

# Fórmulas

- Teorema de Nyquist: Taxa de amostragem deve ser ao menos o dobro da frequência do som.
  - Exemplo: frequência 4KHz -> taxa de amostragem 8KHz
- Taxa de bits: produto entre taxa de amostragem e número de bits.
  - Exemplo: Taxa de amostragem 8KHz, 8 bits por amostra ->  $8000 \times 8 = 64000$  (64Kbps)
- Se  $n$  é a resolução de cor então a quantidade de níveis possível é  $2^n$  níveis.
  - Exemplo: 5 bits por pixel ->  $2^5 = 32$  níveis por componente.
- Requisito de armazenamento de imagens:  $HVP/8$  (bytes)
  - $H$  = número de pixels por linha,  $V$  = número de linhas,  $P$  = bits por pixel.
  - Exemplo: 420 pixels/linha (largura), 512 linhas (altura), 24 bits =  $(420 \times 512 \times 24)/8 = 645120 = 645,12\text{Kbytes}$
- Tamanho em disco:
  - Arquivo ocupa (tamanho do arquivo)/(tamanho do cluster) clusters. Arredonda para cima.
  - Tamanho em disco: (número de clusters) \* (tamanho do cluster)
  - Exemplo: Arquivo de 645174B, clusters de 4096B ->  $645174/4096 = 157,51 = 158$  clusters. Tamanho em disco =  $4096 \times 158 = 647168\text{B}$ .
- Taxa de bits para imagem: calculada a partir dos requisitos de armazenamento e tempo de transferência.
  - $R = HVP/t$  ( $t$  = tempo de transferência)
  - Exemplo: Imagem 420x512, 24 bits, deve ser transmitida em 2s ->  $(420 \times 512 \times 24)/2 = 2580480 = 2,58 \text{ Mbps}$
- Taxa de bits para áudio: número de canais \* taxa de amostragem \* bits por amostra
  - Exemplo: 2 canais, 44100Hz, 16 bits por amostra ->  $2 \times 44100 \times 16 = 1411200 = 1411,2 \text{ Kbps}$

- Espaço ocupado por áudios = número de canais \* taxa de amostragem \* bits por amostra \* duração / 8
  - Exemplo: 1 canal, 8000Hz, 8 bits, 60 segundos - >  $1 * 8000 * 8 * 60 / 8 = 480000 = 480\text{KBytes}$
- Taxa de bits para vídeo = HVP \* fps
  - Exemplo: 30 fps e imagens 720x480 com 24 bits/pixel de 1 minuto ->  $720 * 480 * 24 * 30 = 248.832.000 = 249 \text{ Mbps}$
- Espaço ocupado por vídeos = (HVP/8) \* fps \* duração
  - Exemplo: 30 fps e imagens 720x480 com 24 bits/pixel de 1 minuto ->  $249 * 60 / 8 = 1,87 \text{ GB}$
- Dado um alfabeto com s símbolos, são necessários  $n = \log_2 s$  bits para codifica-los. Apenas se a distribuição de probabilidades for desconhecida ou uniforme.
- Entropia: é definida como:
  - Onde  $p_i$  é a probabilidade de ocorrência do símbolo i do alfabeto.
- Taxa de compressão: tamanho original / tamanho após compressão
- 
- 
- B: número de bits por símbolo
- Tamanho do quadro de voz (payload): Taxa do codec (bit/sec) \* tempo do quadro de voz (sec) / 8 (para bytes)
  - Exemplo: 64kbps, pacote 20ms ->  $(64000 * 0,02) / 8 = 160 \text{ bytes}$
- Pacotes por segundo:  $1000 (1 \text{ segundo em ms}) / \text{tempo do pacote (ms)}$
- Taxa de bits: numero pacotes \* tamanho do pacote (bytes) \* 8
-

# Capítulo 1

- Quais os tipos de mídia? (Estática, dinâmica)
  - Estática (discreta): dimensões unicamente espaciais, não altera com o tempo.
  - Dinâmica (contínua): dimensão temporal, significado depende da taxa.
  - Imersiva: informação interativa em ambiente 3D.
  - Capturadas: capturadas do meio exterior.
  - Sintetizadas: produzidas em computador.
  - Linear: Não há controle no desenrolar do processo.
  - Não-linear: Oferece interatividade.
  - Online e offline.
- Qual a definição de multimídia? (sem consenso)
  - Não há consenso, pois, é um termo utilizado para marketing, mas a definição etimológica é “múltiplos meios de portar a informação”.
  - Sistema que apresenta imagens, sons e texto. Vários meios de comunicação.
  - **Combinação controlada por computador de pelo menos um tipo de mídia estática com pelo menos um tipo de mídia dinâmica.**
- Quais os desafios da multimídia? (redes e sincronismo)
  - Sistemas convencionais tem mais suporte para dados alfanuméricos do que multimídia.
  - Devem ser apresentados em taxas fixas e devem satisfazer requisitos tempo real.
  - Relações temporais e espaciais devem ser mantidas (sincronização).
  - Banco de dados convencionais não possuem suporte efetivo

# Capítulo 2

- Quais são os componentes de frequência? (tons, sons)
  - Frequência: número de períodos em um segundo.
  - Tons: sons com frequência única.
  - Sinal simples: não pode ser decomposto em tons.
  - Sinal composto: soma de sinais periódicos de múltiplas ondas senoidais.
- Quais as características da faixa audível humana? (frequência, capacidade, limiar)
  - Faixa audível humana: 20Hz a 20000Hz.
  - Humanos não tem a mesma capacidade auditiva independente da frequência.
  - Limiar de audição: relação entre frequência e intensidade do som para ser audível ou doloroso.
- Quais as etapas do processo de digitalização de um som?
  - Amostragem: conjunto discreto de valores analógicos é amostrado em intervalos temporais em periodicidade constante.
    - § Quanto maior a taxa de amostragem, melhor é a qualidade.
    - § Se a taxa de amostragem é 1.5 a taxa do sinal, vai produzir frequência alias (pseudonímia).
  - Quantização: O sinal amostrado é quantificado. Técnica que utiliza o mesmo passo de quantização se chama PCM.
    - § Ruído de quantização: ocorre quando o número de bits é insuficiente.
  - Codificação: Conjunto de bits chamado de code-word é associado com cada valor quantificado.
- Quais são os parâmetros importantes para a digitalização?
  - Taxa de amostragem: quantidade de amostras coletadas em um certo período.
    - § Teorema de Nyquist: Taxa de amostragem deve ser ao menos o dobro da frequência do som.

§ Se o sinal tiver componentes de frequência maiores que a frequência de Nyquist, eles são convertidos a frequências mais baixas e geram ruído denominado de Pseudonímia.

- Taxa de bits: produto entre taxa de amostragem e número de bits.

· Quais as diferentes formas de quantização e vantagens?

- Quantização linear: Tamanho do passo de quantização é constante, ou seja, toda a amostra tem o mesmo número de bits.

§ Resulta em qualidade mais elevada na região de mais alta amplitude. Não aumenta qualidade percebida.

- Quantização não linear: tamanho do passo de quantização aumenta logaritmicamente com a amplitude do sinal.

§ Sinal é transformado de linear para não linear. Quantização uniforme é aplicada em sinal transformado. Companding PCM.

· Quais as características luminosas das imagens? (intensidade, comprimento, espectro)

- Três grandezas físicas: intensidade, frequência (comprimento de onda), polarização (direcionamento).
- Espectro visível: 400-700nm.

· Como funciona o sistema visual humano? (cones e bastonetes, RGB)

- Bastonetes: medem a intensidade da luz (luminosidade). Maior quantidade, mais baixa definição.
- Cones: Medem a frequência da luz (cores). Requerem maior luminosidade para serem ativados.
- Tipos de cones: curtos (azul), médios (verde), longos (vermelho).
- Proporção entre longos, médios e curtos é 40:20:1.

· Quais são os sistemas de representação de cores? (aditivo, subtrativo)

- Sistemas de síntese aditiva: cor é percebida diretamente a partir da fonte luminosa. Adotado por dispositivos de emissão de luz.

§ Qualquer cor pode ser reproduzida com uma mistura de RGB.

- Sistemas de síntese subtrativa: cor é percebida a partir do reflexo da luz sobre uma superfície. Adotado por dispositivos de impressão.

§ Usam as cores secundárias: CMY e preto.

- Sistema HLS: Utiliza propriedades mais relevantes do ponto de vista da percepção humana. Luminância (amplitude/intensidade), Tonalidade (frequência), Saturação (grau de pureza).
- Sistema CIE 1931 XYZ: Cor é definida por valores XYZ. Y é a luminância, X e Z a cor. Interessante para a compressão.

· Como funciona a captura analógica de imagens e vídeos?

- Para imagens monocromáticas: lentes da câmera focam imagem em uma superfície fotossensível de sensores CCD, brilho de cada ponto é convertido para uma carga elétrica, superfície fotossensível é rastreada para capturar as cargas elétricas.

§ Para vídeos, apenas um sinal de luminância é produzido. São usadas câmeras de luminância.

- Teoria Tristimulus: câmera divide imagem em componentes RGB. Convertidos em sinais elétricos separados.
- Sinal RGB: sinal separado pelas cores básicas.
- Sinal de vídeo composto colorido: sinais RGB são codificados em um único sinal seguindo um determinado padrão (NTSC, PAL-M, etc).
- Sinal de croma: sinal composto por luminância e croma.
- Sinal YCbCr: um sinal de luminância e dois de croma.

· Quais os tipos de câmera analógicas?

- Câmera de croma: são usados 3 CCDs com filtros separados RGB para cada. 1 passo de captura. Vídeo composto colorido, S-vídeo ou sinal RGB.

§ 1 passo – 3 CCDs: Utilizada em aplicações profissionais. Custo elevado.

§ 1 passo – 1 CCD: Usado um único CCD com filtros R, G ou B em cada célula. Qualidade menor.

§ 3 passos – 1 CCD: uma digitalização para cada filtro RGB. Imagens devem ser estáticas, tem boa qualidade, telescópio.

· Quais os conceitos base da captura de imagens digitais? (bitmap, pixel)

- Bitmap: matriz espacial bidimensional de pixels.

- Pixel: menor elemento da resolução da imagem. Valor numérico de amplitude. Expresso por um número de bits. Preto e branco, nível de cinza ou atributo de cor (3 valores).
  - Resolução: Número de pixels que a imagem possui na vertical e horizontal.
  - Captura digital semelhante a analógica: imagem é digitalizada por CCD e armazenada de forma compactada ou não na memória. Qualidade depende da resolução do CCD e compressão. Podem ser usados sensores CMOS em vez de CCDs.
- Como funciona a digitalização de imagens? (amostragem, PCM, bits por pixel)
- Amostragem: é espacial (áudio é temporal).
  - Quantificação: também pode produzir ruído de quantificação.
  - Codificação: representação digital da luz/cor (RGB).
  - CCD captura 50% de verde, 25% de vermelho e azul. Percepção humana da retina usa cones M e L combinados durante a luz do dia, que é mais responsivo à luz verde.
- Quais os tipos de imagem? (binárias, tons de cinza, diferentes coloridas)
- Binárias: dois níveis (preto ou branco). É necessário apenas um bit por pixel (geralmente 24 em RGB). Cor definida na paleta. Usada para impressão.
  - Tons de cinza: cada pixel define uma intensidade de luminosidade representada em um certo número de bits. Resolução 8 bits representa até 256 cinzas. Padrões mais usados são de 4 e 8 bits por pixel.
  - Coloridas: Tipos dependendo do propósito e dispositivo.
    - § True color: cada pixel é representado por 3 componentes (RGB) com um certo número de bits. Geralmente número de bits é igual. Se  $n$  é a resolução de cor então a quantidade de níveis possível é  $2^n$  níveis.
    - § Cores indexadas: cada pixel é representado por um índice que aponta para uma tabela de cores (paleta). Geralmente 24 bits para representar cada cor em formato RGB. Paleta variável. Número de cores e resolução de cores variável.
    - § Cores fixas: cada pixel é representado por um índice que aponta para uma tabela de cores fixas.

- Como é caracterizado um vídeo digital? (sequência, parâmetros)
  - Vídeos digitais são imagens apresentadas em ordem sucessiva.
  - Quadro (frame): uma imagem individual em uma animação.
  - É definida uma resolução para todas as imagens.
  - Taxa de bits: define a qualidade final do vídeo.
  - Imagens animadas: vídeos. Capturados na vida real com câmeras ou criadas em computador.
  - Gráficos animados: apresentação sucessiva de objetos visuais gerados pelo computador.
- O que é taxa de quadros?
  - Taxa de quadros é a quantidade de quadros apresentada por segundo.
  - Movimentos naturais a partir de 16 fps. “Máximo” em 60.
- O que são mapas de caracteres?
  - Conjuntos de caracteres: tabelas mantidas pelo sistema operacional que consistem em uma correspondência entre os códigos e os caracteres.
  - Contém representações de grafemas (unidades fundamentais de um sistema de escrita) ou unidades similares a grafemas.
  - Tornam o texto revisável, permitem buscas, facilitam comparação.
  - Face: família de caracteres com muitos estilos e tipos (Arial).
  - Fonte: conjunto de caracteres com único tamanho, estilo e face (Arial 15 pontos itálico.)
- Quais são os padrões de codificação? (ASCII, ISO8859, Unicode)
  - ASCII: Primeiro conjunto normalizado. Usa 7 bits para representar 128 caracteres. Adequado a inglês, mas não para muitas línguas.
  - ISO8859: conjuntos de caracteres de 8 bits para cada parte. 10 partes representam sistemas de escrita diferentes. 8 bits ainda insuficientes. Não trabalha com várias línguas simultaneamente.
  - Unicode: Definido por consórcio de empresas. Repertório de mais de 128000 caracteres cobrindo 100 scripts (coleção de símbolos). Codifica caracteres em espaço numérico entre 0 e 10FFFF dividido em 17 planos



(0 a 16). U+aaa. Alguns formatos de codificação como UTF-8: um a quatro bytes por caractere.

- Como calcular taxa de bits e requisitos de armazenamento?
  - Requisito de armazenamento: será adotado o sistema internacional (potências de 10).
  - Requisito de taxa de bits: será adotado o SI (potências de 10).
  - Requisito de armazenamento de imagens: HVP/8
    - §  $H = \text{número de pixels por linha}$ ,  $V = \text{número de linhas}$ ,  $P = \text{bits por pixel}$ .
    - § Tamanho total + cabeçalho (54 bits para BMP)
    - § Tamanho em disco: unidades de alocação necessárias.
  - Tamanho em disco:
    - § Arquivo ocupa  $(\text{tamanho do arquivo}) / (\text{tamanho do cluster})$  clusters. Arredonda para cima.
    - § Tamanho em disco:  $(\text{número de clusters}) * (\text{tamanho do cluster})$
  - Taxa de bits para imagem: calculada a partir dos requisitos de armazenamento e tempo de transferência.
    - §  $R = HVP/t$  ( $t = \text{tempo de transferência}$ )
  - Taxa de bits para áudio:  $\text{número de canais} * \text{taxa de amostragem} * \text{bits por amostra}$
  - Espaço ocupado por áudios =  $\text{número de canais} * \text{taxa de amostragem} * \text{bits por amostra} * \text{duração} / 8$
  - Taxa de bits para vídeo =  $HVP * \text{fps}$
  - Espaço ocupado por vídeos =  $(HVP/8) * \text{fps} * \text{duração}$
- Quais os tipos de sincronização?
  - Relações espaciais e temporais entre mídias definem a sincronização multimídia: apresentação temporalmente correta de dados multimídia.
  - Sincronização intramídia: elementos da mesma mídia apresentados em instantes corretos (áudio, vídeo).

- Sincronização intermídia: relacionamentos temporais corretos entre os dados multimídia de uma aplicação deve ser mantidos (áudio e vídeo).
- Sincronização labial: sincronização entre movimento e voz.
- Sincronização de interação: evento de interação deve produzir efeito desejado dentro de um tempo relativamente curto.
- Quais são as restrições de atraso?
  - Atrasos fim-a-fim: soma de todos os atrasos em todos os componentes de um sistema multimídia.
  - § Conversações ao vivo: 400ms é limite de percepção.
  - Variação de atraso (jitter): mídias contínuas são transmitidas em pacotes que sofrem diferentes atrasos fim-a-fim.
- Como é a tolerância a perdas de informação?
  - Percepção humana tolera alguma perda de informação. Uso de técnicas de recobrimento de erros.

## Capítulo 3

- O que é entropia? (Teorema da codificação da fonte)
  - Dado um alfabeto com  $s$  símbolos, são necessários  $n = \log_2 s$  bits para codificá-los.
  - Apenas se a distribuição de probabilidades for desconhecida ou uniforme.
  - Entropia: é definida como:
  - Onde  $p_i$  é a probabilidade de ocorrência do símbolo  $i$  do alfabeto.
- Quais são os princípios da compressão? (redundância, percepção)
  - Redundância: amostras vizinhas geralmente tem similaridades. Pode se compactar guardando erros em relação a uma amostra base.
  - Percepção: técnicas de compressão pode remover informações que não seriam percebidas, como frequências fora do limiar de audição.
- Quais os três tipos de compressão?

- Sem perda: codificação por entropia, dado original pode ser reconstruído. Não leva em consideração o significado dos símbolos. Baixas taxas de compressão.
  - Com perda: codificação na origem. Utiliza propriedades da percepção humana, leva em conta a semântica. Altas taxas de compressão.
  - Codificações híbridas: combinam técnicas com e sem perda.
- Quais os parâmetros para medir o desempenho das técnicas de compressão?
- Taxa de compressão: tamanho original / tamanho após compressão. Sem perdas.
  - SNR: razão entre sinal e ruído.
  - 
  -
- § B: número de bits por símbolo
- § Se não há perdas, PSNR é infinito.
- Compressão e descompressão acontecem em tempo real, ocorre atraso. Velocidade de compressão importante para aplicações tempo real.
  - Descompressão importante para streaming ou não tempo real.
- Quais os princípios gerais das técnicas RLE e codificação de Huffman?
- Codificação RLE: codificação por entropia. Supressão de sequências de mesmos símbolos. Número de repetições e símbolo repetido.
  - Codificação Huffman: atribui menos bits a símbolos que aparecem mais frequentemente e mais para os que aparecem menos. Codificação custosa, decodificação simples.
    - § A) Colocar os símbolos em uma linha de probabilidade acumulada (aumenta de baixo para cima).
    - § B) Junta dois símbolos com menor probabilidade em um nó para formar dois ramos na árvore.
    - § C) Nova árvore tratada como símbolo único com probabilidade igual a soma dos ramos.
    - § D) Repetir B e C até todos os símbolos inseridos na árvore. Último nó é raiz.

§ E) Partindo da raiz, atribuir bit 0 no ramo de maior prioridade e bit 1 no ramo de menor de cada nó.

§ F) Código de cada símbolo é obtido percorrendo a árvore a partir da raiz e concatenando os bits.

- Quais os princípios gerais de (A)DPCM e LZ\*?
  - DPCM: guarda erros a partir de uma amostra de base.
  - ADPCM: varia o passo de quantização de acordo com tamanho do erro.
  - LZ: armazena repetições de símbolos em um dicionário e substitui pelo índice.
- Como comparar GIF e PNG?
  - GIF: utiliza técnica de compressão LZW. 256 cores sem perda. Em geral taxas 4:1. Extensão permite transparência, entrelaçamento e animação.
  - PNG: múltiplos níveis de transparência, melhor entrelaçamento, 48-bit truecolor ou 16bit cinza. Não suporta animação. LZ77 e Huffman.
- Quais os princípios das técnicas gerais de compressão de áudio, imagens e vídeos?
  - PCM: mesmo passo de quantificação.
  - PCM não linear: passo aumenta com o aumento da amplitude do sinal.
  - DPCM: função de predição baseada em erros.
  - LPC: amostra de áudio prevista com base nas anteriores.
  - ADPCM: variar o tamanho de passo representado pelos erros.
  - Imagens digitais puras por PCM (ou DPCM ou ADPCM).
  - Vídeos: explora redundância entre os quadros.
  - Redução da resolução de um bitmap.
  - Truncagem: eliminação dos bits mais significativos de cada pixel e imagens por segundo.
  - Predição: base + erro.
  - Preenchimento condicional: imagem é segmentada em áreas que são movidas, movimento é registrado como erros DPCM.

- Estimativa e compensação de movimento: imagem dividida em blocos de tamanho fixo, procura blocos similares na imagem anterior, calcula diferença.

· Quais são as etapas do algoritmo de compressão JPEG?

- Compressão é parametrizável: especifica o comportamento do codificador.
  - Codificação sequencial: maior suporte, com perdas baseada em DCT, única varredura esquerda-direita cima-baixo.
  - Codificação progressiva: com perdas baseada em DCT expandido, varreduras sucessivas, arquivos internet exibe resolução menor enquanto carrega.
  - Codificação hierárquica: codificação progressiva que aumenta resolução em estágios. Versões podem ser acessadas antes da imagem inteira ser descompactada. Taxa de compressão menor.
1. Transformar RGB em YCbCr
  2. Redução da resolução das matrizes YCbCr
  3. YCbCr decomposta em blocos 8x8, transformados em frequências usando DCT
  4. Quantificação: alta frequência descartada de acordo com taxa de compressão
  5. Coeficientes DCT são ordenados em sequência zig-zag.
  6. Codificação por entropia (Huffman ou aritmética).

· O que são pacotes de voz e quais suas características?

- Cada pacote de voz tem uma duração de 1 a 30ms.
- Fluxo de áudio acumulado até atingir o tamanho do bloco.
- Blocos maiores = maior atraso fim-a-fim.
- Tamanho do quadro de voz (payload): Taxa do codec (bit/sec) \* tempo do quadro de voz (sec)
- Pacotes por segundo:  $1000 \text{ (1 segundo em ms)} / \text{tempo do pacote (ms)}$
- Taxa de bits:  $\text{numero pacotes} * \text{tamanho do pacote (bytes)} * 8$
- Se o quadro é pequeno, a sobrecarga de protocolos é maior.

- Se o pacote é maior, a perda gera mais silêncio, e gera mais atraso.
- Quais as vantagens e limitações da supressão de silêncio?
  - VAD: detecta quando o usuário está conversando e em silêncio.
    - § Economiza energia de dispositivos a bateria.
    - § Deve ser bastante sensível, mas não tanto por causa de ruído de fundo.
  - DTX: codec para transmitir quadros quando o VAD detecta silêncio.
    - § Libera a banda dinamicamente.
  - Quando a fala é frequente, a supressão não tem ganhos.
  - Como a detecção não é imediata, início da frase pode ser perdido.
  - Se o ruído for muito alto, pode ser empacotado.
- Quais os princípios gerais da compressão MPEG áudio?
  - Padrão até 20KHz, explora a percepção humana.
  - Filtra sons acima disso, e também abaixo do limiar de audição.
  - Mascaramento: descarta um som quando é mascarado por outro (intenso).
  - Mapeamento tempo-frequência: divide a entrada em sub-bandas de frequências múltiplas.
  - Modelo psico-acústico: cria conjunto de dados para controlar quantificador e codificador.
  - Quantificador e codificador: cria conjunto de símbolos. Sub-bandas menos importantes são descartadas.
  - Empacotamento de quadros: monta e formata os símbolos de código com outras informações.
- Quais os princípios da compressão MPEG-vídeo? (redundância espacial, temporal)
  - Quadros representados em YCbCr
  - Redundância espacial: codificação de cada quadro similar a JPEG.
  - Tipos de quadro:

§ I: imagens estáticas independentes, JPEG.

§ P: diferença bloco a bloco com I ou P anterior.

§ B: diferença com quadro anterior e próximo.

§ D: usado para fast forward.

- Quadros P e B exploram redundância temporal. Compensação de movimento por macroblocos (16x16).

· Quais as características do formato MPEG-4 e H.264\*?

- MPEG-4: padrão de compressão de vídeo DCT. 21 perfis.
- ITU-T H.264: criado para telefonia e videoconferência. Compressão de vídeo em tempo real com atraso mínimo. Quadros I e P.