

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Engenharia do campus de Guaratinguetá

**O Padrão de Compressão de Vídeo MPEG - Moving Picture
Experts Group**

GUSTAVO DOS SANTOS ARABURA

Guaratinguetá

2012

GUSTAVO DOS SANTOS ARABURA

O PADRÃO DE COMPRESSÃO DE VÍDEO MPEG – MOVING PICTURE EXPERTS
GROUP

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia elétrica da Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia Elétrica.

Orientador: Prof. Dr. José Feliciano Adami

Guaratinguetá

2012

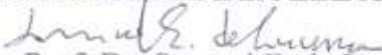
A658p	<p>Arabura, Gustavo dos Santos O padrão de compressão de vídeo MPEG – moving picture experts group / Gustavo dos Santos Arabura. – Guaratinguetá : [s.n], 2012 62 f. : il. Bibliografia : f. 62</p> <p>Trabalho de Graduação em Engenharia Elétrica – Universidade Estadual Paulista, Faculdade de Engenharia de Guaratinguetá, 2012 Orientador: Prof. Dr. José Feliciano Adami</p> <p>1. Compressão de dados (Telecomunicações) 2. Codificação I. Título.</p> <p style="text-align: right;">CDU 621.39</p>
-------	---

O PADRÃO DE COMPRESSÃO DE VÍDEO MPEG – MOVING PICTURE
EXPERTS GROUP

GUSTAVO DOS SANTOS ARABURA


ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO
COMO PARTE DO REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE
GRADUADO EM ENGENHARIA ELÉTRICA

APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO
DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA


Prof. Dr. Samuel E. de Lucena
Coordenador

BANCA EXAMINADORA:


Prof. Dr. JOSÉ FELICIANO ADAMI
Orientador/UNESP-FEG


Prof. Dr. TEÓFILO MIGUEL DE SOUZA
UNESP-FEG


Prof. Dr. DANIEL JULIEN BARROS DA SILVA SAMPAIO
UNESP/FEG

Fevereiro de 2012

DADOS CURRICULARES**GUSTAVO DOS SANTOS ARABURA**

NASCIMENTO	27.07.1987 – SÃO PAULO / SP
FILIAÇÃO	Mauricio Arabura Maria Izabel dos Santos Arabura
2007 / 2011	Curso de Graduação em Engenharia Elétrica Universidade Estadual Paulista – Faculdade de Engenharia de Guaratinguetá

De um modo especial, aos meus pais, Mauricio e Izabel, a minha irmã, Raquel, a minha namorada, Fabiana, e ao meu filho Luis Gustavo, por fazerem parte da minha vida de uma maneira muito especial.

AGRADECIMENTOS

Primeiro de tudo, gostaria de agradecer a Deus por tudo o que ele tem me preparado. Por ter me dado a saúde e a força necessária durante toda a graduação que se encerra com esse trabalho.

De um modo especial, aos meus pais, Mauricio e Izabel, por terem me dado todo o amor e carinho que precisei durante toda a minha vida, principalmente no esforço para me ver um engenheiro.

A minha irmã, Raquel, por ser uma peça fundamental em todo o meu aprendizado como pessoa.

A minha namorada, Fabiana, pela paciência e por ter me apoiado em todas as horas que precisei. E, especialmente, por ter dado o maior presente da minha vida, meu filho Luis Gustavo.

A minha família, especialmente meus primos, e a todos os meus amigos, por sempre estarem do meu lado quando precisei, tanto nos momentos bons quanto nos ruins.

Aos meus “irmãos” da República DV, por todos os momentos vividos, pelas noites de estudo juntos e pelas experiências trocadas. Cada um tem a sua importância ao longo destes anos.

Ao meu orientador, Professor Dr. José Feliciano Adami, por acreditar em meu potencial e me ajudar na elaboração deste trabalho.

"Um fator importante para o sucesso é a autoconfiança. Um fator importante para a autoconfiança é a preparação."

(Arthur Ashe)

ARABURA, G.S. **O Padrão de Compressão de Vídeo MPEG – Moving Picture Experts Group**. 2011. X pg. Monografia (Graduação em Engenharia Elétrica) – Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, Guaratinguetá, 2011.

RESUMO

Este trabalho tem como foco o estudo do padrão de compressão de vídeo MPEG. Para tanto, foi elaborado um estudo partindo dos conceitos básicos de um vídeo digital, abordando os componentes necessários para o entendimento das ferramentas de codificação de vídeo utilizadas pelo padrão MPEG. O *Motion Picture Experts Group* (MPEG) foi formado no final dos anos 80, por um grupo de especialistas, com o intuito de criar padrões internacionais de codificação e decodificação de áudio e vídeo. Este trabalho abordará as técnicas de compressão de vídeo presentes no padrão MPEG, bem como as suas evoluções. São descritos os padrões MPEG-1, MPEG-2, MPEG-4 e o H.264 (MPEG-4 Parte 10), porém, os dois últimos são apresentados com maior ênfase, pois são os padrões presentes nas tecnologias de vídeo mais modernas, como nas transmissões em HDTV (*High Definition television*).

PALAVRAS-CHAVE: MPEG, compressão, codificação, digital, H.264

ARABURA, G.S. **The Standard MPEG video compression – Moving Picture Experts Group**. 2011. X pg. Monograph (Graduation in Electrical Engineering) – Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, Guaratinguetá, 2011.

ABSTRACT

This work focuses on the study of video compression standard MPEG. To this end, a study was undertaken starting from the basics of digital video, addressing the components necessary for the understanding of the tools used by the video coding standard MPEG. The Motion Picture Experts Group (MPEG) was formed in the late '80s by a group of experts in order to create international standards for encoding and decoding audio and video. This paper will discuss the techniques present in the video compression standard MPEG, as well as its evolution. Will be described in the MPEG-1, MPEG-2, MPEG-4 and H.264 (MPEG-4 Part 10), however, the last two will be presented with more emphasis, because the standards are present in most modern video technologies, as in HDTV broadcasts.

KEYWORDS: MPEG, compression, codification, digital, H.264

LISTA DE FIGURAS

FIGURA 1 - SEQUÊNCIA DE IMAGENS (INATEL, 2010).....	16
FIGURA 2 – ESPAÇO DE COR RGB (TEXAMPLE, 2011).....	18
FIGURA 3 – ESPAÇO DE COR YCrCb (WIKIPEDIA, 2011).	19
FIGURA 4 - RELAÇÃO DE LUMINÂNCIA E CROMINÂNCIA (MANZATO, 2004).....	20
FIGURA 5 - REPRESENTAÇÃO DO PIXEL (SAMSUNG, 2010).....	21
FIGURA 6 - REPRESENTAÇÃO DE MACROBLOCO E BLOCO (INATEL, 2010).....	22
FIGURA 7 - REPRESENTAÇÃO DE FATIA (MANZATO, 2004).	22
FIGURA 8 - AMOSTRAGEM DE UM SINAL ANALÓGICO (ARANGO, 2011).	24
FIGURA 9 – VETORES DE DESLOCAMENTO (INATEL, 2010).	30
FIGURA 10 - VETOR DE MOVIMENTO OBTIDO PELO <i>BLOCK MATCHING</i> (INATEL, 2010).	31
FIGURA 11 - PREDIÇÃO BIDIRECIONAL (C. CASTRO; F. CASTRO, 2002).	32
FIGURA 12 - ESTRUTURA DO GOP (MANZATO, 2004).....	33
FIGURA 13 - MATRIZ RESULTANTE DO PROCESSO DCT (INATEL, 2010).	34
FIGURA 14 - MATRIZ RESULTANTE DO PROCESSO DE QUANTIZAÇÃO (INATEL, 2010).	36
FIGURA 15 - PERCEPÇÃO VISUAL HUMANA ÀS FREQUÊNCIAS ESPACIAIS (INATEL, 2010).	36
FIGURA 16 - VARREDURA ZIGUEZAGUE (FAPTECH, 2004).....	37
FIGURA 17 - CODIFICAÇÃO EM BLOCOS MPEG-2 (MACKENZIE DE ENGENHARIA E COMPUTAÇÃO, ANO 5, N. 5).....	41
FIGURA 18 - TELEVISÃO COM INTERATIVIDADE (CANAL BLOOMBERG, 2007).....	46
FIGURA 19 – EXEMPLO DE COMPOSIÇÃO DE CENA (SOLER, 2006).	47
FIGURA 20 – ESTRUTURA HIERÁRQUICA DE VÍDEO MPEG-4 (SOLER, 2006).....	48
FIGURA 21- MODOS DE CODIFICAÇÃO DE UM VOP (VALENTIN, 2004).	50
FIGURA 22 - CODIFICAÇÃO MPEG-4 USANDO GMC (UFF, 2005).	51
FIGURA 23 – VETORES DE MOVIMENTO UTILIZANDO MÚLTIPLOS QUADROS (IMG.LX, 2007)...	55
FIGURA 24 - COMPARAÇÃO ENTRE H.264, MPEG-4 E MOTION JPEG (AXIS COMMUNICATION , 2008).....	58

SUMÁRIO

1.	INTRODUÇÃO.....	15
1.1	Fundamentos de Vídeo.....	16
1.2	Vídeo Digital.....	17
1.3	Espaço de Cores.....	18
1.3.1	RGB.....	18
1.3.2	Y, Cr, Cb.....	19
1.4	Luminância e Crominância.....	20
1.5	Pixel.....	21
1.6	Bloco e Macrobloco.....	22
1.7	Fatia.....	22
1.8	Quadro (Frame).....	23
1.9	Amostragem.....	23
2.	INTRODUÇÃO A COMPRESSÃO DE VÍDEO.....	25
2.1	Perfis e Níveis.....	25
2.2	Requisitos do MPEG para vídeo.....	26
2.2.1	Acesso Randômico.....	26
2.2.2	Buscas à frente/atrás rápidas.....	26
2.2.3	Reprodução para trás.....	26
2.2.4	Sincronização áudio-vídeo.....	27
2.2.5	Resistência a erros.....	27
2.2.6	Atraso de codificação/decodificação.....	27
2.2.7	Editibilidade.....	27
2.2.8	Flexibilidade do formato.....	28
2.3	Problemas da Compressão.....	28
2.4	Técnicas de Compressão de Vídeo MPEG.....	29
2.4.1	Redundância Temporal.....	29
2.4.1.1	Estimativa e Compressão do Movimento.....	29
2.4.1.1.1	<i>Block Matching</i>.....	31
2.4.1.1.2	Predição Bidirecional.....	31
2.4.1.2	GOP (<i>Group of Pictures</i>).....	32
2.4.2	Redundância Espacial.....	33

2.4.2.1	Transformada Direta do Cosseno (DCT)	33
2.4.2.2	Quantização.....	35
2.4.2.3	Varredura Ziguezague.....	37
2.4.2.4	Codificação VLC (<i>Variable Length Coding</i>)	38
3.	O PADRÃO MPEG.....	39
3.1	Padrão JPEG.....	39
3.2	Padrão MPEG-1.....	40
3.3	Padrão MPEG-2.....	40
3.3.1	Perfis e Níveis do MPEG-2.....	42
3.4	Padrão MPEG-4.....	43
3.4.1	Principais características.....	44
3.4.2	Aplicações.....	44
3.4.3	MPEG-4 Visual.....	46
3.4.3.1	Objeto de Vídeo.....	46
3.4.3.2	Ferramentas de Codificação.....	50
3.5	Padrão H.264.....	52
3.5.1	Principais características.....	52
3.5.2	Perfis e Níveis do H.264.....	53
3.5.3	Eficiência do H.264.....	54
3.5.4	Melhorias do H.264.....	55
4.	COMPARAÇÃO DE DESEMPENHO.....	57
4.1	Teste realizado com uma câmera de segurança.....	57
4.2	Estudo comparativo de economia de taxa de bits.....	59
5.	CONCLUSÃO.....	61
6.	BIBLIOGRAFIA.....	62

LISTA DE ABREVIATURAS E SIGLAS

AVC	<i>Advanced Video Coding</i>
CABAC	<i>Context-Adaptive Binary Arithmetic Coding</i>
CAVLC	<i>Context-Adaptative Variable-Lenght Codes</i>
CD	<i>Compact Disc</i>
DCT	<i>Transformada Discreta do Cosseno</i>
DVD	<i>Digital Video Disc</i>
GMC	<i>Global Motion Compensation</i>
GOP	<i>Group of Pictures</i>
GOV	<i>Group of Video Object Planes</i>
HD	<i>High Definition</i>
HDTV	<i>High Definition Television</i>
IDCT	<i>Transformada Discreta do Cosseno Inversa</i>
IP	<i>Internet Protocol</i>
ISO	<i>International Standards Organization</i>
ITU	<i>International Telecommunications Union</i>
ITU-T	<i>Telecommunication Standardization Sector</i>
JPEG	<i>Joint Photographic Experts Group</i>
MPEG	<i>Moving Picture Experts Group</i>
NTSC	<i>National Television System Committee</i>
PAL	<i>Phase Alternating Line</i>
RGB	<i>Red, Green, Blue</i>
RLE	<i>Run-length Encoding</i>
VCD	<i>Video Compact Disc</i>
VCEG	<i>Video Coding Experts Group</i>
VLC	<i>Variable Length Coding</i>
VO	<i>Video Object</i>
VOL	<i>Video Object Layer</i>
VOP	<i>Video Object Plane</i>
VRML	<i>Virtual Reality Modeling Language</i>
VS	<i>Video Session</i>
SNR	<i>Signal to Noise Ratio</i>

1 INTRODUÇÃO

À medida que ocorre a migração das mídias analógicas para as digitais, torna-se necessário contornar a dimensão dos arquivos de vídeo digitais, o que tornou imperativa a adoção de padrões de compressão para permitir a operacionalidade entre os equipamentos diferentes dos diversos fabricantes. As limitações dos meios de transmissão e a necessidade de armazenamento de mídias de forma compacta e rápida deram origem a um grupo de especialistas denominado MPEG (*Moving Picture Experts Group*). Formado pela ISO¹ (*International Standards Organization*), o grupo tinha como meta desenvolver um padrão de compressão de imagens dinâmicas.

A compressão de uma imagem possibilita transmissão eficiente e melhor aproveitamento das mídias de armazenamento digital. A principal meta dos algoritmos de compressão é produzir uma representação digital do sinal de vídeo, para que, depois de decodificado e reproduzido, ele tenha pequena diferença do sinal original, e que essas diferenças esteja dentro de determinado limiar, imperceptível ou indiferente ao receptor, buscando sempre a menor taxa de transmissão (*bit rate*) possível. Como o valor de cada pixel da imagem original é, normalmente, representado por um byte, o processo de compressão, efetivamente, ocorre quando representa-se cada pixel por uma quantidade menor de bits.

Como a tecnologia deve ser desenvolvida em função da programação e não o contrário, a necessidade de técnicas de compressão de vídeo mais eficientes é constante. Por esse motivo, que o padrão MPEG vem se aperfeiçoando desde a sua criação, oferecendo condições para as empresas apresentarem produtos novos para atender a exigência do mercado. O MPEG-4 e o H.264 são os padrões mais recentes, que oferecem desde uma imagem de alta qualidade, como o HDTV, até a interatividade do usuário com a televisão, por isso são descritos com maiores detalhes.

Portanto, por ser um padrão aberto às melhorias e atender as necessidades da programação, o padrão MPEG tem um enorme potencial para continuar sendo referência na representação eficiente de conteúdos audiovisuais.

¹ ISO – *International Standards Organization*: é uma entidade que aprova normas internacionais em todos os campos técnicos, de padronização/normalização, de 170 países (Wikipedia, 2011).

1.1 Fundamentos de Vídeo

Uma sequência de vídeo caracteriza-se por uma série de imagens (quadros) ordenadas no tempo. Quando a taxa de quadros ou campos é suficiente (acima de 20 quadros por segundo)², não percebemos a descontinuidade entre um quadro e outro.

O avanço da tecnologia possibilitou a reprodução da sucessão de imagens, oferecendo ao ser humano a oportunidade de ficar mais próximo do maravilhoso mundo dinâmico descrito por Lindstrom:

Vivemos em um mundo dinâmico – um mundo de movimento. Mesmo quando você se senta em uma sala onde nada se move, ainda pode alterar a perspectiva movendo sua cadeira, sua cabeça ou até mesmo apenas seus olhos. Você é, de certa forma, o cineasta (e o editor) do que vê. (LINDSTROM, 1996)

O efeito de sobreposição de imagens fixas torna possível a ilusão visual de movimento, esse fenômeno biológico é conhecido como persistência visual. Um objeto visto pelo olho humano permanece retido na retina por algum tempo. Portanto, quando a sucessão de imagens consecutivas apresenta uma fluidez com velocidade suficiente, o expectador tem a sensação de natureza dos movimentos, como observado na figura 1.

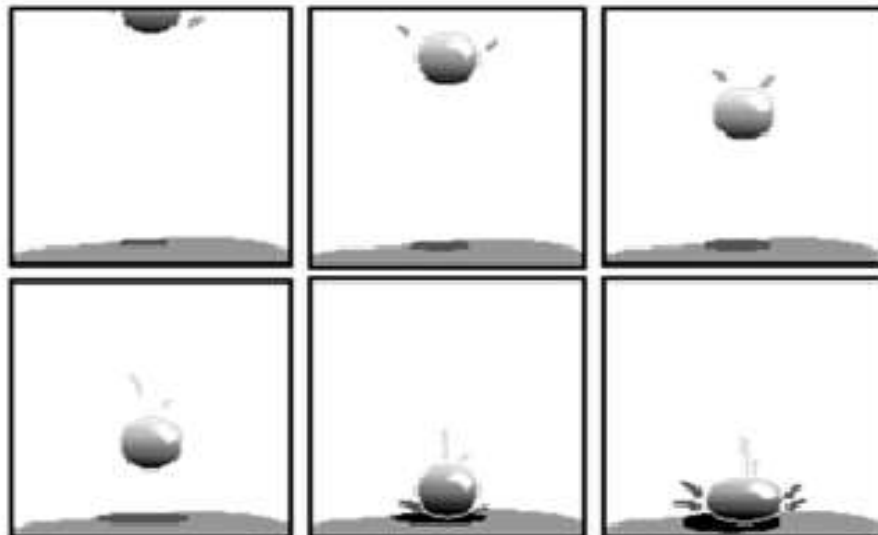


Figura 1 - Sequência de Imagens (INATEL, 2010).

² Em NTSC ou PAL-M existem 29,97 quadros (frames) por segundo. Em PAL europeu são 25.

1.2 Vídeo Digital

Um vídeo digital é caracterizado pela sucessão de imagens digitais ordenadas no tempo. A tecnologia digital é baseada na linguagem binária, formada por sequência de dados numéricos (0 e 1), conhecidos como bits. O conjunto desses sinais ou bits compõe a informação. Esse tipo de registro é mais preciso que o analógico, porém a capacidade de armazenamento da informação ocupa muito mais espaço que nos sistemas analógicos, por isso que a “compressão” das informações é um fator vital para o processo de vídeo digital.

A digitalização do sinal de vídeo traz benefícios em qualidade, flexibilidade e interface com os sistemas digitais. Ao amostrar as linhas de vídeo e convertê-las em centenas de números, a banda passante (faixa de frequência) necessária para o sinal aumenta muito.

As diferentes quantidades de dados podem ser utilizadas para transmitir uma mesma informação. Esta diferença consiste na redundância de dados. Uma imagem digital apresenta três tipos de redundâncias:

- Redundância de código: para qualquer sinal digitalizado, alguns valores codificados ocorrem com maior frequência que outros. Isto pode ser aproveitado pela codificação. Quanto maior a ocorrência de um determinado valor menor será a palavra para representar tal código. O código Morse utiliza o mesmo principio, as letras mais comuns no idioma inglês é o “E” e “T” e são representados por um ponto e um traço e os mais raros são “X” “Y” “Z” que são representados por quatro pontos e quatro traços.
- Redundância interpixel: através dos valores dos pixels vizinhos, pode-se prever o valor de um pixel. Esta correlação espacial está ligada ao relacionamento geométrico entre os objetos na imagem.
- Redundância psicovisual: está relacionada ao limite do detalhe que o olho pode perceber e a habilidade de seguir imagens com velocidade. Esta forma de redundância resulta da maneira como olho e o cérebro trabalham.

A relação entre a quantidade de bytes obtida no processo de conversão do sinal analógico para digital e a quantidade obtida após a compressão é chamada de taxa de compressão. Suponha-se que a conversão de um sinal de vídeo resulte em 200MB. Se após a compressão, obteve-se 10MB, é dito que houve uma compressão de 20:1 (vinte para um).

1.3 Espaço de cores

A cor é uma percepção visual provocada pela ação de um feixe de fótons sobre células especializadas da retina, que transmitem através de informação pré-processada no nervo óptico, impressões para o sistema nervoso (Wikipedia, 2011). Logo, o olho humano não consegue diferenciar componentes e sim a cor resultante.

Um espaço de cor nada mais é do que um modelo matemático usado para descrever cada cor a partir de fórmulas. Basicamente existem dois tipos de espaço de cores: o RGB, que é utilizado principalmente em sistemas computacionais; e o YCrCb, que é utilizado em sistemas televisivos.

1.3.1 RGB

Um espaço de cor é gerado por uma cor origem e três componentes primárias de cor. Dizemos então que uma imagem colorida usa um dado espaço de cor quando os três valores associados a cada pixel da imagem representam respectivamente as intensidades das três componentes primárias deste espaço. Uma imagem colorida pode ser representada em diferentes espaços de cores. O mais comum é o espaço RGB (*Red, Green, Blue*), como observado na figura 2.

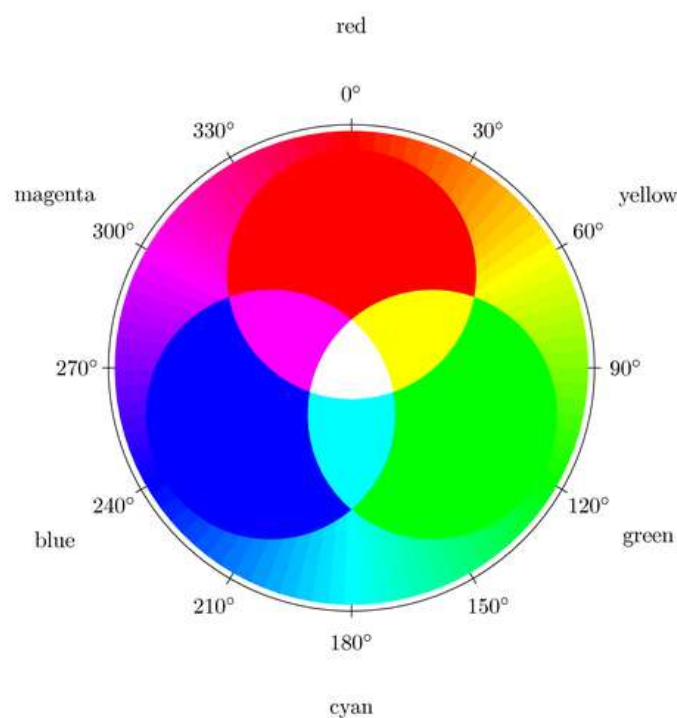


Figura 2 – Espaço de cor RGB (TEXAMPLE, 2011).

1.3.2 Y, Cr, Cb

No espaço de cor YCbCr, a luminância é representada pela componente Y, enquanto a crominância (matiz e saturação) é representada pelas componentes Cb (sinal de crominância diferença azul) e Cr (sinal de crominância diferença vermelho), como ilustra a figura 3. Esse espaço de cor possui a vantagem de ser mais perceptível, devido ao fato de que, assim como nossa visão, ele trabalha separadamente com brilho e a cor. Além disso, os programas que são transmitidos em preto e branco utilizam apenas a componente Y, dispensando as outras que definem as cores.

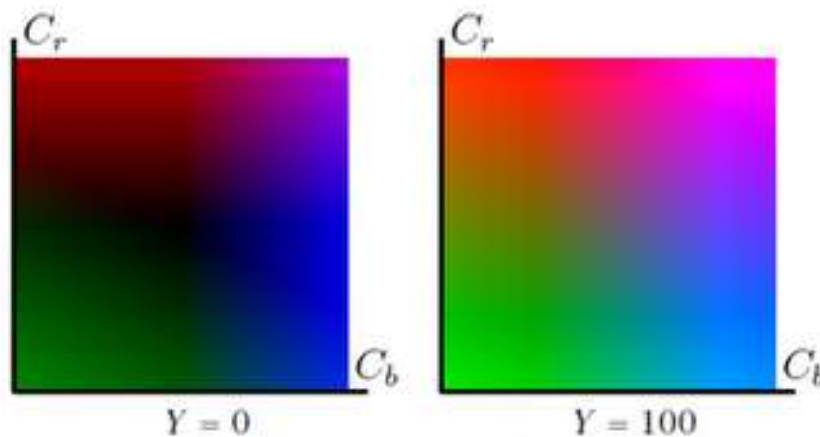


Figura 3 – Espaço de Cor YCrCb (WIKIPEDIA, 2011).

Como o olho humano trata de forma diferente o brilho e a cor da imagem, sendo mais sensível ao brilho do que à cor, a representação RGB torna-se insuficiente para a compressão, sendo necessária a conversão para o espaço de cores YCbCr, como descrita abaixo:

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cb = -0.1687R - 0.3313G + 0.5B$$

$$Cr = 0.5R - 0.4187G - 0.0813B$$

Na qual R representa a componente vermelha, G a verde e B a azul.

1.4 Luminância e Crominância

A forma de onda consiste primeiramente em um sinal de brilho (luminância) e um sinal de cor de alta-freqüência (crominância). A luminância e a crominância são somadas para formar a forma de onda completa. O sinal de luminância é uma série de tensões ou "níveis" que determinam variações de brilho na imagem. Cada uma das cores no sinal de barras tem um nível de luminância diferente, e as barras são organizadas em níveis, do mais alto para o mais baixo (branca, amarela, ciano, verde, magenta, vermelha, azul e preta).

O tamanho do sinal de crominância varia de uma cor para outra. A primeira barra é branca e não tem crominância. Todas as outras barras têm a quantia correta de amplitude de crominância para produzir uma intensidade completa de cor. A última barra, a preta, também não tem crominância (PIZZOTTI, 2009).

Como a visão humana é menos sensível às cores que ao preto e branco, os sinais de cor necessitam de uma banda passante menor. Essa limitação do ser humano permite a redução da banda passante, e é uma forma de compressão, já que não é necessária a transmissão de todas as informações de cores.

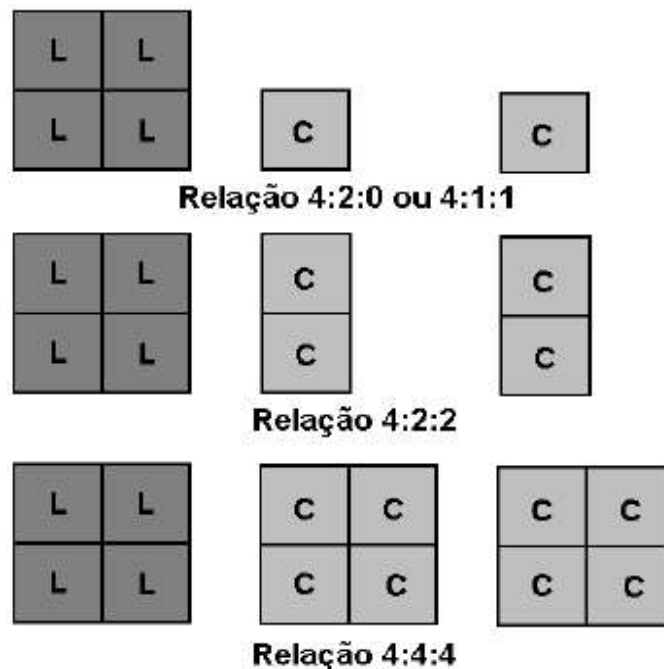


Figura 4 - Relação de Luminância e Crominância (MANZATO, 2004).

As componentes de luminância e crominância são representadas por três matrizes, que combinadas formam a imagem. Essas matrizes possuem tamanho variável, de acordo com as relações de componentes, como mostrado na figura 4. A matriz de luminância, por exemplo, possui o mesmo tamanho da imagem para qualquer relação. As matrizes das componentes de crominância, por sua vez, possuem metade do tamanho, no caso da relação 4:2:2, e um quarto do tamanho, no caso da relação 4:2:0.

1.5 Pixel

O menor elemento de exibição de uma imagem é um pequeno ponto denominado pixel. A palavra vem da combinação das palavras "*PIC*tura" e "*EL*ement", que significam "imagem" e "elemento", respectivamente. O conjunto de milhares de pixels forma a imagem inteira. Logo, ao visualizar uma imagem com alto índice de aproximação é possível identificar pequenos quadrados, ou retângulos, coloridos nela, que somados formam o desenho completo, como é demonstrado na figura 5.

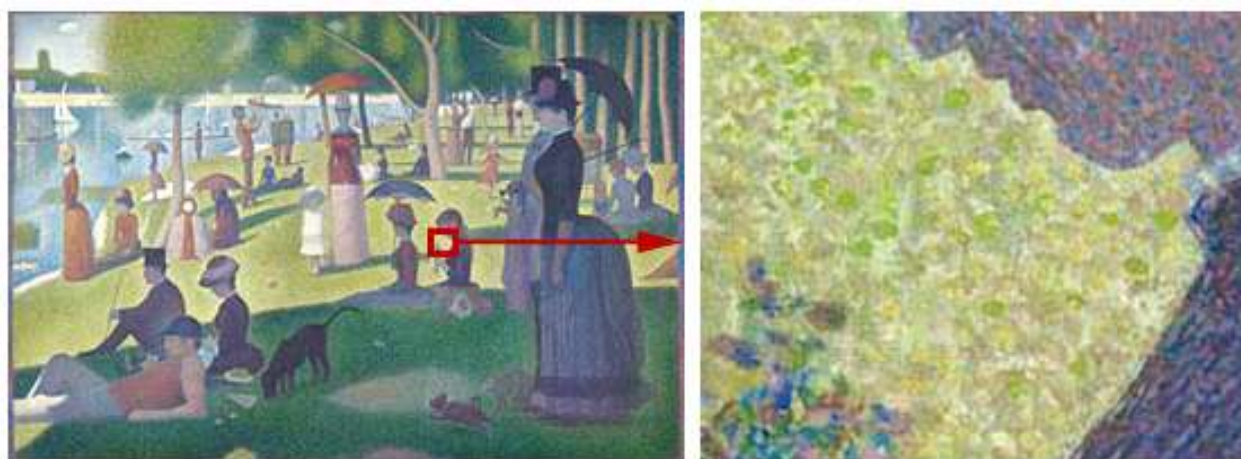


Figura 5 - Representação do Pixel (SAMSUNG, 2010).

Para representar um pixel é definido um número fixo de bits, os quais definem a quantidade de cores possíveis no sistema. Por exemplo, em sistemas *truecolor* cada componente possui 24 bits (oito bits para cada cor primária) para representar um pixel, ou seja, ao combinar as tonalidades dos três pontos é possível obter 2^{24} cores diferentes.

1.6 Bloco e Macrobloco

Um bloco é formado por um grupo de pixels representados em uma matriz 8×8 . O macrobloco é um conjunto de blocos, cuja dimensão é usualmente 16×16 , mas podem ser representados de várias maneiras diferentes de acordo com o fator de subamostragem. A representação de bloco e macrobloco é mostrada na figura 6.

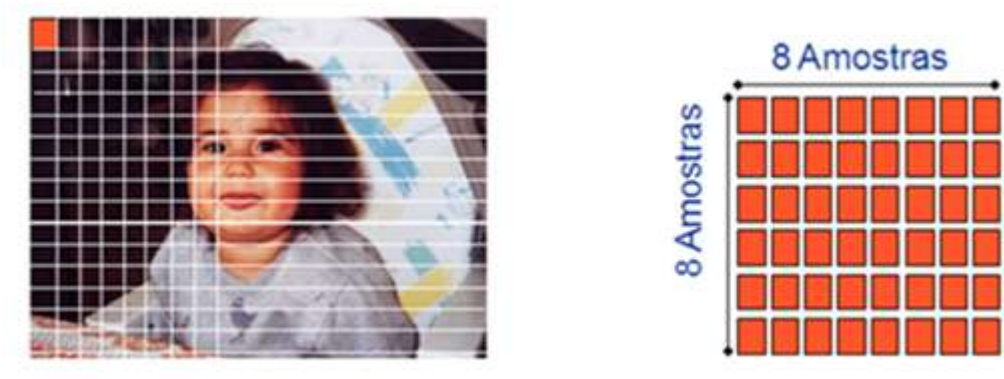


Figura 6 - Representação de Macrobloco e Bloco, respectivamente (INATEL, 2010).

1.7 Fatia

Uma fatia (*slice*) é formada pela organização dos macroblocos em uma mesma linha horizontal, como mostra a figura 7.

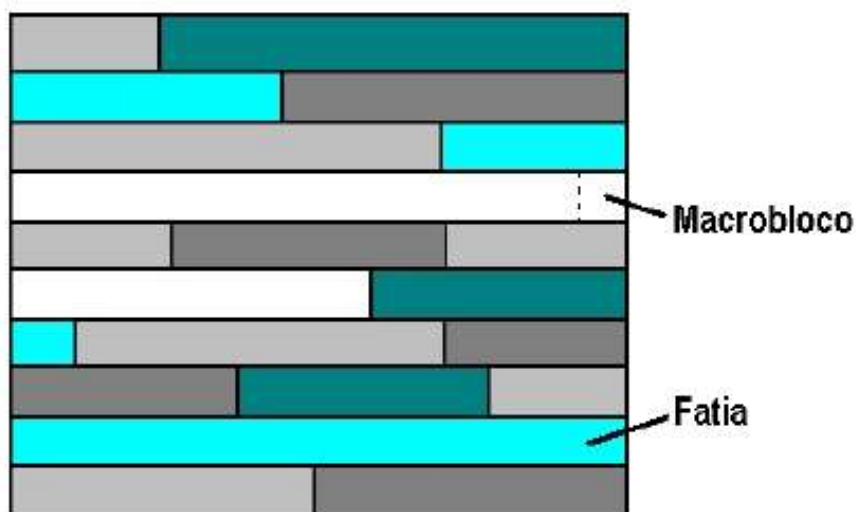


Figura 7 - Representação de Fatia (MANZATO, 2004).

Um vídeo pode conter uma fatia para cada frame e/ou uma fatia para cada macrobloco. A divisão da imagem em fatias possibilita a decodificação de toda a imagem de maneira independente. Desse modo, mesmo com a presença de erros, o restante da imagem pode ser decodificado com sucesso mesmo com a ocorrência de algum erro de codificação em algum trecho da imagem.

1.8 Quadro (Frame)

Na realidade os quadros são imagens, que no padrão MPEG podem ser de três tipos diferentes. O tipo da imagem define qual o modo de predição pode ser usado para codificar cada macrobloco (MANZATO, 2004).

- Quadros I (Intra): contêm toda informação requerida para reconstruir o quadro original. Não apresenta compressão por redução de redundância temporal, mas uma compressão moderada pela redução de redundância espacial. Esse tipo de quadro é muito importante por possibilitar pontos de acesso na sequência de bits onde a decodificação pode começar sem referência a imagens prévias. Permite, no receptor, o início da decodificação em um ponto intermediário na sequência total de bits.
- Quadros P (Preditiva): eles são codificados utilizando predição por movimento a partir de um quadro I ou de outro P passado e podem ser utilizadas como referência para uma predição futura (*forward*). Oferecem maior índice de compressão comparado aos do tipo I.
- Quadros B (Bidirecionalmente Preditiva): podem usar quadros I ou P passados e/ou futuros para compensação de movimento. É o quadro com maior grau de compressão.

1.9 Amostragem

A amostragem pode ser definida como o processo de medição instantânea de valores de um sinal analógico em intervalos regulares, como mostrado na figura 8. O intervalo entre as amostras é determinado por um pulso de sincronismo e a sua frequência é chamada de taxa de amostragem (Wikipedia, 2011).

Em 1928, Henry Nyquist dos Laboratórios Bell, estabeleceu que a representação digital de um sinal analógico seria funcionalmente idêntico à forma de onda original se a taxa de amostragem fosse pelo menos duas vezes a maior frequência presente na forma de onda analógica. A frequência de amostragem é dada em amostras por segundo.

$$\text{Frequência de amostragem} = 2 * \text{Frequência máxima}$$

Quanto maior for a taxa de amostragem, mais medidas do sinal são realizadas em um mesmo intervalo de tempo, e assim, maior será a fidelidade do sinal digital em relação ao sinal analógico. É, portanto, assim como a quantização, que atribui valores discretos para um sinal cuja amplitude varia entre infinitos valores, uma propriedade que define a fidelidade da conversão.

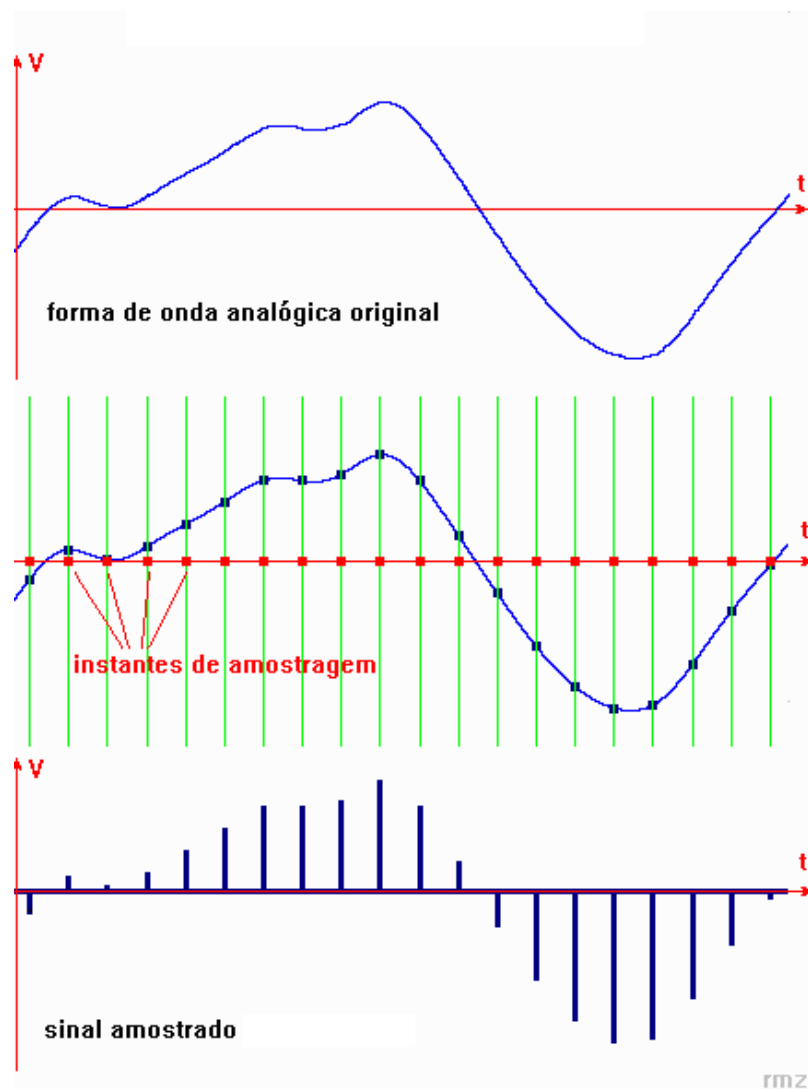


Figura 8 - Amostragem de um sinal analógico (ARANGO, 2011).

2 INTRODUÇÃO A COMPRESSÃO DE VÍDEO

Antes de abordar os métodos/algoritmos de compressão de vídeo, algumas informações importantes são citadas neste capítulo, baseado no tutorial sobre compressão digital de vídeo, publicado por Hugo Melo.

Em vídeo há dois tipos de componentes de sinal: os novos, que são imprevisíveis, e os que podem ser previstos. O componente novo é chamado de entropia, formando as informações reais do sinal. O restante é chamado de redundância porque não é essencial. A redundância pode ser espacial, como em áreas da imagem onde vários pixels adjacentes têm o mesmo valor. A redundância também pode ser temporal, quando há semelhanças entre imagens sucessivas. Os sistemas de compressão separam a entropia da redundância no codificador (*encoder*). Apenas a entropia é gravada ou transmitida e o decodificador (*decoder*) calcula a redundância a partir do sinal recebido.

Um codificador ideal extrai toda a entropia e apenas ela precisa ser transmitida para o decodificador. Na prática não se consegue isso. Uma parte da redundância também tem de ser transmitida. Um codificador ideal seria complexo e caro demais e causaria um retardo muito grande, impedindo o uso da redundância temporal. Em gravação ou em broadcasting pode haver algum retardo no processamento do sinal, mas em transmissões ao vivo, não.

Na prática, é necessário utilizar um conjunto de codificadores com uma faixa de retardos e diferentes complexidades. O poder do MPEG vem do fato de não ser só um formato de compressão, mas um conjunto de ferramentas de codificação padronizadas que podem ser combinadas de forma flexível para atender a diversas aplicações. A forma como a codificação é feita é incluída nos dados comprimidos, informando ao decodificador como reconhecer os dados.

2.1 Perfis e Níveis

A codificação MPEG se divide em diversos perfis, cada um com um grau de complexidade. E cada perfil pode ser classificado em diferentes níveis, dependendo da resolução da imagem de entrada. Um perfil de codificação define os requisitos do codificador ou decodificador e especifica um conjunto de ferramentas de codificação de acordo com a finalidade do perfil.

Existem diversos formatos de vídeo digital e cada um tem uma taxa de bits (bit rate) diferente. Um sistema de alta definição pode ter uma taxa de bits seis vezes maior que um

sistema de TV *standard*. Logo, não basta saber a taxa de bits do codificador. O importante é o fator de compressão, que é a razão entre a taxa de bits de entrada e a taxa de bits comprimida, por exemplo, 5:1.

Mas o número de variáveis envolvidas torna difícil determinar um fator de compressão adequado. Para um codificador ideal, se toda a entropia for enviada, a qualidade é boa. Mas se o fator de compressão for aumentado para reduzir a taxa de bits, nem toda a entropia é enviada e a qualidade cai. Quando ocorre perda de qualidade em um sistema comprimido, a compressão é alta.

Para identificar a entropia perfeitamente, um compressor ideal seria extremamente complexo. Um compressor prático pode ser menos complexo por motivos econômicos e deve enviar mais dados para garantir o envio de toda a entropia. Quanto maior o fator de compressão, mais complexo é o codificador.

2.2 Requisitos do MPEG para vídeo

O padrão MPEG é um padrão genérico, ou seja, a padronização é independente de uma aplicação em particular, isso não significa que os requerimentos das aplicações são ignorados. A seguir são mostradas as principais características necessárias a aplicação do MPEG:

2.2.1 Acesso Randômico

Acesso randômico é uma característica essencial para um armazenamento médio, sendo ou não a mídia um acesso randômico como um CD ou um disco magnético ou ainda uma mídia sequencial nos drives. Acesso randômico implica na existência de pontos de acesso, isto é, segmentos de informação codificados em relação a uma referência.

2.2.2 Buscas à frente/atrás rápidas

Se o meio de armazenamento permitir tal característica, torna-se possível o mapeamento da sequência de bits comprimidos e utilizando a propriedade de pontos de acesso, apresentar figuras selecionadas para obter o efeito de um avanço rápido ou uma reversão rápida.

2.2.3 Reprodução para trás

As aplicações interativas e edição requerem que o sinal de vídeo seja apresentado no modo reverso. Embora não seja necessário para todas as aplicações manter a total qualidade no modo reverso ou até mesmo ter o modo reverso em todas as aplicações, foi possível perceber que esta característica era possível de ser aplicada sem um alto custo adicional de memória.

2.2.4 Sincronização áudio-vídeo

O sinal de vídeo pode ser sintonizado associado com a fonte de áudio. O mecanismo pode ser provido para permanente ressincronização de áudio e de vídeo, podendo os dois sinais serem derivados de *clocks* (temporização) levemente diferentes. Esta característica é definida pelo grupo de sistema MPEG que dita a definição de ferramentas para sincronização, bem como a integração de múltiplos sinais de vídeo e áudio.

2.2.5 Resistência a erros

A maioria dos meios de armazenamento digital e canais de comunicação não são livres de erros, e isso é esperado no esquema apropriado de codificação de canais. Assim, a fonte do esquema de codificação deve ser resistente aos erros que não foram corrigidos e aos comportamentos catastróficos, devendo haver uma reação prevista à presença de erros.

2.2.6 Atraso de codificação/decodificação

Aplicações como a videotelefonía devem manter o tempo de atraso total do sistema abaixo de 1,5 ms para que se possa manter a conversação normal, a natureza “*face a face*” da aplicação. Por outro lado, aplicações de publicação e *broadcasting* podem ter um tempo de atraso maior, aproximadamente 1 segundo, que não são prejudiciais ao resultado da transmissão.

2.2.7 Editibilidade

O formato deve permitir comprimir/descomprimir um pequeno número de imagens com eficiência e rapidez, para viabilizar a edição eletrônica (MELO, 2004).

Sabe-se que toda figura não pode ser comprimida independentemente, sendo desejável a possibilidade de construção de unidades de edição de pequenas durações, e com isso um aceitável nível de editabilidade na forma comprimida é obtida.

2.2.8 Flexibilidade do Formato

O paradigma do computador “*video in a window*” pressupõe uma enorme flexibilidade de formatos em termos de tamanho (altura e largura) e taxa de moldura. Diferentemente da resolução padronizada da TV (número de linhas e quantidade de quadros/segundos) e da relação de aspecto padrão (4x3, 16x9), o formato deve permitir resoluções e relações de aspecto compatíveis com diversas mídias e padrões.

2.3 Problemas da Compressão

As técnicas de compressão de vídeo digital sujeitas a perdas, como o MPEG, podem criar artefatos digitais durante o processo de compressão. Os artefatos digitais podem se distorção da cor, degradação da cor, degradação do movimento, aumento do ruído, duplicação de quadros, varredura e geração de blocos (OLIVEIRA; CABRAL, 2007).

O telespectador não está interessado nesses problemas, o que ele vai exigir é um padrão de qualidade de imagem e som muito superior ao atual, em uma transmissão digital. Ou seja, diferentes padrões de compressão serão utilizados na transmissão e no processamento de sinais na produção e pós-produção de programas.(MELO, 1999).

O artefato digital mais comum gerado pelo comprimido MPEG é o efeito de *blocking* ou mosaico. Esse efeito é representado pela presença de blocos padrão 8x8 (pixels) no fluxo de vídeo comprimido que não fazem parte do vídeo original, causado pelo uso do algoritmo DCT (Transformada Discreta do Cosseno), descrito no próximo capítulo.

Os artefatos de vídeo digital podem ser eliminados recorrendo-se a um variado conjunto de técnicas. A maioria deles pode ser eliminada aumentando o bit rate (quantidade de informação) médio usado na compressão dos conteúdos. A filtragem do fluxo de vídeo para eliminar o ruído de alta frequência é também uma técnica comum para reduzir os artefatos.

2.4 Técnicas de Compressão de Vídeo MPEG

Em geral, seqüências de vídeo contêm uma quantidade significativa de redundância estatística e subjetiva entre e nos quadros. O objetivo de um sistema de codificação de vídeo é a redução do número de bits a serem armazenados ou transmitidos explorando as redundâncias e as informações menos relevantes (definido pelas limitações do ser humano), além de codificar um conjunto mínimo de informação usando técnicas de codificação por entropia.

O desempenho das técnicas de compressão de vídeo depende da quantidade de redundância contida na imagem, bem como nas técnicas usadas na codificação. A codificação de vídeo MPEG é estatística por natureza. A propriedade estatística básica sobre a qual esta codificação se baseia é a correlação entre pixels, incluindo a suposição de simples correlação de movimento translacional entre quadros consecutivos. Diversos algoritmos/métodos são utilizados para a compressão de vídeo MPEG e são mostrados neste capítulo.

2.4.1 Redundância Temporal

A redundância temporal explora as familiaridades entre os quadros consecutivos. Em um vídeo qualquer, se cada quadro for analisado separadamente, perceber-se-á que entre dois quadros ou mais consecutivos existem muitos dados semelhantes entre si. Em uma seqüência de imagens sem ação, como um telejornal, onde são poucas as cenas em que existem movimentos bruscos, essa característica é mais visível.

Isso pode ser utilizado na compactação do vídeo, pois os dados que se repetem são armazenados (ou transmitidos) apenas uma vez, e são compartilhados posteriormente por vários quadros na reconstrução da imagem. Essa técnica é chamada de Estimativa e Compensação de Movimento, e está presente em todos os codificadores que suportam a eliminação da redundância temporal.

2.4.1.1 Estimativa e Compensação do Movimento

O desempenho da predição do movimento é satisfatório para imagens com baixa dinâmica, no entanto, para imagens animadas reais deve-se utilizar técnicas mais eficientes. Para melhorar o esquema de predição entre imagens, utilizamos vetores de deslocamentos para indicar a nova posição dos detalhes na imagem corrente. Conseqüentemente, o vetor de movimentos indica as coordenadas de todos os blocos comprimidos no quadro anterior que

deverão ser repetidos no quadro atual em uma nova posição, como mostrado na figura 9. Esse esquema é conhecido como previsão com compensação de movimento.

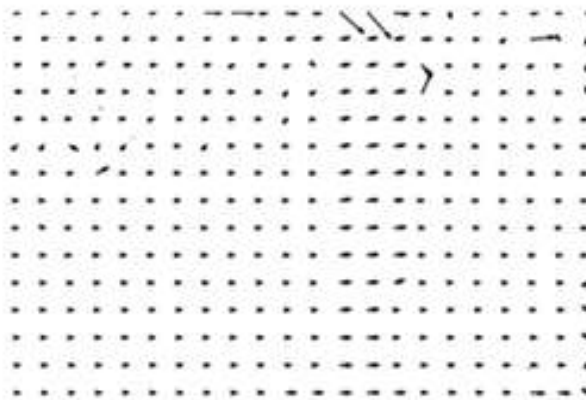
A estimação de movimento é obtida tomando uma aproximação do macrobloco em apenas sinais de luminância. O deslocamento do vetor é estimado por cada macrobloco. Os vetores de movimento e diferença codificada em DCT (transformada discreta do cosseno) do macrobloco atual e o macrobloco de referência, apontado por esse vetor de movimento, são transmitidos juntos e possibilitam uma solução mais eficiente que a simples codificação do macrobloco com DCT.



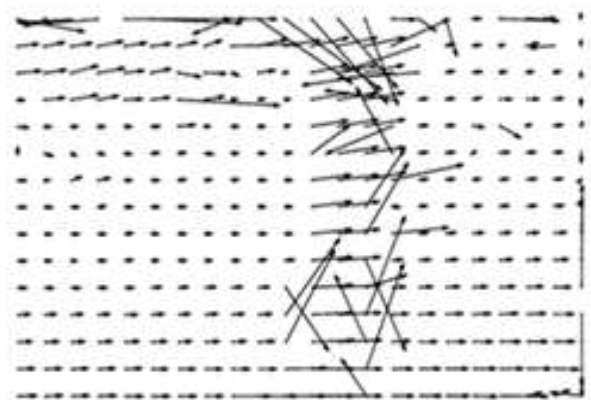
Quadro no. 2



Quadro no. 6



Quadro 2 para 3



Quadro 2 para 6

Figura 9 – Vetores de deslocamento (INATEL, 2010).

2.4.1.1.1 *Block Matching*

Diferentes técnicas são utilizadas para determinar esse vetor de deslocamento. Um dos métodos é o chamado *Block Matching*. Nele o bloco de pixels, chamado bloco de referência, no quadro atual é movido ao redor de sua posição dentro de uma área de busca no quadro anterior. Os valores do bloco DCT de referência são, então, comparados aos valores dos blocos da área de busca, a fim de encontrar o bloco que mais se parece, como mostrado na figura 10. Então, quando a diferença é mínima, temos o melhor casamento entre blocos. Um vetor de movimento que melhor mostra as coordenadas do bloco é então gerado e, então, transmitido com o bloco DCT de coeficientes da diferença.

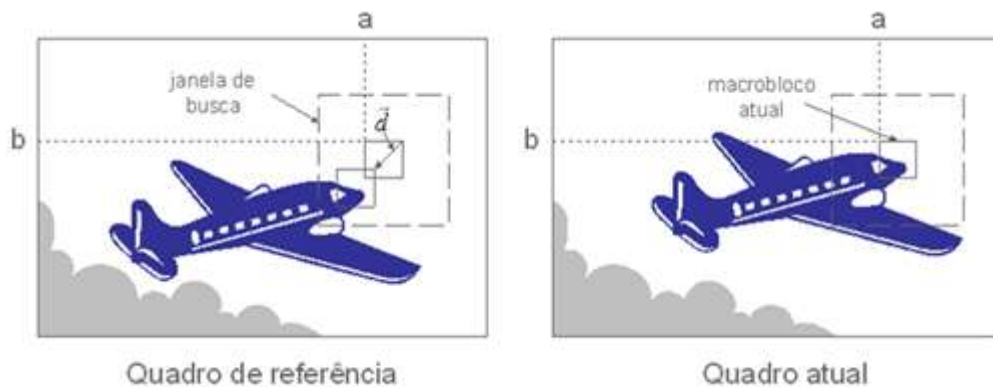


Figura 10 - Vetor de movimento obtido pelo *Block Matching* (INATEL, 2010).

2.4.1.1.2 *Predição Bidirecional*

Outra forma de predição utilizada no padrão MPEG é 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, como mostrado na figura 11 (C. CASTRO; F. CASTRO, 2002).

A predição bidirecional torna-se possível quando áreas de detalhes podem ser visíveis com o movimento das cenas. Para permitir a predição *backward*, a partir de um frame futuro, o codificador ordena as imagens de modo a transmiti-las em uma ordem diferente das mostradas. Isso causa um pequeno *delay* (atraso) na transmissão, sendo prejudicial para algumas situações, como em transmissões ao vivo.

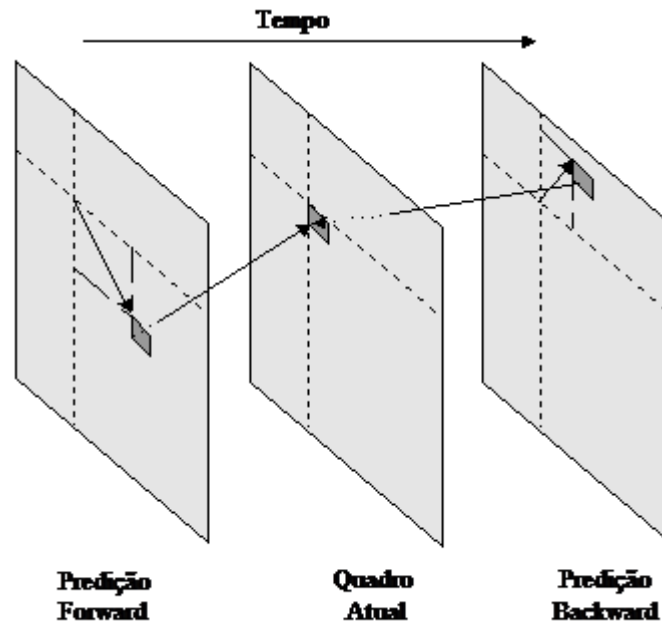


Figura 11 - Predição bidirecional (C. CASTRO; F. CASTRO, 2002).

2.4.1.2 GOP (*Group of Pictures*)

Um fluxo de vídeo MPEG é composto por grupos de imagens (GOP's). O MPEG utiliza a predição temporal para, a partir de um quadro, estimar o próximo. Um grupo de imagens é uma sequência de quadros comprimidos que começa com uma imagem chamada quadro I. O quadro I representa a imagem completa, os quadros P e B só guardam as mudanças do quadro anterior para o posterior. O grau de compressão resultante do processo de codificação MPEG varia, dependendo do conteúdo do fluxo de vídeo.

Os frames são organizados de modo que as imagens a serem codificadas como sendo do tipo I permaneçam no início do grupo de frames; já aquelas que serão codificadas como do tipo P, devem ficar em uma posição posterior a uma já existente do tipo I; e aquelas que serão codificadas como do tipo B devem permanecer entre duas outras que são do tipo I ou do tipo P, conforme mostrado na figura 12.

Para uma dada qualidade de imagem decodificada, a codificação usando cada tipo de imagem produz um número diferente de bits. Em uma sequência típica de vídeo, uma imagem codificada do tipo I gera três vezes mais bits que uma imagem codificada do tipo P, a qual por sua vez gera 50% mais bits que uma imagem codificada do tipo B (C. CASTRO; F. CASTRO, 2002).

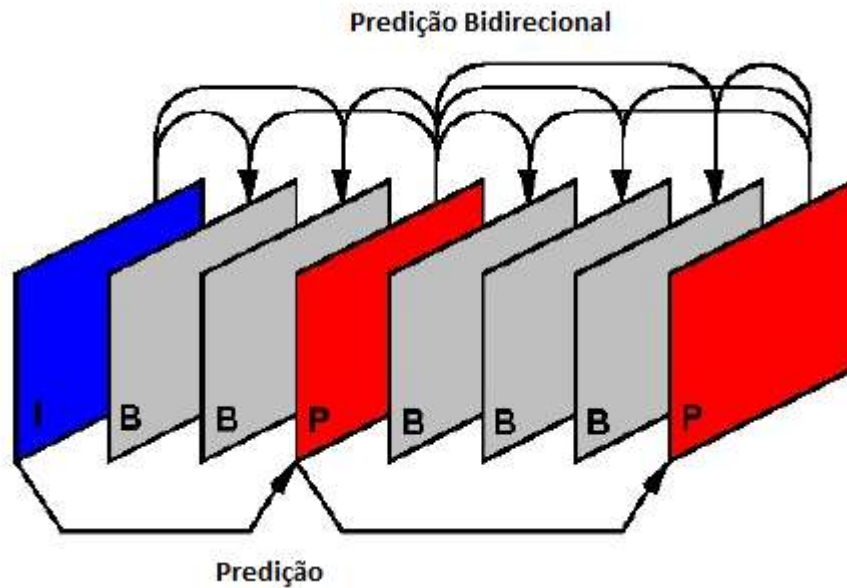


Figura 12 - Estrutura do GOP: a ponta da seta indica a imagem resultante do processo de predição e a origem da seta indica a imagem de referência usada no processo de predição (MANZATO, 2004).

2.4.2 Redundância Espacial

Após a etapa de redundância temporal finalizada, o erro de predição resultante é submetido às etapas da redundância espacial, onde as redundâncias da própria imagem são exploradas e reduzidas ao máximo, tendo como a principal etapa o processo da transformada discreta do cosseno (DCT).

2.4.2.1 Transformada Discreta do Cosseno (DCT)

Os algoritmos de compressão MPEG usam técnicas de Transformada Discreta do Cosseno em blocos de imagens para explorar de forma eficiente a correlação espacial entre pixels próximos na mesma imagem. Esse processo não gera perdas, e também é completamente reversível, utilizando a IDCT (transformada discreta do cosseno inversa).

A transformada discreta do cosseno é um caso particular da transformada de Fourier aplicada a sinais discretos (amostras), que decompõe um sinal periódico em uma série de funções harmônicas de senos e cossenos. O sinal pode ser representado por uma série de coeficientes de cada uma dessas funções.

A DCT decompõe o sinal em apenas uma série de harmônicos de funções cosseno em fase com o sinal, o que reduz pela metade o número de coeficientes necessários para se descrever o sinal comparado com a transformada de Fourier.

No caso de figuras, o sinal original é uma amostra bidimensional, e então pode-se também ter uma DCT bidimensional, que transformará os valores discretos de luminância (ou crominância) de um bloco de $N \times N$ pixels em um outro bloco (ou matriz) de $N \times N$ coeficientes representando a amplitude de cada função harmônica do cosseno. A transformada discreta do cosseno é calculada utilizando a equação (1):

$$Z_{k,l} = \frac{1}{4} C_k C_l \sum_{i=0}^{i=7} \sum_{j=0}^{j=7} x(i,j) \cos \frac{\pi(2i-1)k}{16} \cos \frac{\pi(2j+1)l}{16} \quad (1)$$

Se a energia de vídeo da imagem é de baixa frequência espacial, ou seja, varia lentamente no espaço, então uma transformada pode ser usada para concentrar a energia em poucos coeficientes. Para simplificar a complexidade da transformada, a imagem é dividida em pequenos blocos de, geralmente, 8×8 pixels onde é aplicada uma DCT bidimensional, como observado na figura 13.

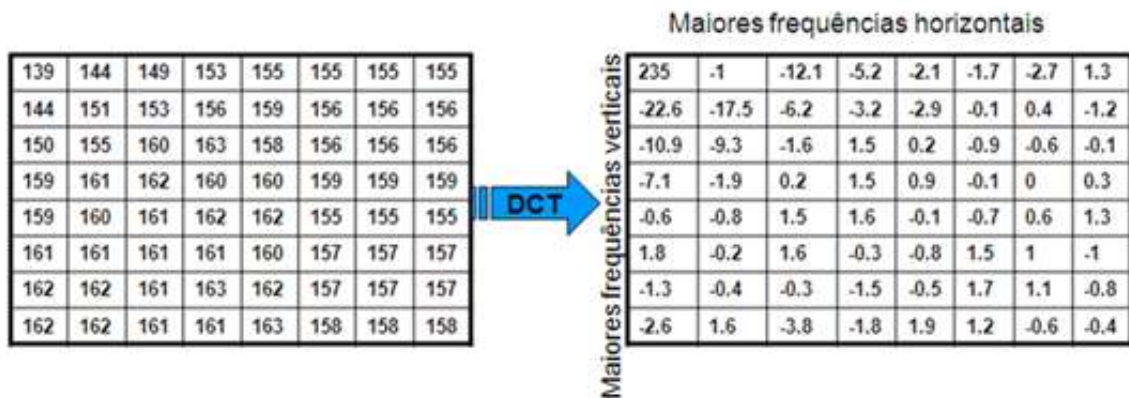


Figura 13 - Matriz resultante do processo DCT (INATEL, 2010).

No bloco transformado, como o mostrado na figura 13, coeficientes na coordenada horizontal representam acréscimo nas frequências horizontais da esquerda para a direita, e na coordenada vertical eles representam acréscimo nas frequências de cima para baixo. O

primeiro coeficiente no canto superior esquerdo representa frequência nula nas frequências horizontais e verticais, e é então chamado de coeficiente DC.

A saída transformada da DCT bidimensional é ordenada de forma que o valor médio, ou coeficiente DC, está no canto superior esquerdo e os coeficientes de altas frequências aumentem com a distância ao coeficiente DC. As altas frequências verticais são representadas por índices grandes para as linhas e as altas frequências horizontais são representadas por índices grandes para as colunas.

Dependendo do número de detalhes que o bloco original contém, os coeficientes de alta frequência serão grandes ou pequenos, mas em geral a amplitude decresce rapidamente na maioria das imagens naturais. Dessa forma o DCT possui a propriedade de concentrar a energia do bloco em um número relativamente baixo de coeficientes, situados nos blocos do canto superior esquerdo da matriz. Essas duas propriedades são usadas como benefício nos próximos passos do processo de compressão.

2.4.2.2 Quantização

No processo de compressão a quantização é aplicada nos coeficientes gerados pela transformada do cosseno. Uma matriz de pixels original é composta de elementos inteiros, mas após a transformada tem-se uma matriz de coeficientes reais. Para reduzir o espaço necessário para armazenar esses valores aplica-se a quantização, que é o processo de redução do número de bits necessários para armazenar um valor reduzindo sua precisão para um inteiro. É nesta redução da precisão que acontece o processo irreversível de perda da informação (NELSON, 1991).

Esse processo é obtido por meio de uma matriz de quantização que vai produzir redução de precisão em cada elemento. Cada elemento da matriz de coeficientes é dividido pelo elemento correspondente da matriz de quantização e seu valor arredondado para o inteiro mais próximo.

Quanto maior o valor do fator de quantização, maiores serão as perdas. O fator de quantização é considerado entre 2 e 25 e os índices iniciam no zero. Para valores maiores que 25, haverá um comprometimento muito severo na qualidade da imagem reconstruída. Dessa maneira é feita uma quantização por zona, sendo os coeficientes associados às mais altas frequências quantizados com mais severidade.

$$Q[i, j] = 1 + (1 + i + j) * (\text{fator de quantização})$$

A figura 14 ilustra uma matriz resultante do processo de quantização.

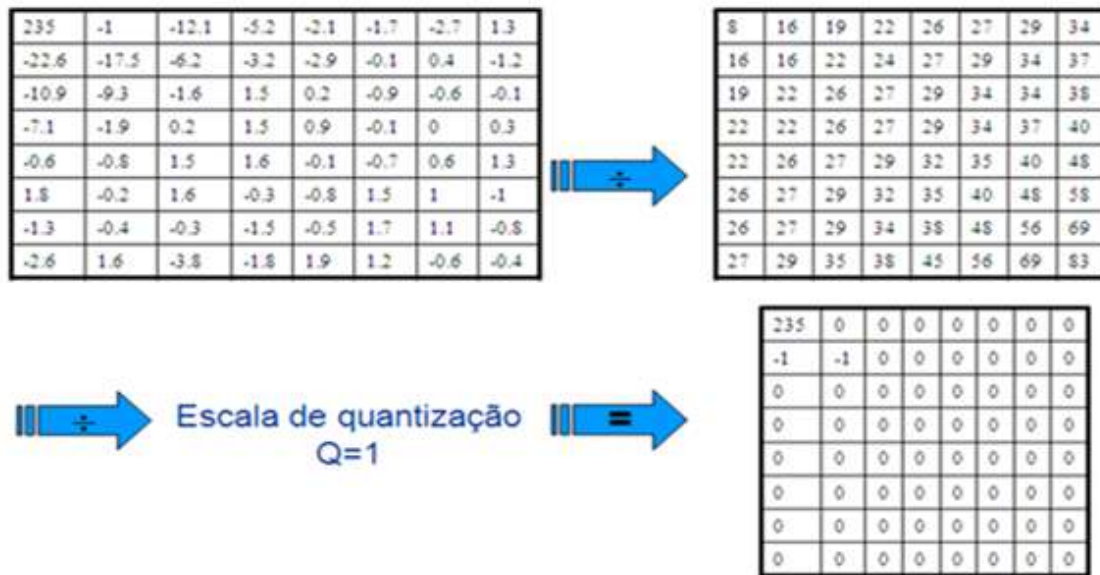


Figura 14 - Matriz resultante do processo de quantização (INATEL, 2010).

Os blocos resultantes do processo de DCT passam por um processo de quantização, em que os bits com valores insignificantes são arredondados para zero. A eliminação de dados psicovisualmente redundantes resulta numa perda de informação. Como essa operação é irreversível, ou seja, informação visual é perdida, a quantização resulta em compressão de dados com perda (GONZALEZ; WOODS, 2000).



Figura 15 - Percepção visual humana às frequências espaciais (INATEL, 2010).

A percepção visual humana do ruído em uma figura não é uniforme, mas é uma função da frequência espacial. Quanto maior a frequência espacial mais ruído pode ser tolerado, como observado na figura 15. Além disso, o ruído de vídeo é mascarado pelo detalhe fino na imagem, apesar de que, em imagens planas serem altamente visíveis. Com isso, podemos quantizar as altas frequências sem que isso nos traga algum prejuízo na qualidade da imagem.

Os coeficientes da DCT são quantizados para reduzir sua amplitude e aumentar o número de coeficientes iguais a zero descartando informações que são visualmente insignificantes.

2.4.2.3 Varredura Ziguezague

Os coeficientes da DCT são quantizados para reduzir sua amplitude e aumentar o número de coeficientes iguais a zero descartando informações que são visualmente insignificantes. Após a quantização, os blocos 8x8 pixels passam por uma varredura em ziguezague que é feita diagonalmente, do topo para esquerda, com intuito de gerar o maior número possível de zeros consecutivos.

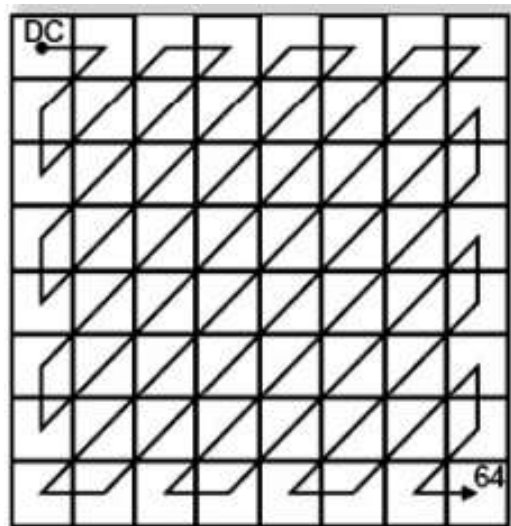


Figura 16 - Varredura ziguezague (FAPTECH, 2004).

As baixas frequências tendem a ter coeficientes maiores e as altas frequências são, pela natureza da maioria das imagens, predominantemente zero. Como a mostra a figura 16, os coeficientes AC passam por um processo de ziguezague, da posição (0,1) à posição (7,7).

Dessa forma, obtém-se uma longa série de zeros consecutivos, facilitando as próximas etapas de codificação.

2.4.2.4 Codificação VLC (*Variable Length Coding*)

Esta codificação é um processo que identifica elementos que ocorram mais frequentemente e usa palavras com poucos bits para codificar, aumentando-se a número de bits, até 12, inversamente à probabilidade das ocorrências (SILVA, 2008). Como a quantificação, a codificação VLC produz tabelas que permitem mapear as figuras a codificar. Essas tabelas, combinadas com os mapas dos códigos, permitem, geralmente, diminuir a quantidade de dados. Nota-se que o processo VLC (ou RLE) não provoca perdas.

Cada palavra do código VLC contém uma série de zeros seguida por um coeficiente com um nível particular diferente de zero. A codificação VLC reconhece que são mais freqüentes as pequenas séries de zeros que as longas e que os pequenos coeficientes são mais comuns que os grandes.

Para permitir ao decodificador distinguir quando é que uma palavra do código termina e começa outra, no VLC não existem palavras de código que tenham o mesmo prefixo. As combinações que não se encontram na tabela de codificações VLC são codificadas pela combinação de um código de escape e dois códigos de comprimento fixo, sendo uma palavra de 6 bits para representar o tamanho do código e uma palavra de 12 bits para representar o caractere (C. CASTRO; F. CASTRO, 2002).

3 O PADRÃO MPEG

O *Moving Picture Experts Group* (MPEG) foi formado em janeiro de 1988, como um grupo de trabalho para criar padrões internacionais de codificação/decodificação de áudio e vídeo (ALENCAR, 2007). A necessidade de armazenamento e reprodução de filmes (imagem animada) e o formato digital de áudio em várias plataformas levaram a ISO (*International Standard Organization*) a formar um grupo de especialistas nas mesmas linhas do JPEG (*Joint Photographic Expert Group*), com membros vindos dos vários setores envolvidos.

É importante ressaltar que 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 dependem 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 (PUC-Rio, 2007).

Este capítulo abordará as características de cada padrão MPEG, dando maior ênfase aos padrões MPEG-4 e H.264, que são os padrões mais utilizados nas reproduções de vídeo com alta qualidade.

3.1 Padrão JPEG

Em 1983, a ISO (*International Standard Organization*) criou um grupo de trabalho internacional chamado JPEG (*Joint Photographic Expert Group*) que tinha a tarefa de elaborar um padrão internacional de compressão para imagens estáticas para várias resoluções em Y, Cr, Cb ou RGB. O padrão internacional resultante (comumente conhecido como JPEG) foi publicado na referência ISO/IEC 10918, e pode ser considerado uma referência de compressão de imagens e o grande responsável pela viabilização do MPEG. O JPEG não será descrito com riqueza de detalhes, pois não é o alvo de estudo deste trabalho.

Quando a empresa *C-Cube Microsystems* desenvolveu uma das primeiras placas de compressão JPEG, notou-se que essa placa conseguia descompactar em 1/30 de segundos, coincidentemente a mesma velocidade necessária para reproduzir vídeos no padrão NTSC³. Essa placa compressora possibilitou a realização das primeiras experiências de compressão de vídeo, em que se compactava quadro a quadro usando padrão JPEG, sem nenhuma

³ NTSC - *National Television Standard Committee*: sistema de televisão em cores utilizado pelo Japão e EUA.

preocupação com redundância temporal. Esse foi o ponto de partida para o desenvolvimento de métodos mais elaborados para compressão de vídeo, como o MPEG (Revista Mackenzie de Engenharia e Computação, Ano 1, n.1).

3.2 Padrão MPEG-1

O primeiro resultado deste trabalho foi o MPEG-1. Publicado em 1989 e aprovado como padrão internacional em 1992, o principal objetivo era permitir o armazenamento de imagens com áudio e vídeo integrados, previamente digitalizadas em CD-ROM ou VCD (*Video Compact Disc*) de vídeo e som estéreo, que implicava numa máxima taxa de 1,5Mb/s (JURGEN, 1997).

Além da exploração da redundância espacial demonstrada pelo JPEG, o MPEG-1 explora a redundância temporal entre sucessivas imagens que compõem uma seqüência de vídeo, podendo atingir uma taxa de compressão de 200:1 (POYNTON, 2003).

Essas e outras características tornaram o MPEG-1 o formato para distribuição de vídeo pela web, representando a primeira oportunidade concreta encontrada pela indústria de microeletrônica para investir em vídeo digital.

Com a taxa máxima as imagens ficam distorcidas, porque a compressão introduz artefatos na imagem, presentes em maior ou menor número, dependendo da qualidade do algoritmo de compressão e da taxa de compressão utilizados. Assim, a maioria dos vídeos comprimidos com MPEG-1 utiliza compressão menor que 50:1. Mesmo com essa taxa, a resolução horizontal obtida após a compressão é baixa, cerca de 320 linhas, e semelhante ao formato VHS (OLIVEIRA; CABRAL, 2007).

O fato é que o padrão MPEG-1 não era adequado para aplicações broadcast visto que, além disso, não contemplava imagens com exibição entrelaçada e a evolução para o HDTV. O MPEG, então, trabalhou na definição de um padrão mais flexível e aplicável ao broadcast. Esse novo padrão chamou-se MPEG-2.

3.3 Padrão MPEG-2

Em 1990, iniciaram os estudos para um segundo padrão de vídeo, capaz de codificar imagens entrelaçadas diretamente, com capacidade de suportar alta qualidade a taxas de transmissão de 5 a 10 Mbits/s. Este segundo padrão foi denominado MPEG-2.

Aprovado em 1994, o padrão MPEG-2 foi desenvolvido para vídeo com vários níveis de qualidade e aplicações em radiodifusão terrestre ou por satélite, TV a cabo, estúdio e HDTV. Ele foi desenvolvido com base no MPEG-1, proporcionando mudanças significativas na codificação de vídeo, como mostrado na figura 17. Foi o primeiro padrão a introduzir os conceitos de perfis e níveis.

Esta norma mantém todas as características do MPEG-1, contudo, acrescenta várias formas de escalabilidade. A escalabilidade é útil para, por exemplo, transmitir o mesmo vídeo a decodificadores normais, que o devem decodificar com a resolução básica, e a decodificadores HD (*High Definition*), que o podem decodificar com a resolução mais alta. Também é útil para garantir, no mesmo decodificador, uma qualidade básica em condições difíceis e uma melhor qualidade com melhores condições de recepção.

O padrão MPEG-2 é muito utilizado até hoje pela possibilidade de codificar imagens entrelaçadas, como a compatibilidade com os aparelhos de televisão analógica, por ser usado na transmissão da televisão digital, também em definição normal e HD, em vários canais: satélite, cabo e antenas terrestres. Outra grande aplicação da norma é a gravação de vídeo em DVD (*Digital Video Disc*).

Inicialmente, eram previstas duas normas: uma para a resolução normal e uma para a alta resolução. A norma MPEG-2 acabou por incluir, também, a alta resolução e, como consequência, não surgiu a norma MPEG-3

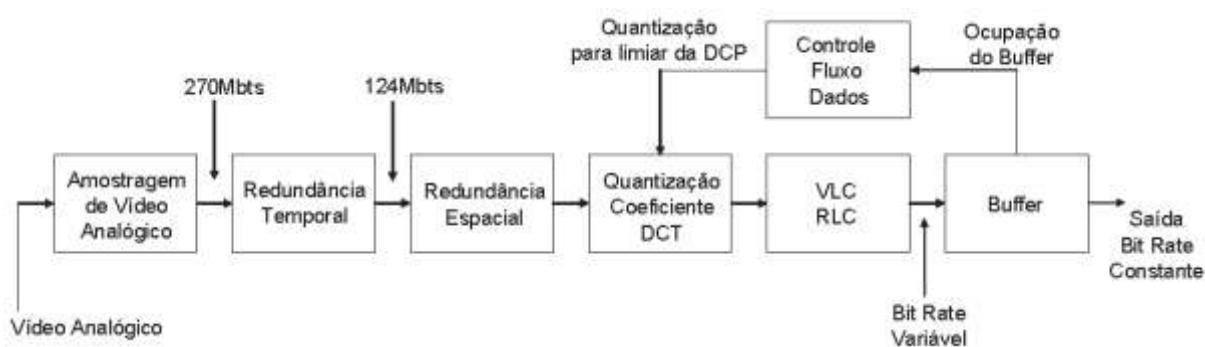


Figura 17 - Codificação em blocos MPEG-2 (Mackenzie de Engenharia e Computação, Ano 5, n. 5).

3.3.1 Perfis e Níveis do MPEG-2

O padrão MPEG-2 pretendia ser genérico, permitindo uma gama variada de informações. Foi o primeiro padrão a introduzir os conceitos de perfis e níveis. Diferentes algoritmos de compressão, desenvolvidos para variadas situações, foram integrados em uma única sintaxe da sequência de bits codificada.

A incorporação 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 tem sido definido. Ainda, dentro de um determinado perfil, um nível é definido, o qual especifica um conjunto de restrições aos parâmetros do perfil, tal como a resolução de imagem.

Em geral, os perfis definidos se ajustam de tal forma que um perfil mais alto é um subconjunto 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 conformidades.

Segue uma lista dos principais perfis (C. CASTRO; F. CASTRO, 2002):

- Perfil Simples: não utiliza frames B, e, portanto, não utiliza predição *backward* ou interpolada. Consequentemente, 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 suportes para imagens B, o que melhora a qualidade de imagem para uma dada taxa de transmissão de bits, mas aumenta o atraso. Presentemente, o maior 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 camadas do padrão MPEG. Cada camada transporta um tipo de informação diferente.

- Perfil Espacial: inclui suporte para ajuste da resolução espacial diferenciadas nas diversas camadas.
- Perfil Alto: inclui suporte para vídeo amostrado 4:2:2.

3.4 Padrão MPEG-4

O MPEG-4 absorve muito das funcionalidades do MPEG-1 e do MPEG-2 e outros padrões relacionados, adicionando novas funcionalidades como o suporte ao VRML (*Virtual Reality Modeling Language*) estendido para renderização 3D, arquivos compostos orientados a objeto (incluindo áudio, vídeo e objetos VRML), suporte a Gerenciamento de Direitos Digitais especificado externamente e vários outros tipos de interatividade.

Inicialmente, o MPEG-4 era destinado a vídeos de baixo *bit-rate*, entretanto, a sua abrangência foi expandida, posteriormente, para ser muito mais que um padrão de codificação multimídia. O MPEG-4 é eficiente através de uma variedade de *bit-rates* indo desde poucos kbits por segundo até dezenas de Mbits por segundo. Ele fornece as seguintes funcionalidades:

- Eficiência de codificação melhorada;
- Possibilidade de codificação de diferentes mídias (vídeo, áudio, fala);
- Flexível a erros para possibilitar transmissões robustas;
- Possibilidade de interação com a cena áudio-visual gerada através do receptor.

Devido aos objetivos ambiciosos com os quais foi concebida, a norma MPEG-4 especifica um numeroso conjunto de ferramentas relacionadas com várias áreas tecnológicas. Para facilitar o seu uso no desenvolvimento de produtos de acordo com a norma, o MPEG-4 está organizado em 23 documentos conhecidos como partes, que podem ser utilizadas em conjunto ou separadamente. Para este trabalho, as partes que se destacam são:

- Parte 2: Visual – define um algoritmo de compressão para informação visual (vídeo, texturas, imagens sintéticas);
- Parte 10: *Advanced Video Coding* (AVC) – descreve um codificador para sinais de vídeo que é tecnicamente igual ao padrão ITU-H H.264.

3.4.1 Principais Características

O padrão MPEG-4 abrange uma vasta gama de aplicações de comunicações visuais, oferecendo suporte com um conjunto de ferramentas de codificação de informação visual. As principais características do padrão que o diferencia dos padrões anteriores são (Richardson, 2003):

- Compressão eficiente de sequências de vídeo entrelaçado ou progressivo;
- Codificação de objetos de vídeo;
- Controle efetivo da transmissão de dados;
- Codificação de objetos visuais animados;
- Codificação para aplicações específicas;
- Compressão de texturas.

3.4.2 Aplicações

A norma MPEG-4 encontra aplicação em muitas áreas. Algumas das mais importantes são citadas abaixo:

- Vídeo na Internet - A transmissão de vídeo através da Internet assumiu um papel cada vez mais importante, como o demonstra a popularidade das notícias, filmes e shows ao vivo que são transmitidos. Como acontece nos canais móveis, na Internet a largura de banda é muito limitada e variável, e pode existir perda de pacotes. Nestes tipos de ambientes, uma melhor tolerância aos erros e uma maior eficiência de codificação fornecidas pela norma MPEG-4 trazem enormes vantagens em termos da qualidade recebida.
- Serviços móveis multimídia - A enorme popularidade dos telefones móveis e dos dispositivos tipo *palm* evidencia o grande interesse provocado por este tipo de equipamento, e pelas correspondentes aplicações. As aplicações nesta área ganham muito com a introdução de informação multimídia. A norma MPEG-4 fornece as ferramentas para tornar as aplicações multimídia em terminais móveis uma realidade.

- Produção de televisão - A utilização de técnicas de produção virtuais, em conjunto com as bem conhecidas técnicas de “fundo azul”. O fundo e os atores são gravados separadamente e podem ser misturados, juntando-lhes ainda efeitos especiais, gerados por computador.
- Televisão digital interativa - O grande crescimento da Internet provocou um aumento do interesse pela interatividade com o conteúdo que a televisão digital pode proporcionar. Texto, imagens, áudio e gráficos podem ser manipulados pelo utilizador para personalizar o conteúdo, como observado na figura 18. Podem-se incluir logótipos das televisões, publicidade escolhida para diferentes tipos de espectador, estatísticas de desporto ou as cotações da bolsa em tempo real.
- Jogos - A popularidade dos jogos de vídeo em computadores pessoais ou em consoles demonstra o interesse existente pela interação entre utilizador e conteúdo. A maioria dos jogos utilizam gráficos tridimensionais quer para o ambiente do jogo quer para as personagens controladas pelo jogador. A integração de objetos de vídeo tornaria estes jogos ainda mais realistas. O acesso a objetos de vídeo e a utilização de técnicas normativas de codificação tornam possível a personalização dos jogos, por exemplo através da utilização de bases de dados de vídeo personalizadas ligadas aos jogos em tempo real.
- Menus inteligentes – ao contrário dos menus de DVD, que ao serem ativados aparecem ocupando toda a tela com o aparelho retornando ao início do disco, os menus inteligentes do MPEG-4 podem ser exibidos da forma que o criador de vídeo quiser. Isso porque a interatividade não é dependente do aparelho reproduzidor de vídeo, mas faz parte do próprio vídeo.



Figura 18 - Televisão com interatividade (Canal Bloomberg, 2007).

3.4.3 MPEG-4 Visual

A norma MPEG-4 é a primeira norma de codificação audiovisual que representa o conteúdo como um conjunto de objetos audiovisuais que compõem a cena em questão e que apresentam um determinado comportamento no espaço e no tempo.

O modelo de representação baseado na composição de objetos audiovisuais está na base de todas as novas funcionalidades oferecidas pela norma MPEG-4 e é a maior diferença conceitual relativamente às normas MPEG anteriores. No lado do emissor, os vários objetos audiovisuais e ainda a informação de composição da cena são codificados separadamente.

3.4.3.1 Objetos de Vídeo

Como descrito anteriormente, o conceito principal da norma MPEG-4 é o de objeto audiovisual, que é o elemento principal do modelo de dados baseado em objetos. Este tipo de representação é apropriado para aplicações interativas e proporciona o acesso de um modo simples ao conteúdo das cenas, permitindo aos usuários acesso e manipulação a eles (cortar, colar, zoom).

Uma cena de vídeo MPEG-4 é construída com objetos independentes que possuem relacionamento no tempo e no espaço. Diferentes objetos podem receber vários tipos de

codificação, adequando-se às características do próprio objeto e sua relevância dentro da cena ou sequência de vídeo. A figura 19 exemplifica uma composição de cena.



Figura 19 – Exemplo de composição de cena (SOLER, 2006).

Existem cinco tipos diferentes de objetos de vídeo usados para representar informação natural (MPEG, 2000):

- *Simple Object* – objeto de vídeo retangular de origem natural, de relação altura/largura arbitrária, desenvolvido para ser usado em baixas taxas de transmissão. Usa ferramentas de codificação simples e de baixo custo.
- *Simple Scalable* – é uma extensão escalável do tipo anterior, provendo escalabilidade temporal e espacial usando o *Simple Object* como base.
- *Core Object* – usa um superconjunto de ferramentas do tipo *Simple Object*, possibilitando melhor qualidade pela utilização de predição bidirecional (B-VOP). Suporta escalabilidade e a utilização de um tom de transparência no objeto.
- *Main Object* – possibilita a mais alta qualidade. Comparado ao tipo *Core Object*, suporta níveis de transparência variáveis e também conteúdos entrelaçados.

- *N-bit Object* – possui todas as características do tipo *Core Object* com a adição da possibilidade de variação da resolução de pixel de 4 a 12 bits para os planos de luminância e crominância.

O padrão MPEG-4 Visual oferece suporte à representação de objetos de origem natural, capturados por um dispositivo tal como uma câmera, ou artificial, produzidos por unidades computacionais. Esses objetos de vídeo são codificados como entidades separadas dentro do fluxo de bits. O modelo de fluxo de bits do padrão MPEG-4 Visual consiste na seguinte hierarquia de classes, como observado na figura 20 (SOLER, 2006):

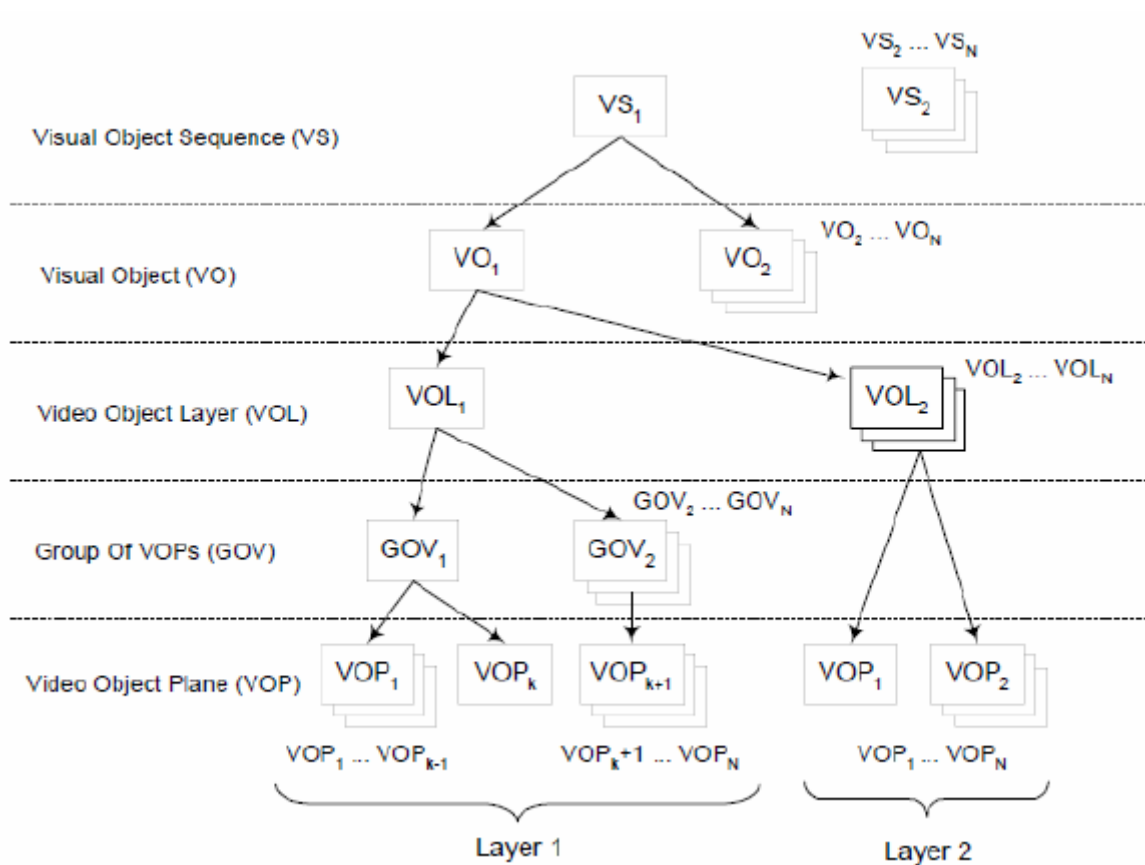


Figura 20 – Estrutura hierárquica de vídeo MPEG-4 (SOLER, 2006).

- Sessão de vídeo (*Video Session - VS*) - É a estrutura sintática que está no topo da hierarquia do fluxo codificado. Inclui objetos visuais 2D ou 3D, naturais ou sintéticos,

e as respectivas camadas de melhoramento. Este nível hierárquico indica quais os objetos visuais presentes na cena e qual o perfil e o nível dessa cena.

- Objeto de vídeo (*Video Object - VO*) - Um objeto visual é uma entidade presente na cena que o utilizador pode ter acesso e manipular. Pode ser um objeto visual (de origem natural ou sintética), uma textura estática, uma malha 2D ou 3D ou um objeto facial 3D. Habitualmente, um objeto visual corresponde a um objeto com significado semântico ou ao fundo de uma cena.
- Camadas de objetos de vídeo (*Video Object Layer - VOL*) - Este nível contém toda a informação referente a uma camada de um objeto de vídeo, nomeando os parâmetros de codificação e o fluxo binário, usadas para melhorar a resolução espacial e temporal de um objeto de vídeo.
- Plano de objeto de vídeo (*Video Object Plane - VOP*) - Uma instância de um objeto de vídeo num determinado instante de tempo denomina-se por VOP, exemplificado na figura 21. O processo de codificação gera representações codificadas de VOPs, juntamente com informação de composição que permite compô-los numa cena:
 - I-VOP – Um VOP pode ser codificado independentemente de qualquer outro VOP (*Intra VOP* ou I-VOP);
 - P-VOP – Um VOP pode ser codificado utilizando predição com base noutra VOP descodificado anteriormente. A estes VOPs chamam-se *Predicted VOP* ou P-VOP;
 - B-VOP – Um VOP pode ser codificado utilizando como predição VOPs passados e/ou futuros, em termos de apresentação. Estes VOPs chamam-se *Bidirectional VOPs* ou B-VOPs.
- Grupo de planos de objetos de vídeo (*Group of Video Object Planes - GOV*) – Os GOVs agrupam VOPs e são essenciais para o acesso aleatório à informação de vídeo.

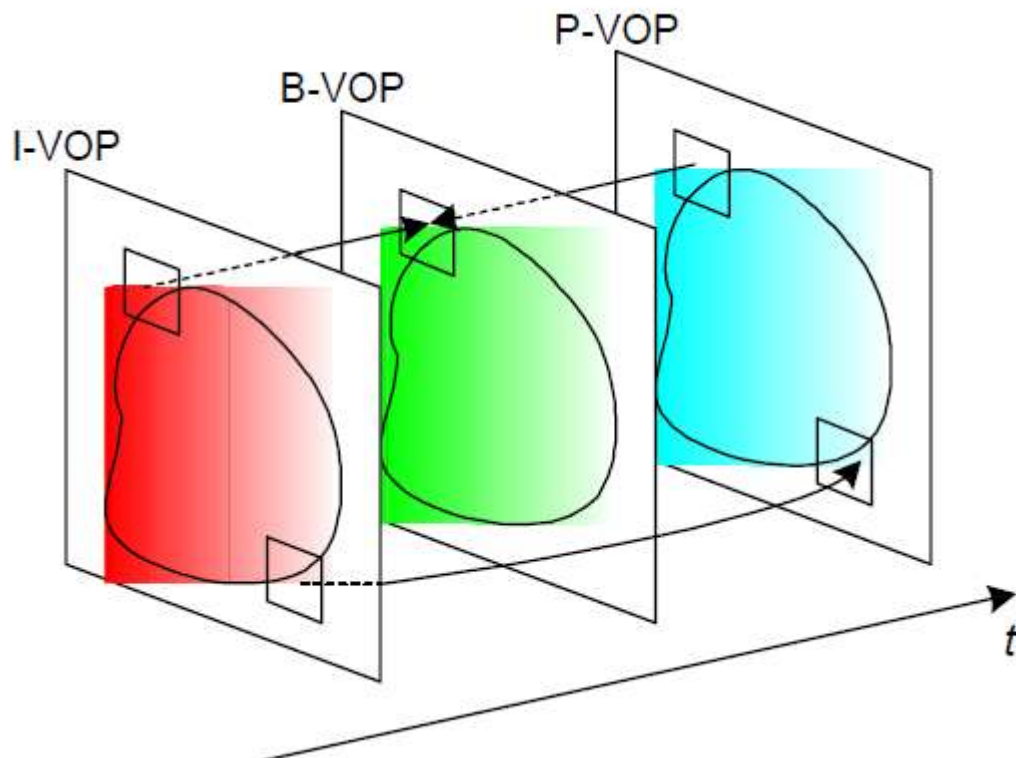


Figura 21- Modos de codificação de um VOP (VALENTIN, 2004).

3.4.3.2 Ferramentas de Codificação

O processo de codificação de vídeo especificado pelo padrão MPEG-4 possui a vantagem de que a eficiência de compressão pode ser aprimorada para algumas sequências de vídeo, dado que as ferramentas e técnicas de predição de movimento baseadas em objetos sejam adequadas às características do vídeo.

A sintaxe da codificação de vídeo na norma MPEG-4 inclui codificação escalável (várias camadas) e não escalável (uma única camada). Em seguida, encontram-se as principais técnicas de codificação de vídeo que podem ser usadas para garantir uma compressão eficiente e uma boa apresentação dos objetos de vídeo (MPEG, 2000):

- Quatro vetores de movimento por macrobloco - blocos de 8 x 8 para a componente de luminância e 4 x 4 pixels para a componente de crominância, com quatro vetores de movimento por macrobloco.

- Vetores de movimento com precisão de $\frac{1}{4}$ de pixel – as amostras dos VOP's de referência são interpoladas duas vezes. Uma descrição do movimento mais precisa implica em menor erro de predição, gerando um vídeo de melhor qualidade.
- Compensação de movimento global (*Global Motion Compensation - GMC*) para objetos de vídeo - Essa técnica é baseada na codificação da trajetória do movimento da câmera e codificação de texturas.
- VLC reversível – minimiza a área da imagem afetada por um erro.
- *Alternate Quant* – propõe uma nova forma de reescalonamento (quantização inversa) dos coeficientes da DCT, exceto os coeficientes DC.

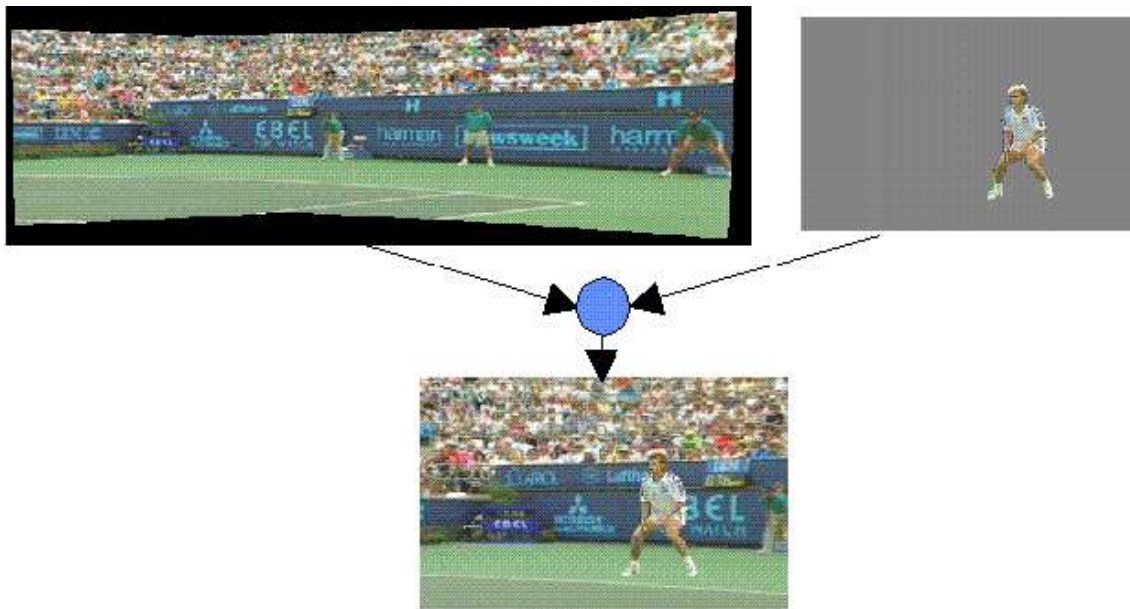


Figura 22 - Codificação MPEG-4 usando GMC (UFF, 2005).

A figura 22 ilustra o conceito básico de sequência de vídeo MPEG-4. É assumido que a imagem pode ser segmentada em plano de fundo e tenista. O plano de fundo define uma grande área apresentada por toda a extensão da sequência de vídeo. O plano de fundo pode ser maior do que o enquadramento da imagem e é transmitido uma única vez, ficando armazenado em uma memória especificamente selecionada no receptor.

Em cada quadro consecutivo são transmitidos apenas os parâmetros de movimentação da câmera. O tenista que é um objeto de vídeo com bastante movimentação, é transmitido

separadamente como um objeto de vídeo de forma arbitrária. Para reconstruir a sequência original, o receptor faz uma composição do plano de fundo com o tenista.

3.5 Padrão H.264

O H.264 é o resultado de um projeto conjunto com especialistas oriundos do VCEG (*Video Coding Experts Group*) da *International Telecommunications Union* (ITU) e o ISO/IEC *Moving Picture Experts Group* (MPEG). O ITU-T é o setor que coordena os padrões de telecomunicações em nome da *International Telecommunication Union*. ISO significa *International Organization for Standardization* (Organização Internacional para Padronização) e IEC é uma organização que supervisiona padrões eletrônicos e elétricos, bem como tecnologias relacionadas. O H.264 é o nome usado pela ITU-T, enquanto a ISO/IEC o denominou MPEG-4 parte 10/ AVC, pois ele é apresentado como uma nova parte em sua plataforma MPEG-4. A plataforma MPEG-4 inclui, por exemplo, o MPEG-4 parte 2, que é um padrão que tem sido utilizado pelas câmeras de rede e codificadores de vídeo baseados em IP (Axis Communications, 2008).

A intenção do projeto H.264 era criar um padrão capaz de fornecer boa qualidade de vídeo com uma taxa de bits muito baixa, em relação aos padrões já existentes, mas sem aumentar a complexidade do projeto. Outra meta do projeto era fazer um padrão que fosse compatível a todas as necessidades, isto é, compatível com vídeos de baixa e alta taxa de bits ou com baixa e alta resolução.

3.5.1 Principais Características

Desenvolvido para solucionar várias deficiências nos padrões de compactação de vídeo anteriores, o H.264 atende suas metas de suporte a:

- Implementações que fornecem uma redução média de taxa de bits de 50%, com base em uma qualidade vídeo fixa em comparação a qualquer outro padrão de vídeo.
- Robustez de erro, de modo que os erros de transmissão em várias redes sejam tolerados.
- Recursos de baixa latência e melhor qualidade para maior latência.
- Especificação clara de sintaxe que simplifica as implementações.

- Decodificação de correspondência exata, que define exatamente como os cálculos numéricos são feitos por um codificador e decodificador, a fim de evitar erros decorrentes de acúmulos.

Outra função desta norma é a integração com as normas MPEG já existentes. Dessa forma, o H.264 tira o melhor proveito das técnicas já existentes e apresenta melhorias para atingir uma codificação de melhor desempenho.

3.5.2 Perfis e níveis do H.264

O grupo conjunto participou da definição do H.264 com foco na criação de uma solução simples e clara, limitando as opções e os recursos a um mínimo. Um importante aspecto do padrão, como com outros padrões de vídeo, é fornecer os recursos em perfis (conjuntos de recursos algorítmicos) e níveis (classes de desempenho) que idealmente são compatíveis com produções populares e formatos comuns.

O H.264 tem vários perfis, cada um direcionado a uma classe específica de aplicações. Cada perfil define qual conjunto de recursos o codificador pode usar e limita a complexidade de implementação do decodificador.

As câmeras de rede e os codificadores de vídeo provavelmente utilizarão um perfil denominado linha de base, que se destina principalmente a aplicações com recursos limitados de computação. O perfil da linha de base é mais adequado de acordo com o desempenho disponível em um codificador em tempo real, que é incorporado em um produto de vídeo de rede (Axis Communication, 2008).

O padrão ITU-T H.264 define um conjunto de três perfis. Um perfil de codificação define os requisitos do codificador ou decodificador e especifica um conjunto de ferramentas de codificação de acordo com a finalidade do perfil. São eles (OLIVEIRA; CABRAL, 2007):

- *Baseline profile*: oferece suporte à codificação intra e inter (usando frames I e P respectivamente) e codificação de entropia com códigos de comprimento variável de contexto adaptativo (*Context-Adaptive Variable-Length Codes – CAVLC*). Aplicado em videotelefonia, videoconferência, comunicações sem fio.
- *Main profile*: suporta vídeo entrelaçado, codificação inter usando frames do tipo B, codificação inter usando perdição ponderada e codificação aritmétrica de contexto

adaptativo (*Context Based Arithmetic Coding* - CABAC). Aplicado em radiodifusão televisiva e armazenamento de vídeo.

- *Extended profile*: não suporta vídeo entrelaçado e CABAC, mas oferece suporte a métodos de permutação de diferentes fluxos de bits codificados e resistência e recuperação de erros aprimorada. Usado principalmente em aplicações de streaming de mídias.

3.5.3 Eficiência do H.264

O H.264 eleva a tecnologia de compactação de vídeo para um novo patamar. Com o H.264, um novo e avançado esquema de previsão intra é introduzido para codificação de quadros I. Esse esquema pode reduzir significativamente o tamanho de bits de um quadro I e manter uma alta qualidade ao permitir a previsão sucessiva de blocos menores de pixels em cada macrobloco de um quadro.

Isso é feito ao tentar localizar pixels de correspondência entre os pixels codificados anteriormente que limitam um novo bloco de pixels 4x4 para ser intracodificado. Ao reutilizar os valores de pixel que já foram codificados, o tamanho de bits pode ser drasticamente reduzido. A nova previsão intra é uma parte fundamental da tecnologia do H.264 que é comprovadamente muito eficiente.

Para fins de comparação, se somente os quadros I fossem usados em um fluxo de H.264, ele teria um tamanho de arquivo muito menor do que um fluxo de Motion JPEG, que usa somente quadros I.

A compensação de movimento com base em blocos – utilizada na codificação de quadros P e B – também foi aprimorada no H.264. Um codificador H.264 pode optar por pesquisar blocos de correspondência – abaixo para precisão de subpixel – em algumas ou muitas áreas de um ou vários quadros de referência.

O tamanho e o formato do bloco também podem ser ajustados para aprimorar uma correspondência. Em áreas em que não há blocos de correspondência em um quadro de referência, são utilizados macroblocos intracodificados. O alto grau de flexibilidade em compensação de movimento com base em bloco do H.264 é válido em cenas de vigilância com muitos elementos, em que a qualidade pode ser mantida para aplicações exigentes. A compensação do movimento é o aspecto mais exigente de um codificador de vídeo e as

diferentes formas e graus com os quais ele pode ser implementado por um codificador H.264 podem ter um impacto na eficiência de compactação de vídeo.

3.5.4 Melhorias do H.264

O padrão H.264 trouxe novidades na codificação de vídeo em relação aos padrões anteriores. Abaixo segue algumas melhorias, considerando estimativa e compensação de movimento (OLIVEIRA; CABRAL, 2007):

- Mais flexibilidade no tamanho do bloco para realizar compensação de movimento variando de 16x16 a 4x4.
- Possibilidade de uso de um quarto de pixel (*quarter sample*) para realizar a compensação de movimento, característica citada no padrão MPEG-4 Visual.
- Vetores de movimento podem ultrapassar os limites de um quadro, como mostrado na figura 23.
- Compensação de movimento usando múltiplos quadros de referência (para P e B).
- Desacoplamento da ordem de referência da ordem de apresentação.
- Possibilidade de uso de quadros B como referência para codificação de outros quadros.
- Predição de quadros de maneira ponderada pelo codificador (cenas com fades).
- Filtragem anti-bloco (*deblocking filtering*) – evita efeitos de visualização da imagem em blocos (*blocking artifacts*).

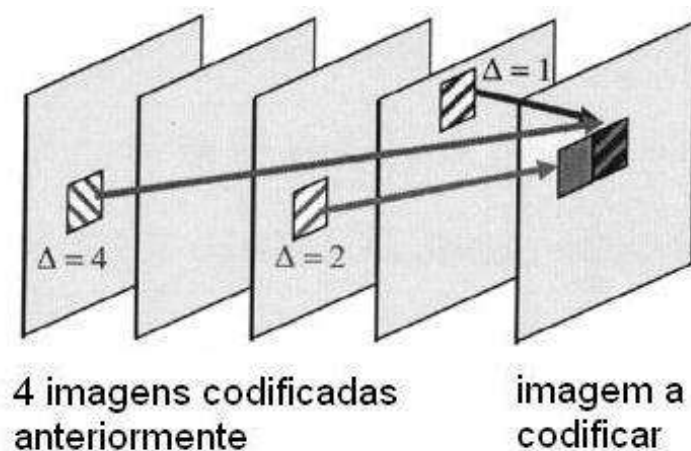


Figura 23 – Vetores de movimento utilizando múltiplos quadros (IMG.LX, 2007).

As principais novidades adicionais na codificação de vídeo em relação a padrões anteriores, que ajudam a aumentar a eficiência na codificação são listadas abaixo (J.F.F. de Oliveira e P.A. Cabral, 2007):

- Menor tamanho e particionamento de blocos para aplicação da transformada (8x8 ou 4x4).
- Transformada hierárquica.
- Transformada com palavras de tamanho pequeno – padrões anteriores usavam processamento de 32 bits, H.264/AVC usa aritmética de 16 bits.
- Transformada inversa exata, permitindo obter a mesma qualidade de decodificação em todos os decodificadores:
- Codificação por entropia usando codificação aritmética (CABAC – *Context-Adaptive Binary Arithmetic Coding*).
- Codificação por entropia adaptativa ao contexto: ambos os métodos CAVLC e CABAC utilizam codificação adaptativa ao contexto para melhorar o desempenho do codificador.

4 COMPARAÇÃO DE DESEMPENHO

Ao comparar o desempenho dos padrões MPEG – como o MPEG-4 e o H.264 –, é importante observar que os resultados podem variar entre codificadores que utilizam o mesmo padrão. Isso ocorre porque o criador de um codificador pode optar por conjuntos diferentes de ferramentas definidas por um padrão. Se a saída gerada por um codificador estiver de acordo com o formato e o decodificador de um padrão, é possível realizar diferentes implementações.

Portanto, um padrão MPEG não pode garantir uma determinada velocidade de transmissão ou qualidade, e não é possível realizar comparações corretas sem antes definir como os padrões são definidos em um codificador. Um decodificador, ao contrário de um codificador, deve possuir todas as partes obrigatórias de um padrão para decodificar um fluxo de bits compatível. Um padrão especifica exatamente como um algoritmo de descompactação deve restaurar cada bit de um vídeo compactado.

4.1 Teste realizado com uma câmera de segurança

O teste apresentado a seguir, foi realizado por uma empresa de segurança, já que o H.264 também deve acelerar a adoção de câmeras com altos valores de megapixels, pois a tecnologia de compactação extremamente eficiente pode reduzir os tamanhos de arquivos grandes e das taxas de bits geradas sem comprometer a qualidade da imagem. No entanto, há concessões. Ao mesmo tempo em que o H.264 possibilita economia nos custos de armazenamento e largura de banda de rede, ele exigirá câmeras de rede e estações de monitoramento de desempenho superior.

A figura 24 apresenta uma comparação da velocidade de transmissão necessária para uma mesma sequência de vídeo, gravada por uma câmera de segurança em uma portaria, realizadas por codificadores da *Axis Communication*⁴, entre os seguintes padrões de vídeo: *Motion JPEG*, MPEG-4 Visual (sem compensação de movimento), MPEG-4 Visual (com compensação de movimento) e H.264 (perfil básico).

⁴ A Axis Communications é uma empresa de TI que oferece soluções de vídeo em rede para instalações profissionais. A empresa é líder do segmento no mercado global e impulsiona a contínua migração do sistema analógico de vigilância em vídeo para o sistema digital. Seus produtos e soluções são baseados em plataformas tecnológicas novas e abertas e são focados em vigilância e monitoramento remoto (website da empresa).

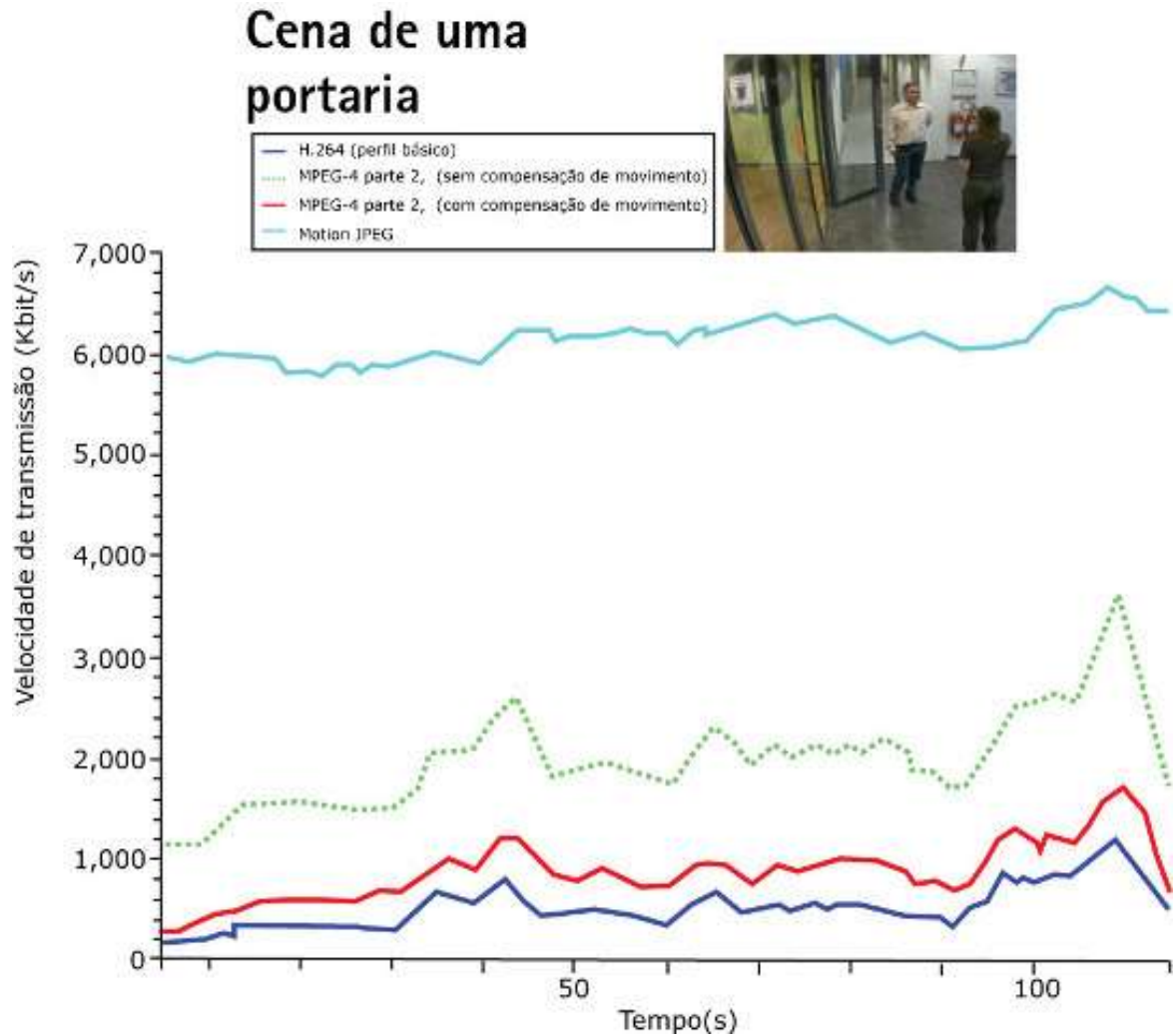


Figura 24 - Comparação entre H.264, MPEG-4 e Motion JPEG (Axis Communication , 2008).

Em uma seqüência de vídeo gravada pela mesma câmera de segurança de uma portaria, o codificador H.264 da *Axis Communication* gerou até 50% menos bits por segundo (velocidade de transmissão necessária) do que um codificador MPEG-4 com compensação de movimento. O codificador H.264 foi pelo menos três vezes mais eficiente do que um codificador MPEG-4 sem compensação de movimento, e pelo menos seis vezes mais eficiente do que com o Motion JPEG.

Com o H.264, os artefatos geralmente em bloco – vistos em vídeo extremamente compactado usando padrões Motion JPEG e MPEG que não sejam H.264 – podem ser reduzidos usando um filtro de eliminação de blocos em loop. Esse filtro suaviza as bordas de blocos usando uma potência adaptável para fornecer um vídeo descompactado quase que perfeito.

4.2 Estudo comparativo de economia de taxa de bits

Para demonstração da eficiência de codificação do H.264, foram realizados testes comparativos com outros importantes, e bem-sucedidos, padrões de codificação de vídeo. O teste demonstrado a seguir, é um estudo disponível na literatura (ASCENSO, 30 CAP4). Os codificadores envolvidos nos testes foram os seguintes:

- MPEG-2
- H.263
- MPEG-4 Visual, com precisão da estimação de movimento em $\frac{1}{4}$ de pixel.
- H.264 *Main Profile*, configurado para usar cinco quadros de referência.

Foram escolhidas sequências de testes especialmente usadas com esse intuito. Apenas a primeira imagem de cada sequência foi codificada como um quadro I, e foram inseridos dois quadros B entre dois quadros P sucessivos. Para o H.264, os quadros B não foram armazenados nas memórias de quadros, impossibilitando-as de serem usadas como quadros de referência. Foi usada a busca exaustiva no processo de estimação de movimento com um raio de busca de 32 bits com controle *lagrangeano* (SULLIVAN, 1998). A quantização manteve-se constante para toda a sequência de vídeo.

Tabela 1 - Comparação de economia de taxa de bits.

Codificador	MPEG-4	H.263	MPEG-2
H.264	38,62%	48,80%	64,46%
MPEG-4	-	16,65%	42,95%
H.263	-	-	30,61%

A partir da tabela 1, é possível destacar o poder do H.264 na evolução das normas de codificação de vídeo, pois apresenta um desempenho mais que duas vezes superior com

redução da taxa de bits maior que 50% para o mesmo nível de qualidade, em comparação com o popular MPEG-2. Mesmo quando comparado com o MPEG-4, que apresenta complexidade parecida, o H.264 obteve um economia de bits superior a um terço.

O desempenho superior do padrão H.264 está ligado às ferramentas de compensação do movimento e a codificação aritmética de contexto adaptativo, o que mostra a eficiência dos métodos de compressão de vídeo, quando comparada aos padrões anteriores.

Com este incremento de eficiência, novas aplicações e serviços podem ser desenvolvidos. O desenvolvimento tecnológico de novas memórias e processadores são quem dita o ritmo desta norma, já que apresenta uma complexidade no desenvolvimento dos codificadores e decodificadores que gera custo.

5 CONCLUSÃO

O H.264 é um padrão aberto e licenciado que é compatível com as técnicas mais eficientes de compactação de vídeo disponíveis atualmente. Sem comprometer a qualidade de imagem, um codificador H.264 pode reduzir o tamanho de um arquivo de vídeo digital em mais de 80% em comparação com o formato Motion JPEG e em até 50% mais do que o tradicional padrão MPEG-4 parte 2. Isso significa que muito menos largura de banda de rede e espaço de armazenamento são necessários para um arquivo de vídeo. Ou ainda, pode ser atingida uma qualidade de vídeo bem mais superior para determinada taxa de bits.

Nota-se que todos os métodos de compressão visam reduzir as redundâncias presentes na informação, e com isso diminuir de forma significativa o tamanho do arquivo que contém essas informações, aumentando a velocidade de transmissão e economizando espaço no armazenamento dos mesmos.

O poder do MPEG vem do fato de não ser um só formato de compressão, mas um conjunto de ferramentas de codificação padronizadas que podem ser combinadas de forma flexível para atender a diversas aplicações. A forma como a codificação é feita é incluída nos dados comprimidos, informando ao decodificador como reconhecer os dados.

É importante lembrar que o padrão MPEG especifica somente a sintaxe e semântica da taxa de transmissão e do processo de decodificação. Ele não especifica o processo de codificação. Muito da codificação é deixado a cargo dos projetistas de codificadores e isso cria espaço para melhorias das técnicas já criadas e para desenvolvimento de novas técnicas de codificação.

Logo, o padrão MPEG não é um padrão específico. É um conjunto de padrões adequados a diferentes situações, mas que se baseiam em princípios semelhantes. Essa padronização facilita sua difusão no mercado.

6 BIBLIOGRAFIA

ALENCAR, Marcelo Sampaio de. **Televisão digital**. São Paulo: Érica, 2007.

GOMES, Jonas; LUIZ, Velho. **Computação Gráfica: Imagem**. Rio de Janeiro, IMPA/SBM, 1994.

MELO, Hugo. **Tutorial: Compressão Digital de Vídeo**. São Paulo: Revista Engenharia de Televisão – Parte 1, 1999.

CASTRO, F. C. C. de; CASTRO, M. C. F. de. **Apostila: Codificação de Sinais**. PUCRS – Faculdade de Engenharia – Departamento de Engenharia Elétrica, 1999.

ANATEL. Disponível em <<http://www.anatel.gov.br>>. Acesso em 08 de Setembro de 2011.

VALENTIM, João Manuel Malveiro. **Modelo Alternativo de Complexidade para a Norma MPEG-4 Visual**. Dissertação para obtenção do Grau de Mestre em Engenharia Eletrotécnica e de Computadores – Universidade Técnica de Lisboa, 2002.

MANZATO, Marcelo Garcia. **Técnicas e Padrões de Codificação de Vídeo**. Trabalho de Conclusão de Curso – Departamento de Computação, Universidade Estadual de Londrina, 2004.

Padrão de compactação de vídeo H.264. Disponível em <<http://www.axis.com>>. Acesso em 20 de setembro de 2011.

O que é espaço de cores? Disponível em <<http://www.tecmundo.com.br/2481-o-que-e-espaco-de-cores.htm#ixzz1fC5IYWRR>>. Acesso em 10 de outubro de 2011.

Digitalização de um sinal analógico. Disponível em <<http://www.qsl.net/py4zbx/teoria/digitaliz.htm>>. Acesso em 23 de agosto de 2011.

INTERNATIONAL TELECOMMUNICATIONS UNION. *ITU-T Recommendation H.264 : Advanced video coding for generic audiovisual services*. Maio 2003.