenta o atraso. Presentemente, a maior parte dos codificadores de vídeo MPEG-2 implementados em *chip* suportam este perfil.

- **Perfil SNR:** inclui suporte para ajuste diferenciado da SNR (*Signal-to-Noise Ratio*) de quantização (a qual define a qualidade de vídeo) nas diversas *layers* do padrão MPEG. Cada *layer* transporta um tipo de informação diferente. Assim vídeo de alta qualidade é transportado em um tipo de *layer* enquanto que vídeo de baixa qualidade é transportado em outra.

- **Perfil Espacial:** inclui suporte para ajuste da resolução espacial diferenciado nas diversas *layers*.

- **Perfil Alto:** inclui suporte para vídeo amostrado 4:2:2.

Todos os decodificadores MPEG-2 irão também decodificar imagens codificadas utilizando o padrão MPEG-1, mas não vice-versa.

5.5 Conclusões

O grupo MPEG (http://www.mpeg.org) têm obtido sucesso em definir padrões para codificação de compressão de vídeo, servindo a uma ampla gama de aplicações, taxas de bits, qualidades e serviços. Os padrões são baseados em um conjunto flexível de técnicas de redução de taxa de bits. A especificação somente define a sintaxe das cadeias de bits e o processo de decodificação; o processo de codificação não é especificado e o desempenho de um codificador irá variar dependendo, por exemplo, da qualidade da medida vetor-de-movimento, e os processos usados para seleção do modo de predição.

A qualidade da imagem de um codificador/decodificador MPEG depende fortemente do conteúdo da imagem; mas à medida que aumenta a experiência com codificação MPEG, as taxas de bits necessárias para uma dada qualidade de imagem provavelmente devem diminuir.

5.6 Referências

[1] ISSO/IEC 13818-2:1995, Information Technology - Generic coding of moving pictures and associated audio information:Video.

[2] SARGINSON, P. A., 1996, MPEG-2: Overview of the systems layer. BBC Research & Development Department Report Nº. 1996/2.

[3] STOLL, G. and GILCHRIST, N. H. C. , 1996. ISSO/IEC MPEG-2 AUDIO: Bit-rate-reduced coding for two-channel and multichannel sound. BBC Research & Development Department Report Nº. 1996/4.

[4] TUDOR, P., 1995. MPEG-2: What it is and what it isn't. MPEG-2 Video Compression Tutorial, IEE Colloquium, January.

[5] WELLS, N., 1994. Bit-rate reduction of digital television for transmission: na introdutory review. BBC Research & Development Department Report N^{o}. 1994/2.

[6] WELLS, N., BARBERO, M. and HOFMANN, H., 1992. DCT source coding and current implementations for HDTV, EBU Technical Review N^{o} 251, Spring.

[7] SANDBANK, C. P. (editor), 1990. Digital television, John Wiley, Chichester, ISBN 0 471 92360.

[8] TUDOR, P. N., and WELLS, N. D., 1994. Digital video compression: standardisationn of scalable coding schemes. BBC Research & Development Department Report N^{o}. 1994/10.

[9] CCIR Recommendation 601-1 . Encoding parameters of digital television for studios.