

Aluno: Lucas Mikilita Ribeiro
#USP: 5908703

Listas de Multimídia

LISTA 1

1. Dê a sua própria definição de multimídia para a área computacional, com base no que foi discutido em aula.

O termo multimídia se refere à apresentação ou recuperação de informações que se faz, com o auxílio do computador, de maneira multissensorial, integrada, intuitiva e interativa, envolvendo diferentes tipos de mídia.

2. Você classificaria um dispositivo tal como o iPod Classic como um sistema multimídia? Se sim, liste os diferentes tipos de mídia que ele suporta. Para cada mídia listada, classifique-a de acordo com a finalidade (percepção, representação, apresentação, armazenamento e transmissão) e dependência do tempo (discreta ou contínua).

Sim, o iPod Classic pode ser considerado um sistema multimídia, pois reproduz som e imagens. Ambas são mídias de percepção. Os padrões mp3 e jpg, usados para codificar som e imagens, são mídias de representação. A tela de LCD do iPod e os fones de ouvido são mídias de apresentação. A memória interna do iPod é a mídia de armazenamento. O ar e o cabo do fone de ouvido são as mídias de transmissão.

A imagem do iPod é uma mídia discreta e o som reproduzido é uma mídia contínua.

3. Aponte qual a motivação em comprimir dados digitais.

Apesar da afirmação "Compressão vai se tornar redundante em breve, conforme as capacidades de armazenamento e transmissão aumentem." ter sido repetida nos últimos 20 anos, os meios de transmissão e armazenamento continuam limitados, além das mídias terem se sofisticado com o passar dos anos. A compressão ajuda a diminuir a demanda por espaço e banda.

4. É aconselhável realizar compressão lossy em textos? Por quê?

Não, pois a perda de informação em textos impossibilita o seu entendimento. Essa compressão torna-se irreversível.

5. Seja um determinado tipo de informação que é usualmente codificado por longas strings que alternam zeros e uns, como, por exemplo, 0101101010100101010101110101. É possível adaptar o algoritmo de codificação por carreira (run-length) para obter boas taxas de compressão para esse tipo de

informação? Explique e mostre como seria o resultado da compressão para a string dada como exemplo. (Dica: pense em substrings de tamanho maior que 1)

Sim, é possível. Adotando uma substring de tamanho 2 (01, 10 ou 11), teremos:

01,2 - 10,4 - 01,6 - 11,1 - 10,2

6. Dê exemplo de uma situação em que a codificação por diferenças pode ocasionar perdas.

Se o número de bits usado não for adequado para o tamanho máximo das diferenças, ocorrerá perda de informação. Por exemplo, usando 8 bits para as diferenças e tendo um nível com mais de 256 unidades de diferença para o próximo nível, a codificação não conseguirá reproduzir essa diferença, sendo necessário aumentar o número de bits.

7. O que significa dizer que a árvore de Huffman tem a propriedade do prefixo? Qual é a vantagem dessa propriedade?

É o mesmo que dizer que nenhum código é prefixo de outro código. Essa propriedade facilita a codificação/decodificação, pois basta substituir os caracteres por seu código ao codificar, e percorrer a árvore de Huffman com o código para decodificar.

8. Sabe-se que uma das vantagens da codificação aritmética é que ela sempre atinge o valor da entropia. Cite uma de suas principais desvantagens, intrinsicamente relacionada com o hardware em que ela é executada.

Como o número de dígitos do código cresce linearmente com a entropia, o número máximo de caracteres de uma string é determinado pela precisão de ponto flutuante na máquina destino.

9. Decodifique a string representada pelo código 0,44220. Dado o alfabeto formado por A, C,

O e S cujas probabilidades são:

$P(A) = 0,4$

$P(C) = 0,3$

$P(O) = 0,1$

$P(S) = 0,1$

$P(.) = 0,1$

R: CAASO

10. Explique como uma palavra é decodificada utilizando o algoritmo LZW.

As strings são codificadas e armazenadas em um dicionário, que é construído dinamicamente. Inicialmente os dicionários contêm apenas a tabela ASCII. A primeira palavra é codificada usando os índices dos caracteres que a compõe, Em seguida, A palavra é armazenada no dicionário. Na próxima ocorrência o codificador envia seu código e o decodificador compara com o dicionário.

Lista 2

1. A digitalização de áudio se baseia em transformar um sinal de áudio analógico, com valores expressos em Volts, para um sinal de áudio digital, com valores expressos em bits. Isto é necessário para que o computador possa operar sobre o sinal de áudio. A digitalização pode ser dividida em duas etapas: a amostragem e a quantização. Explique o que ocorre com o sinal de áudio em cada um delas.

Amostragem: ocorrem leituras periódicas e instantâneas da voltagem do sinal em espaços de tempo uniformes. Este intervalo de amostragem é calculado levando-se em conta o teorema de Nyquist, que exige que a frequência de amostragem seja no mínimo o dobro da frequência do sinal amostrado.

Quantização: os valores analógicos obtidos na amostragem são convertidos em valores digitais. Quanto mais bits forem utilizados para quantizar o sinal amostrado, mais fiel ele será ao sinal original.

2. Para que fosse possível fazer o processo inverso da digitalização, isto é, através do sinal digital reconstruir o analógico, seriam necessárias amostras infinitas do sinal de modo que não houvesse distorções que causassem perda da qualidade. Felizmente, há um teorema que dita qual a taxa de amostragem mínima para que um sinal não sofra distorções ao ser reconstruído.

a. Qual é este teorema e qual a taxa de amostragem mínima a ser adotada?

Nyquist, deve ser o dobro da frequência do sinal amostrado.

b. Qual o fenômeno que ocorre caso não seja obedecido o teorema? Explique.

Ocorre perda de informação e o sinal amostrado não será inteiramente codificado, pois a convolução não será completa.

3. Diferente do que acontece na amostragem, a quantização do sinal de áudio não possui nenhum teorema.

a. Neste caso, qual a abordagem a ser seguida para obter um bom número de bits a ser usado por amostra?

Deve-se escolher uma quantidade de bits que não permita grandes distorções, as mais comuns são 4, 8, 16 e 24 bits.

b. Explique o que acontece quando se utilizam poucos bits por amostra.

Quando se utiliza poucos bits na quantização, ocorre uma distorção do som, pois há menos informação e portanto ele não será fiel ao sinal original.

4. Podemos considerar o PCM (Pulse Code Modulation) como um padrão para codificação de áudio. Uma versão melhorada desse padrão, o ADPCM (Adaptive Differential Pulse Code Modulation), possibilita a codificação do áudio com boa compressão utilizando-se a codificação por diferenças modificada para fazer lookahead. Explique como essa codificação é feita e o porquê da utilização de um lookahead.

Como as amostras de áudio adjacentes costumam ser parecidas, o ADPCM codifica a amostra inicial e depois passa a codificar apenas a diferença entre as amostras, de modo a tentar economizar informação a ser armazenada. O lookahead é necessário para se adaptar a escala de diferenças de acordo com o tamanho da mudança entre as amostras.

5. O MP3 (MPEG-1 Audio Layer 3) é um dos padrões mais conhecido e utilizado para codificação de áudio. A boa taxa de compressão atingida envolve a mistura de diferentes técnicas de compressão lossy e lossless, bem como explora características do modelo psico-acústico humano, sendo elas: a sensibilidade do ouvido, o mascaramento de frequências e o mascaramento temporal.

a. Explique cada uma das características do modelo psico-acústico humano.

A Sensibilidade do ouvido humano varia entre 20Hz a 20KHz, sendo mais sensível de 2KHz a 4KHz. Após ouvir um som alto há um delay até que o ouvido possa perceber um som mais baixo. Esse delay é de aproximadamente 50ms.

b. Como o algoritmo do MP3 explora cada uma dessas características para obter a compressão do áudio?

O algoritmo do MP3 explora essas características de forma a excluir frequências que o ouvido humano não consegue escutar e também tirando vantagem de utilizar as frequências mais sensíveis do ouvido e descartar sons baixos logo após sons altos (dentro do delay de 50ms). Isto permite arquivos até 12 vezes menores com perdas que não são perceptíveis ao ouvido humano.

Lista 3 – Percepção, cor e imagens

1. Cones e bastonetes são fotorreceptores importantes para o sistema visual humano. Cada um está presente em diferentes regiões e é sensível a características diferentes. Por nós possuímos quantidades bem diferentes de cada um, somos mais sensíveis a uma característica do que a outra, o que é explorado na compressão de vídeos e imagens. Defina a função dos cones e dos bastonetes e a qual característica o olho humano é mais sensível.

Cones são fotoreceptores para cor e bastonetes são fotoreceptores para intensidade luminosa. Por possuímos 20 vezes mais bastonetes do que cones, somos mais sensíveis ao brilho e intensidade da luz.

2. O CCD é o dispositivo responsável pela captura de imagens.

a. Explique seu funcionamento.

A imagem é “entendida” como uma matriz de pontos. A luz proveniente de cada ponto da imagem é capturada por um sensor (CCD). É composto por uma malha de material fotossensível. Cada célula da malha corresponde a um ponto da imagem. A intensidade da luz incidente em cada célula é convertida em sinal elétrico. Amostragem e quantização. Informação digitalizada é armazenada no Frame Buffer como uma matriz de pontos.

b. No que difere a captura pelo CCD de imagens em tons de cinza em relação a imagens coloridas?

Nas imagens coloridas, utiliza-se um CCD para cada primária RGB, e um filtro separa as componentes para o CCD correto.

3. Quanto tempo seria necessário para transmitir uma imagem truecolor sem compressão com resolução 1024x768 em uma rede a 1.5 Mbps? (Dica: pesquisa quantos bits por canal existem para uma imagem truecolor)

Truecolor: 24 bits, 8 bits por canal. Levaria 12,6 segundos.

Ao realizar os cálculos do exercício 3, você consegue perceber a real necessidade de compressão em imagens (e qualquer outro tipo de mídia digital). Felizmente, imagens possuem muitos dados redundantes que podem ser explorados visando compressão. Os próximos exercícios são relacionados às técnicas que permitem explorar os três tipos de redundância encontrados em imagens: redundância estatística, redundância espacial e redundância psicovisual.

4. Explique o que é a redundância estatística e quais técnicas já estudadas anteriormente podem ser utilizadas para explorá-la?

Toda imagem possui uma Função de Densidade de Probabilidade (pdf). Valores dos pixels em uma imagem tem pdf não uniforme. Os métodos de

codificação estatística (Huffman, codificação aritmética) podem ser usados para compressão de imagens.

5. Qual a diferença entre redundância estatística e redundância espacial? Dê uma técnica sem perdas e uma técnica com perdas que pode ser utilizada para remoção da redundância espacial.

Na redundância espacial utiliza-se a correlação entre pixels vizinhos, enquanto que na redundância estatística utiliza-se a frequência de cada pixel em toda a imagem. Lossless: codificação por diferença. Lossy: codificação preditiva.

6. A utilização de transformadas DCT é um passo importante na remoção de redundâncias psicovisuais. Apenas sua utilização, porém, não comprime os dados. Explique então o porquê de ela ser importante.

Após aplicar a DCT, as regiões da imagem que possuem uma única cor geram matrizes com coeficientes DC idênticos (ou próximos) e poucos coeficientes AC. As regiões da imagem que possuem transições de cores geram matrizes com coeficientes DC distintos e muitos coeficientes AC. Isso ajuda a detectar e eliminar informações redundantes.

7. Explique o processo de compressão JPEG, destacando em que momentos ocorrem as remoções de redundâncias espacial, estatística e psicovisual.

Preparação da imagem/bloco; transformada DCT (psicovisual); quantização; codificação por entropia (espacial); construção do quadro; decodificação.

Lista 4 – Vídeo

1. Um sinal de vídeo nada mais é do que uma sequência de imagens (ou quadros) que, quando amostrada e reproduzida a certa frequência, provoca a sensação de movimento.

a. Qual a frequência (Hz ou fps – frames per second) utilizada pelo padrão de TV?

A taxa depende do padrão adotado, variando de 25 Hz nos padrões PAL e SECAM e 29,97 Hz no padrão NTSC.

b. Qual a vantagem e desvantagem da amostragem progressiva e da entrelaçada?

A amostragem progressiva possui uma imagem com menos flickering, porém a taxa de refresh da tela é limitada de acordo com o número de amostras. A amostragem entrelaçada permite que se aumente a taxa de refresh da imagem

sem aumentar o número de amostras, fazendo com que a tava vertical seja o dobro da tava horizontal.

2. Em relação à compressão, qual a vantagem de se transformar o espaço de cores de um quadro de RGB para YCbCr?

A vantagem é que a cor é representada de modo mais eficiente e permite mais adiante uma redução da largura de banda necessária ao manipular a crominância.

3. Explique as diferenças e consequências de utilização dos formatos de amostragem YcbCr 4:4:4, 4:2:2 e 4:2:0.

As diferenças destas amostragens consistem na quantidade de componentes de crominância em relação aos de luminância nas proporções dadas pelos números das amostragens. Sabendo que o olho humano é mais sensível à luminância, podemos diminuir as amostragens de crominância, diminuindo a banda usada sem que a mudança seja muito perceptível para o olho humano.

4. Pelo fato de o vídeo se tratar de uma sequência de quadros, podemos aplicar a cada um as técnicas de compressão relacionadas a imagens, para remoção de redundâncias estatísticas, espaciais e psicovisuais. Além disso, o vídeo traz um novo tipo de redundância que pode ser explorado. Explique qual é este tipo e como ele pode ser explorado para aumentar a compressão.

Vídeos por consistirem em sequência de imagens podem possuir características bem peculiares. Quadros adjacentes costumam ser parecidos com pequenas alterações em alguns pontos e portanto este tipo de redundância temporal pode ser explorada de modo a enviar uma imagem inteira e em seguida mudar apenas os pontos onde haja diferença.

5. Explique o porquê de os métodos de previsão de movimento serem classificados como não perfeitos. O que se pode fazer para melhorar a estimativa?

Os métodos de previsão são classificados como não perfeitos pois pode ocorrer propagação de erros entre os quadros. Para melhorar a estimativa usam-se algoritmos de estimativa e compensação de movimento que tentam evitar que estes erros se propaguem.

6. Durante a codificação, cada quadro recebe um nome especial, dependendo de como será codificado.

a. Defina o que são quadros I, P e B.

I: Intracoded Frames - são codificados sem nenhuma referência a outros quadros.

P: Predictive Frames - são codificados em relação a um quadro I ou P anterior.

B: Bidirectional Frames - são codificados em relação a um quadro I ou P anterior e um posterior.

b. Defina GOP span e Prediction span.

GOP span define o número de quadros entre 2 quadros I sucessivos.

Prediction span define número de quadros entre um quadro P e o quadro I ou P imediatamente anterior.

c. Qual a limitação encontrada nos quadros P?

Como propagam erros a quantidade de quadros P entre quadros I deve ser limitada.. isso acaba limitando também a taxa de compressão do vídeo.

d. Qual a vantagem e desvantagem da utilização de quadros B?

Permite uma taxa de compressão maior e a não propagação de erros. Isto ao custo de possuir um aumento no tempo (delay) para codificação e decodificação.

7. A evolução das técnicas de codificação, bem como das tecnologias de transmissão e reprodução multimídia, resultou na criação dos padrões MPEG-4 e H.264.

a. Cite duas características do padrão MPEG-4 que o diferenciam de seus antecessores.

Visa fornecer padrões para integrar a produção, distribuição e acesso ao conteúdo audiovisual e insere o conceito de Objeto

b. Cite vantagens da utilização do padrão H.264.

Possui técnicas de compressão não disponíveis em outros padrões o que o torna o padrão mais avançado disponível atualmente.

8. Muitas pessoas erroneamente entendem que a extensão de arquivo de vídeo (.AVI, por exemplo) representa a compressão que está sendo utilizada, sendo que na verdade, ela é somente um formato de representação. Explique a diferença entre formatos de representação e formatos de codificação.

O Formato de representação serve apenas como tipo de pacote de dados que esta sendo usado e não tem relação com a sua codificação. Arquivos com mesmo formato de representação podem possuir codificações diferentes e portanto qualidades diferentes.