

Acuidade Visual, Codificação de Vídeo e Quantização

UFABC - ESTI019 – Codificação Multimídia

Profs. Celso Kurashima e Mário Minami

mario.minami@ufabc.edu.br

celso.kurashima@ufabc.edu.br



Acuidade Visual e Parâmetros de Visualização de Imagens

Acuidade Visual

- Muitos animais, particularmente pássaros, possuem visão que produz resolução espacial muito mais alta do que os humanos.
- Visão Humana consegue a resolução espacial mais alta apenas numa pequena área no centro do campo de visão (fovea), onde a densidade de cones sensíveis é a mais alta.
- Numa distância de 50-cm de vista, detalhes com uma largura de 1 mm representa um angulo de pouco mais que um décimo de um grau.
- A acuidade (resolução espacial) é normalmente especificada em unidades de ciclos por grau (*cycles per degree*).

Acuidade Visual

- O limite superior (detalhamento fino) visível com o olho humano é cerca de 50 *cycles per degree* (cpd), que seria correspondente a graduação em que o brilho varia do mínimo ao máximo aproximadamente cinco vezes sobre o mesmo 1 mm.
- No espaçamento fino, contraste de 100% seria necessário, em outras palavras linhas pretas e espaços brancos.
- É disto que advem a especificação comum de que as linhas mais finas distinguíveis sem ajuda óptica são de aproximadamente 100 μm .

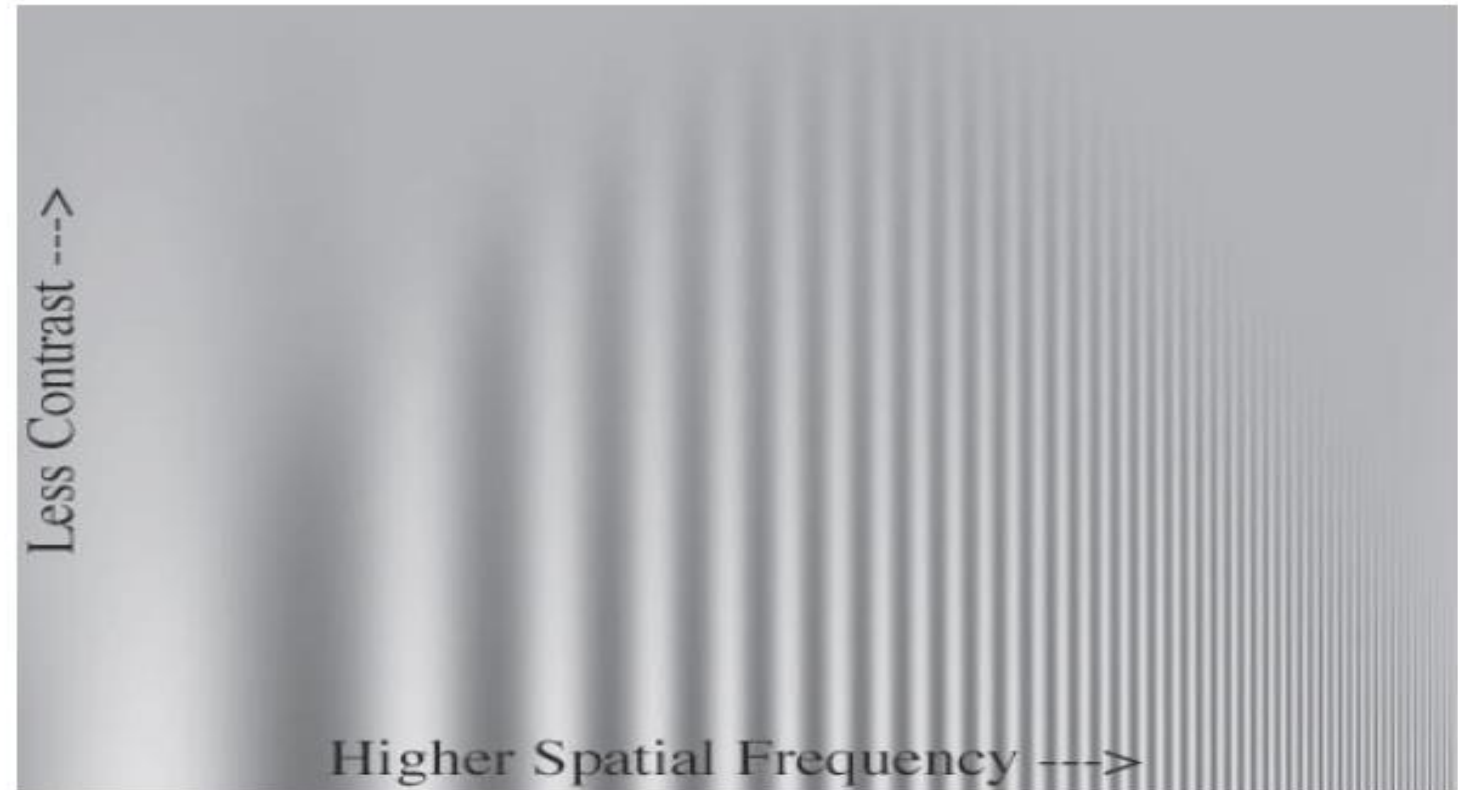
(RUSS, J. C., 2007, p. 94)

Exemplo

- Numa distância normal de visão de aproximadamente 50 cm, 1 mm na imagem é o tamanho ótimo para detectar a presença de detalhes.
- Num monitor de computador típico, isto corresponde a +/- 4 pixels.
- A medida que a frequência espacial diminui (detalhes tornam-se maiores), o contraste necessário aumenta, de modo que quando a distância em que o brilho varia do mínimo ao máximo é de 1 cm, o contraste necessário é dez vezes maior.
- A variação da resolução espacial (“acuidade”) com o contraste requerido é denominado de função de transferência de modulação (vide figura).

Contraste requerido vs Frequencia espacial

Figure 2.9 *Illustration of the modulation transfer function for human vision, showing that the greatest ability to resolve low-contrast details occurs at an intermediate spatial frequency and becomes poorer for both smaller and larger details.*



Resolução retina

- Um sistema com resolução retina (*retina display*) é aquele que atinge o limiar da capacidade da visão humana.
- A visão humana tem uma resolução espacial de 50 CPD (*cycle per degree*).

Resolução retina

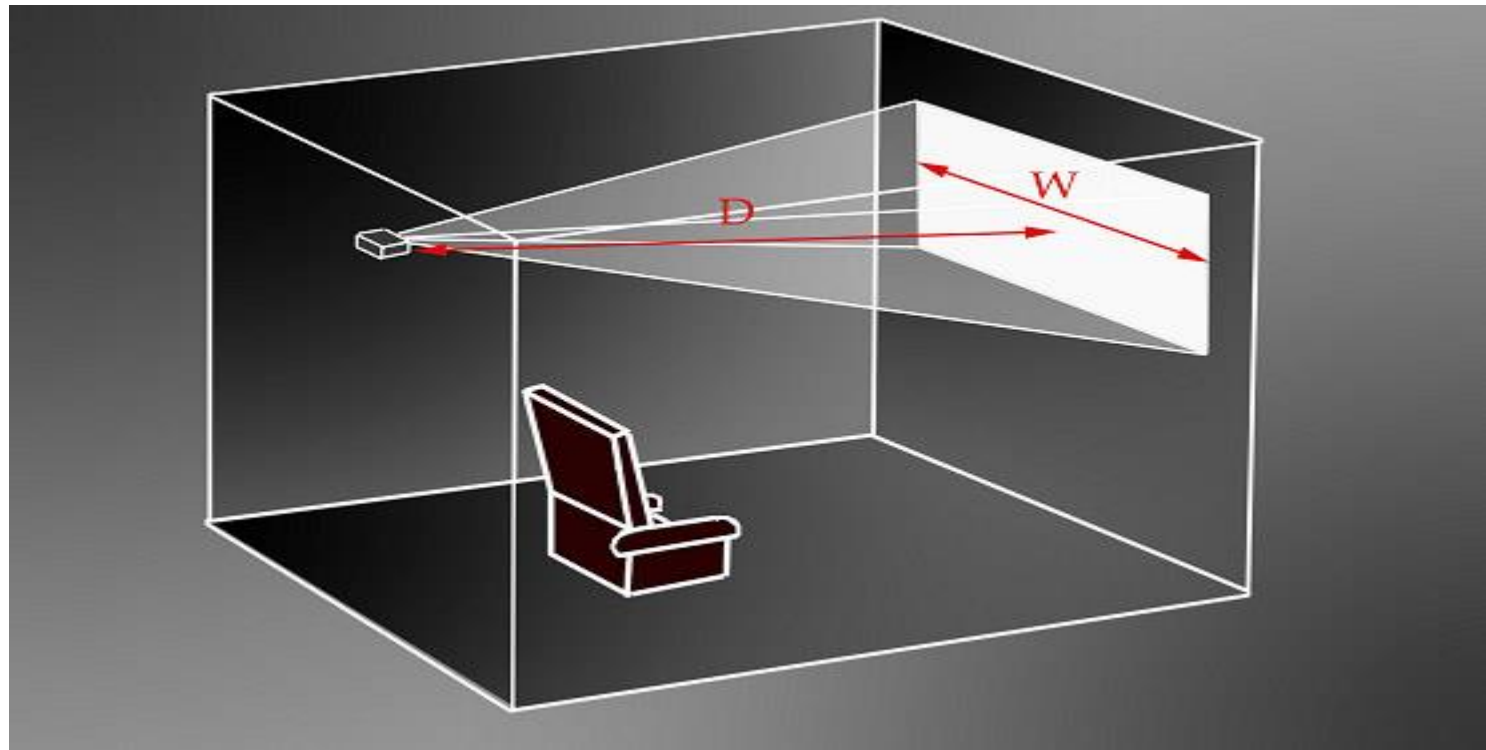
- a resolução retina deve ser calculada através da relação entre a distância do usuário e o tamanho da tela.

$$\begin{aligned} \text{resolução_retina} &= \frac{\text{dimensão}}{\text{distância_observador} \times \tan\left(\frac{1}{50}\right)} \\ &\cong \frac{2865 \times \text{dimensão}}{\text{distância_observador}} \end{aligned}$$

Razão de Tiro de Projetor

- Razão de tiro de um projetor é definido como a distância (**D**), medida da lente à tela, que um projetor que é colocado a partir da tela, dividido pela largura (**W**) da imagem que vai projetar (**D / W**).
- (<http://www.theprojectorpros.com/>)

Razão de Tiro de Projetor

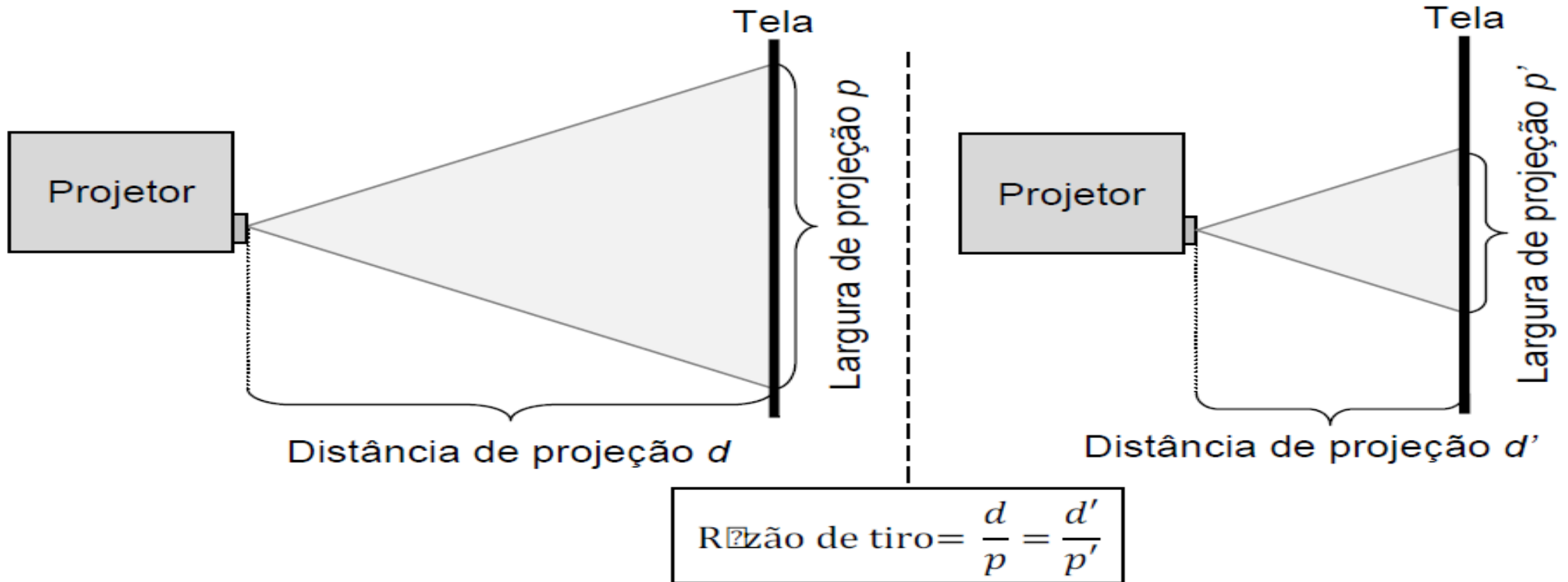


http://www.theprojectorpros.com/learn-s-learn-p-theater_throw_ratios.htm

Distância dos projetores

- A diagonal (ou largura, ou altura) de projeção é linearmente proporcional à distância do projetor em relação à tela (considerando que a razão de tiro seja constante), ou seja, quanto menor a dimensão de projeção, menor a distância do projetor em relação à tela.

Relação entre a distância e a largura de projeção





Codificação de Video

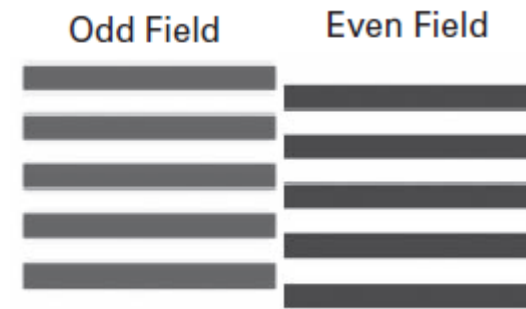
Moving Pictures = Imagens em Movimento = Video?

Moving Pictures = Movie?

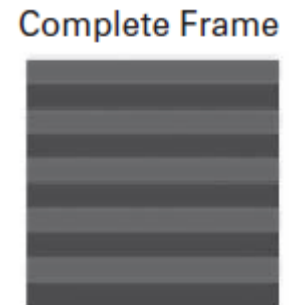
Movie = filme = video + audio?

Taxa de quadros na transmissão de televisão

- No sistema PAL-M (Brasil) de TV analógica a taxa de quadros de vídeo é 29,97 fps.
 - Resolução da imagem: 525 linhas (transmissão entrelaçada).



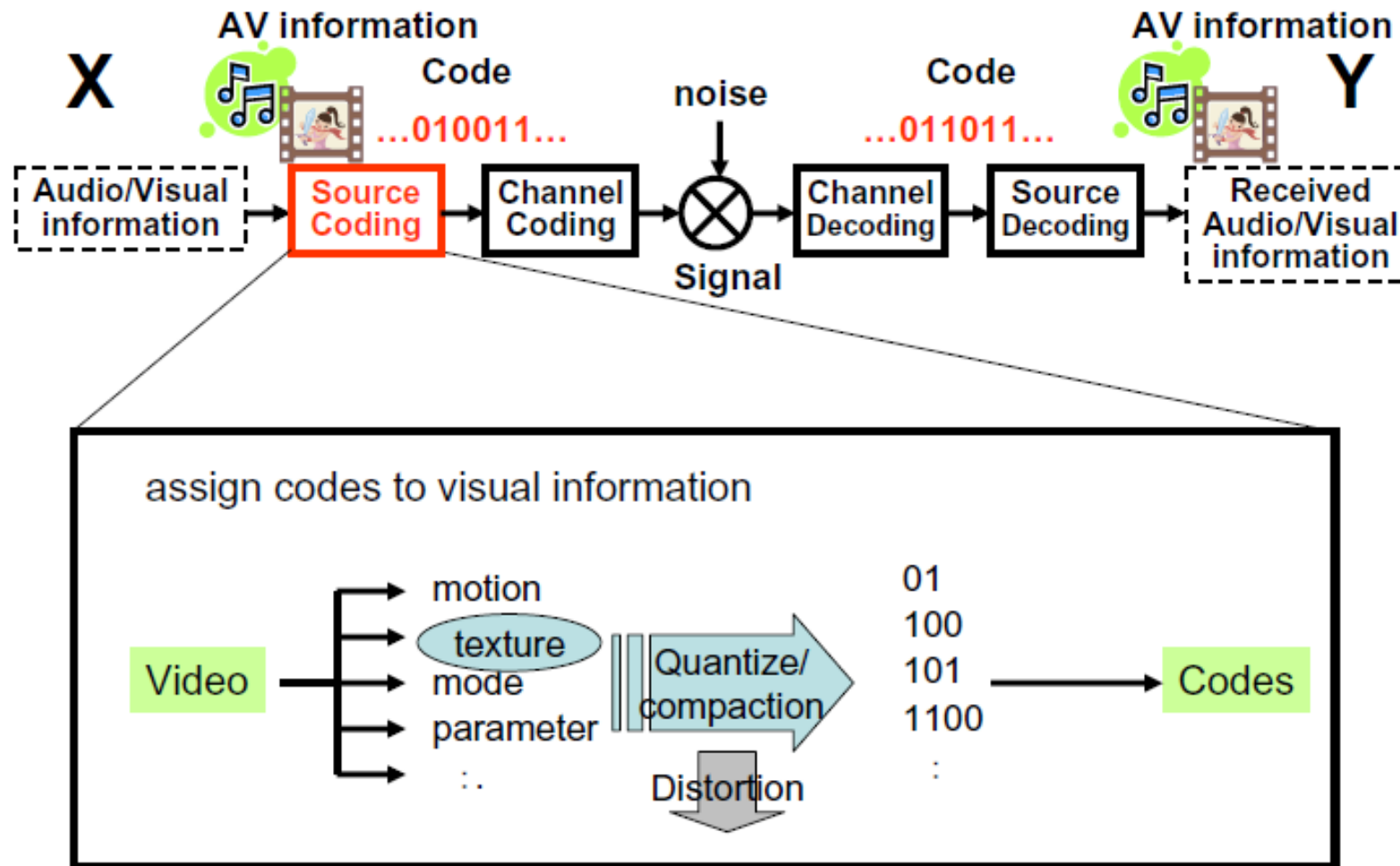
- Na TV Digital, sistema SBTVD ou ISDB-TB, a taxa de quadros de vídeo é 59,94 fps.
 - Atualmente tem-se:
 - HD: 1280x720p em 16:9 (1280 x 720 - 60p) - progressivo
 - Full HD: 1920x1080i em 16:9 (1920 x 1080 - 60i) – entrelaçado.



Codificação (compressão) de vídeo

- Codificação de vídeo visando a compressão de informações digitais começou seu desenvolvimento a partir da necessidade de transmitir uma grande quantidade de dados de vídeo digital através de linha de comunicação com banda estreita.
- Portanto, é possível traçar uma analogia entre o modelo de informação e de telecomunicações de Shannon e a composição do sistema de transmissão de codificação de vídeo digital.
- No lado emissor de um sistema de transmissão de vídeo digital, o sinal de vídeo analógico adquirido a partir de uma câmara é digitalmente amostrado e quantizado.

Source Coding – Codificação de fonte



Visual information is distorted by both source coding and channel error.
→ Quality is measured by PSNR versus Bitrate.

A necessidade de compressão de vídeo

- Ex. ■ Digital video studio standard ITU-R Rec. 601

	Y	Cb	Cr
Sampling rate	13.5 MHz	6.75 MHz	6.75 MHz
Quantization	8 bit	8 bit	8 bit
Raw bit rate	216 Mbps		
w/o blanking intervals	166 Mbps		

- Some interesting bit-rates (some evolving with time)
 - Terrestrial TV broadcasting channel ~20 Mbps
 - DVD (4.7...17 GB/length of movie) 5...20 Mbps
 - Ethernet/Fast Ethernet <10/100/1000 Mbps
 - DSL downlink 0.2 ... 10+ Mbps
 - Wireless cellular data 0.02 ... 1+ Mbps

Principais Componentes da Codificação de Video

What are the key technology of video coding?



Principais Componentes da Codificação de Vídeo

- Codificação de vídeo geralmente consiste em quatro componentes principais, incluindo a predição, transformada, quantização, e codificação por entropia.
- **Predição** reduz a redundância relativa explorando a correlação dentro de uma imagem ou em várias imagens.
- O sinal residual que representa a diferença entre o sinal original e o sinal predito (estimado) é codificado.
- **Transformada** é um processo para compactação de energia do sinal para reduzir a correlação dos símbolos.
- Na prática, o sinal é transformado de um domínio espacial para um domínio de frequência.

Componentes da Codificação de Video

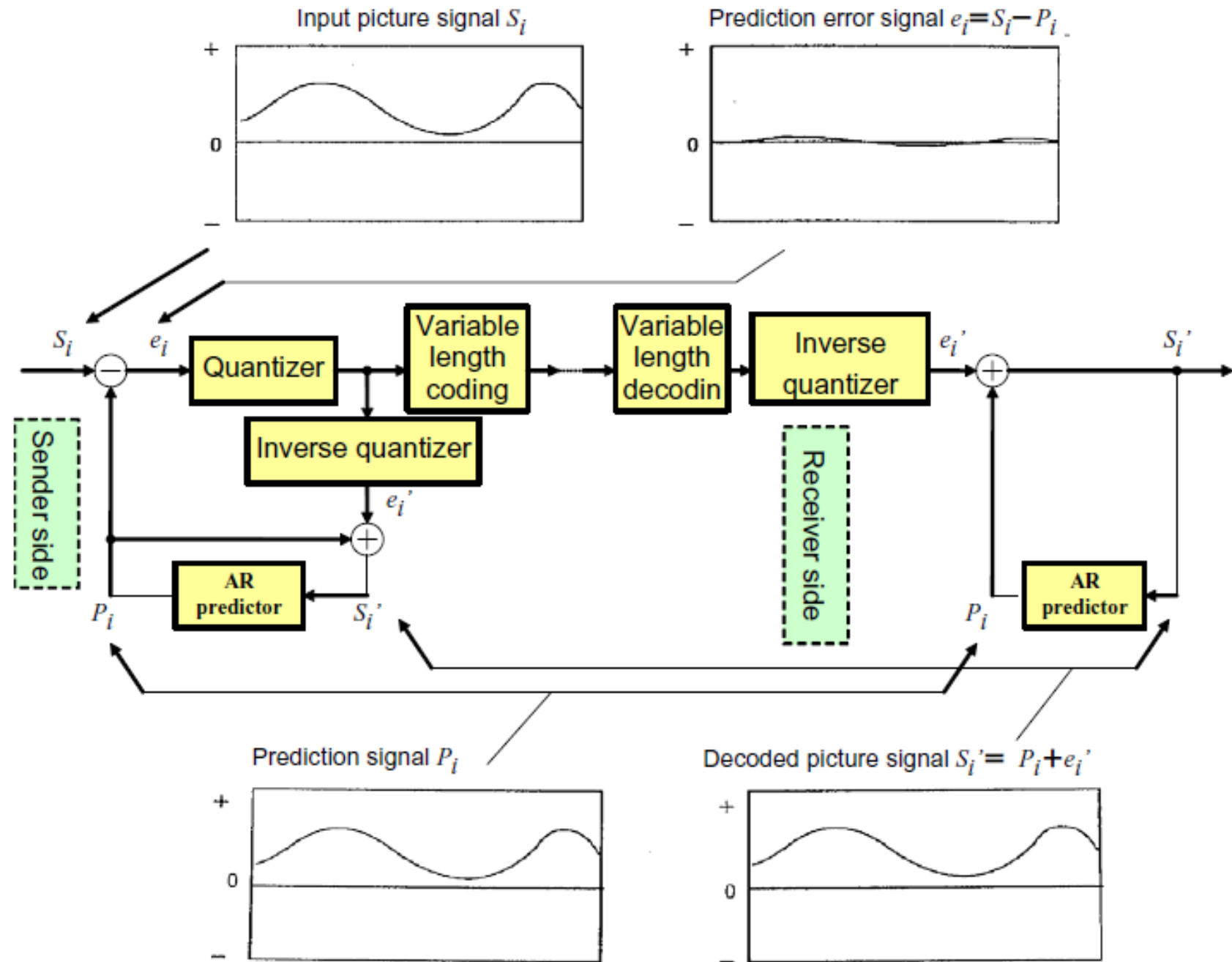
- Existem várias transformações que têm sido utilizadas em padrões típicos de codificação de imagem e vídeo, incluindo Transformação de Coseno Discreto (DCT) e Transformada Wavelet Discreta (DWT).
- **Quantização** é uma técnica que reduz a quantidade de informação diretamente. Existem dois métodos principais de quantização incluindo quantização escalar e quantização vetorial.
- **Codificação por Entropia** é um método de codificação reversível baseado na caracterização estatística dos símbolos a serem codificados. A codificação de Huffman e a codificação aritmética são exemplos típicos de esquemas de codificação de entropia.

Predição

- Uma imagem tem alta correlação entre pixels vizinhos em ambas as direções espaciais e temporais.
- Consequentemente, a quantidade de informação pode ser reduzida pela combinação da predição entre pixels e a codificação do erro de predição (sinal residual).
- A predição que explora a correlação espacial dentro de uma imagem é conhecida como **Intra prediction**, enquanto a predição que explora a correlação temporal entre duas ou mais imagens é conhecida como **Inter prediction**.
- Um método para explorar mais a correlação entre quadros é utilizar a predição de movimento, que é referida como **Predição Compensada por Movimento** - **Motion Compensated Prediction**.

Predição

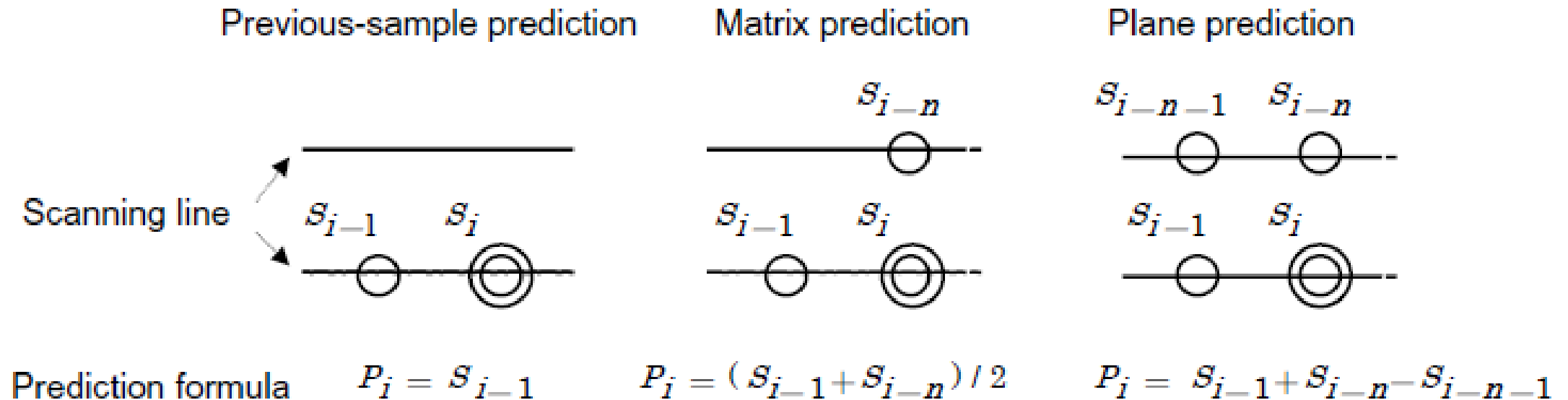
- Esta figura mostra a diferença de potência entre vários sinais num sistema de codificação de vídeo típico.



Intra Frame Prediction

- **Predição Intra Frame** é uma técnica de predição que usa os pixels vizinhos dentro de um quadro.
- Três métodos de predição, incluindo Predição de amostra prévia, Predição de matriz, e Predição de plano, são mostrados como exemplos de Predição Intra-Frame na Fig. 9.
- Predição de amostra prévia usa pixels vizinhos na direção horizontal como um pixel de predição, Predição de matriz usa pixels vizinhos em ambas direções horizontais e verticais e Predição de plano usa pixels vizinhos na direção horizontal e subtrai os valores de pixel nas mesmas posições na linha anterior .

Intra Frame Prediction



P_i : Prediction value

n : Number of pixels per line

⊙ : Predicted pixel

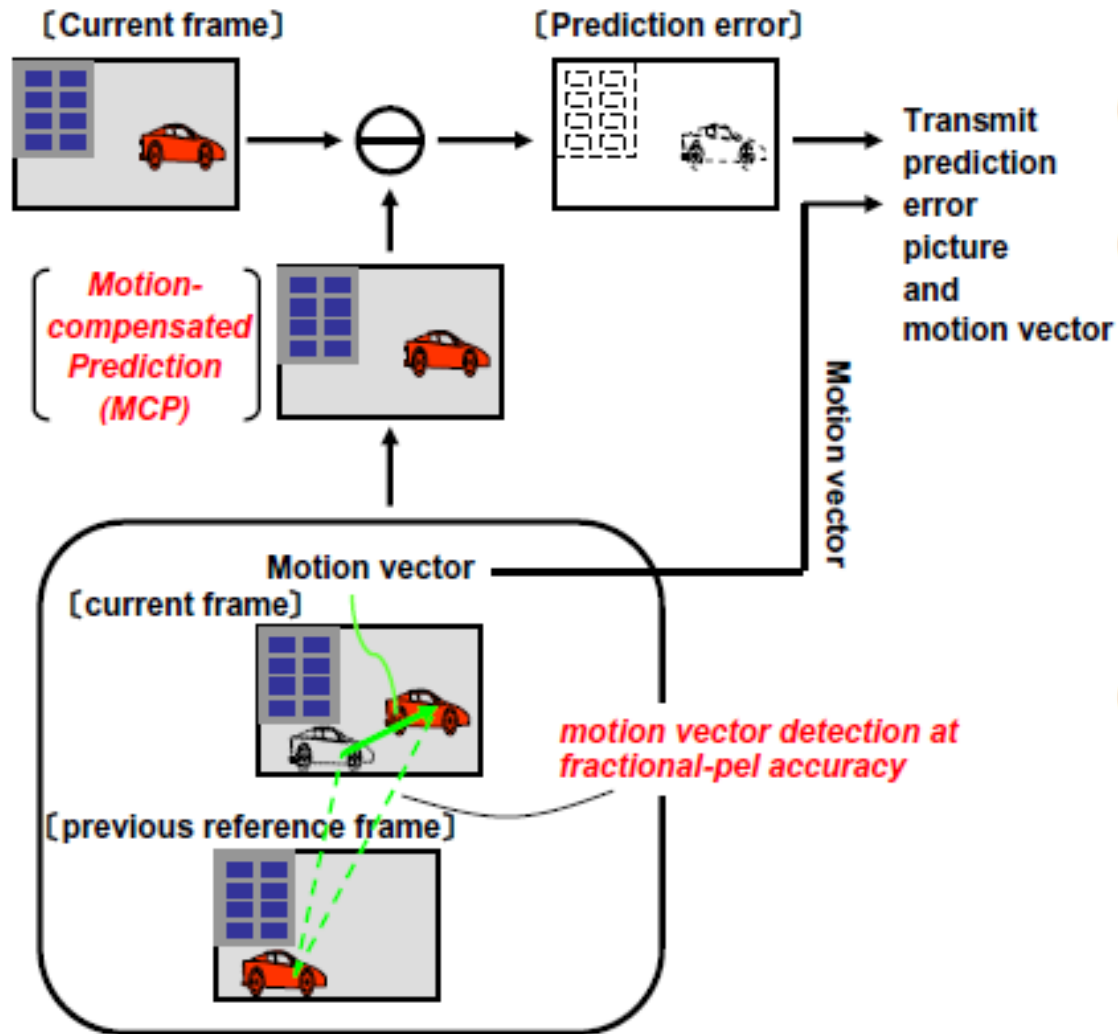
○ : Reference pixel

Fig. 9 Examples of Intra Frame Prediction

Motion Compensated Prediction

- Predição compensada por movimento é uma técnica que cria uma imagem de predição que se assemelha à imagem atual por translação linear de um bloco dentro de uma imagem de referência que já foi transmitida e decodificada.
- A compressão é obtida codificando a diferença entre as imagens preditas (estimadas) e as originais.
- O princípio da Predição Compensada por Movimento é mostrado na Fig. 10.
- A *Motion Compensated Prediction* pode reduzir a energia do sinal residual em comparação com a diferença simples entre quadros.
- O sinal de diferença diminui dramaticamente quando a Predição Compensada por Movimento é utilizada.

Motion Compensated Prediction



- ① Divide current frame into block $x_n(i, j)$
- ② Calculate distortion (Sum of Absolute Difference: D^L) between $x_n^L(i, j)$ and $x_{n-1}^L(i, j)$ of current and previous reference frames, respectively.

$$D^L = \sum_{i,j} |x_n(i, j) - x_{n-1}^L(i, j)|$$

- ③ Find the block with minimum distortion in the previous reference frame and obtain motion vector which indicates translational movement

Fig. 10 Principle of Motion Compensated Prediction

Motion Compensated Prediction

	Original signal	Differential signal between adjacent pixels	Differential signal between frames	Motion compensated differential signal between frames
Entropy [bit/pixel]	7.12	4.37	6.05	4.16
Signal Power (Variance)	118.22	47.89	10.63	0.58

- Esta tabela mostra um exemplo das características de Entropia e Potencia do Sinal para os sinais original, com predição Intra Quadro, com predição Inter-Quadro e com Predição Compensada por Movimento de uma imagem de HDTV.
- Observe que a potência do sinal diminui acentuadamente quando a Predição Compensada por Movimento é utilizada.

Transformada

- Transformada é o método de converter um sinal de imagem em outro domínio de sinal e centralizar a potência do sinal em bandas de frequência específicas.
- Existem a DCT e a DWT para esta finalidade, que são usados nos padrões atuais de codificação de imagem.
- A DCT converte o sinal do domínio espacial no domínio da frequência utilizando uma janela com largura fixa para a transformação.
- Normalmente, uma imagem é dividida em blocos de pixels $N \times N$ (N pixels de largura tanto horizontal como vertical) e a transformação é executada para cada bloco de pixels.

Transformada DCT

- Depois de realizar a DCT de um sinal de vídeo, uma porção significativa de energia tende a ser concentrada nos coeficientes de DCT nas bandas de baixa frequência, mesmo se não houver desvio estatístico num bloco de pixels propriamente dito.
- Portanto, a codificação é realizada de acordo com o sistema visual humano e o desvio estatístico no domínio do coeficiente DCT de um sinal de imagem.
- Um exemplo de uma imagem após transformação por DCT, no caso de 8x8 pixels, é mostrado na figura seguinte.

Transformada DWT

- DWT é um dos métodos de transformação que utiliza a transformação base feita pela operação de expandir e mover uma função localizada no domínio de frequência.
- DWT permite o uso de janelas cujos tamanhos são diferentes de acordo com as frequências, e tem a característica de alta resposta para ambas as porções de baixa frequência e de alta frequência dos sinais.

Quantização

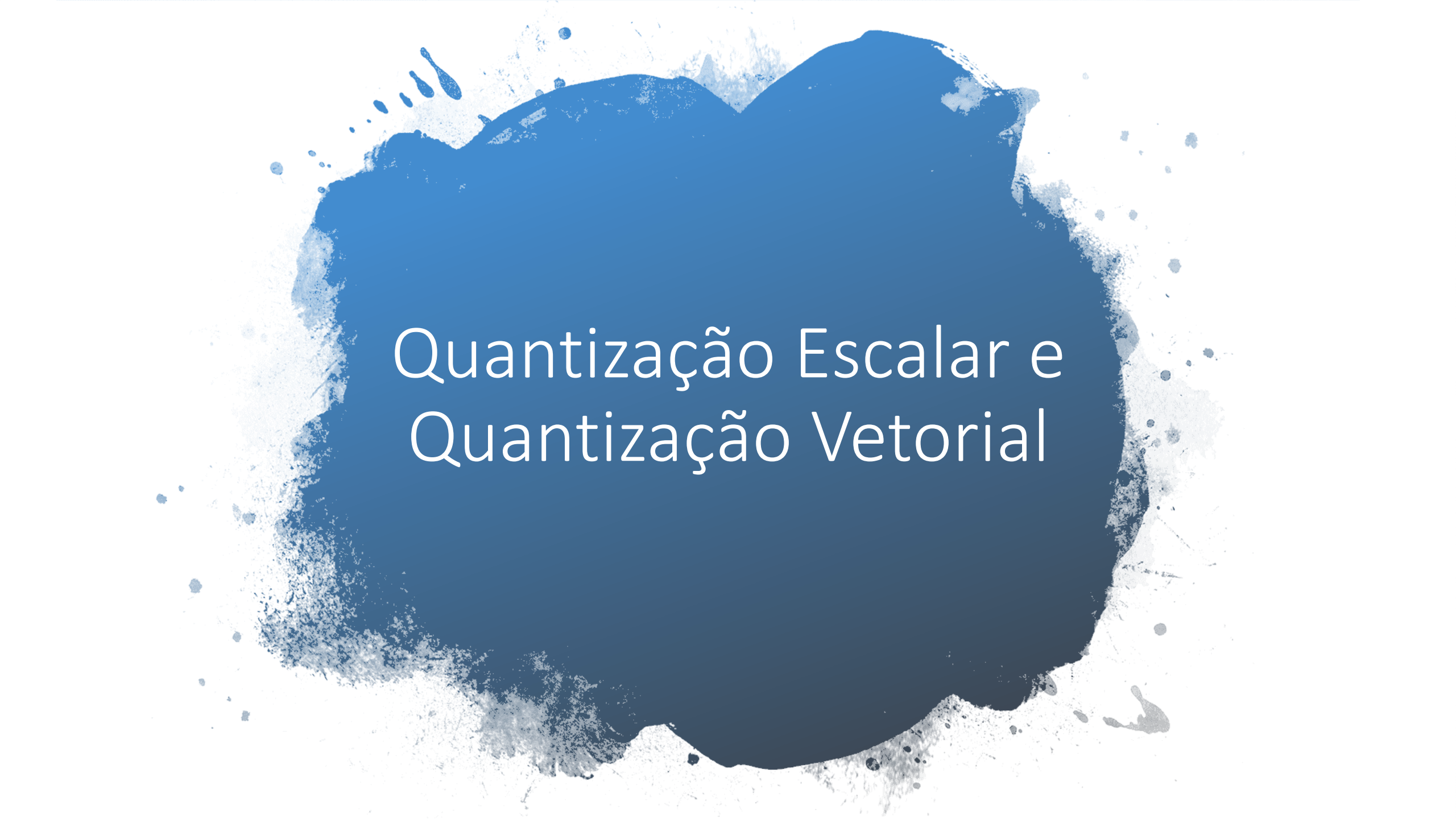
- Quantização é uma técnica de reduzir a quantidade de informação diretamente, e existem principalmente dois métodos bem conhecidos para a compressão de vídeo, que são Quantização Escalar e Quantização Vetorial.
- A Quantização Vetorial é uma operação que quantiza várias amostras ao mesmo tempo e as expressa com um vetor representativo que dá a melhor aproximação das amostras.
- As fontes de informação que consistem em muitas dimensões são quantizadas por um dos pontos representativos de um espaço multidimensional por Quantização Vetorial.

Codificação por Entropia

- A Codificação por Entropia é um método de descrever a informação de modo, informação de vetor de movimento, valores quantizados, etc. como uma série de sinais binários que consiste apenas em 0 e 1 (binarização).
- A quantidade total de códigos é redutível pela atribuição de palavras codificadas de acordo com a probabilidade de ocorrência de símbolos.

Codificação Huffman

- A codificação Huffman e a codificação aritmética são métodos de codificação de entropia típicos utilizados na codificação de vídeo.
- A codificação Huffman é um método de formação e utilização de uma tabela de código de comprimento variável que associa símbolos e palavras-código.
- Este método pode encurtar o comprimento médio do código, atribuindo códigos curtos a símbolos com probabilidade de ocorrência alta e códigos longos a símbolos com baixa probabilidade de ocorrência.

A large, irregular blue ink splash or blotch serves as the background for the text. The splash is centered and has a textured, painterly appearance with various shades of blue and white. The text is white and centered within the splash.

Quantização Escalar e Quantização Vetorial

Quantização Escalar

- Muitas das ideias fundamentais de quantização e compressão são facilmente introduzidas no contexto da quantização escalar
- Por exemplo, qualquer número real x pode ser arredondado para o número inteiro mais próximo, digamos

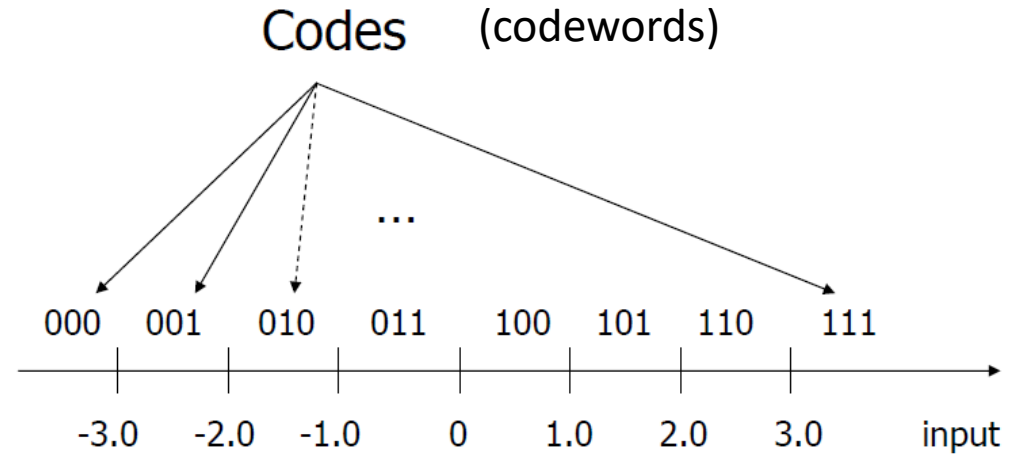
$$Q(x) = \text{round}(x)$$

- Mapeia a reta \mathbb{R} (um espaço contínuo) em um espaço discreto

Componentes de um Quantizador

- Mapeamento do codificador: divide o intervalo de valores que a fonte gera em um número de intervalos.
- Cada intervalo é então mapeado para uma palavra-código (*codeword*).
- É um muitos casos um mapeamento irreversível. A *codeword* apenas identifica o intervalo, não o valor original.
- Se o valor da fonte for analógico, é chamado de conversor A/D

Mapeamento num Codificador Escalar de 3 bits



(codewords)

Input Codes	Output
000	-3.5
001	-2.5
010	-1.5
011	-0.5
100	0.5
101	1.5
110	2.5
111	3.5

Mapeamento num conversor D/A de 3 bits (Decoder Escalar)

Componentes de um Quantizador II

- Decodificador: dada a palavra de código, o decodificador dá um valor estimado que a fonte pode ter gerado
- Normalmente, é o ponto médio do intervalo mas uma estimativa mais precisa irá depender da distribuição dos valores no intervalo.
- Ao estimar o valor, o decodificador pode gerar alguns erros.

Compressão de Imagem



2bits/pixel



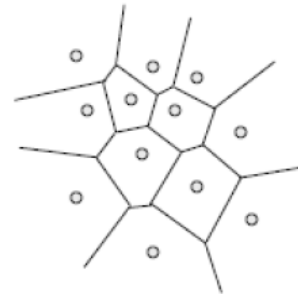
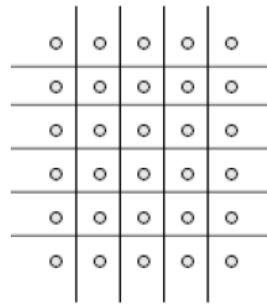
1bit/pixel

Quantização Vetorial (VQ)

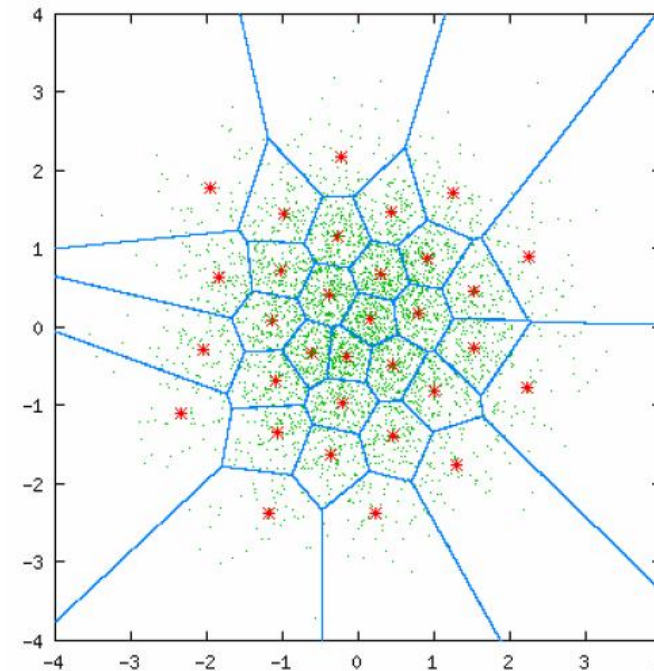
- A quantização vetorial (VQ) é método de compressão com perdas (*lossy*)
- No passado, o projeto era uma tarefa muito difícil, pois a abordagem era uma integração multidimensional
- Linde, Buzo e Grey (LBG) propuseram um algoritmo de projeto VQ baseado em sequências de treinamento (treinamento supervisionado).
- O uso de uma sequência de treinamento ignora a necessidade de integração multidimensional

Quantização vetorial - centróides

- Para definir regiões no espaço, um conjunto de pontos contidos em cada região sendo projetado em um vetor representativo (centróide)
- Exemplos de espaços 2D:



Regiões de Voronoi em
linhas contínuas, centróides
nos pontos circulares

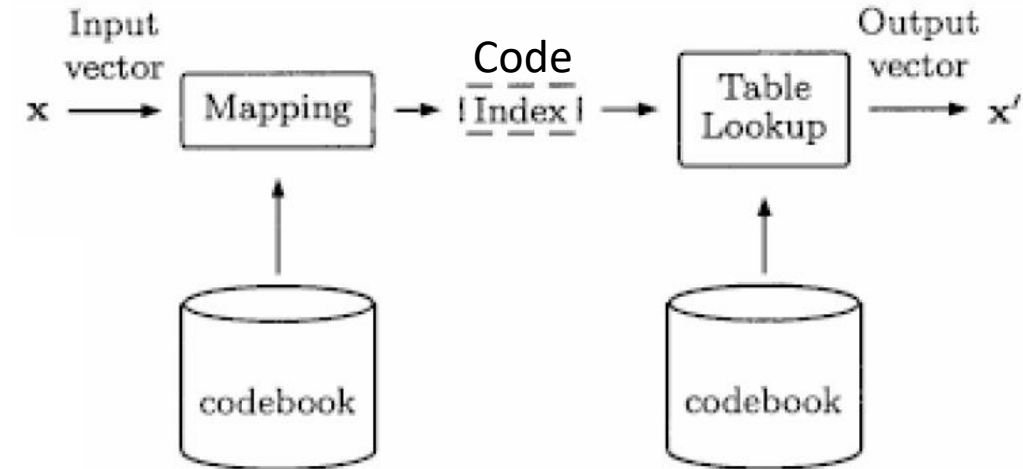
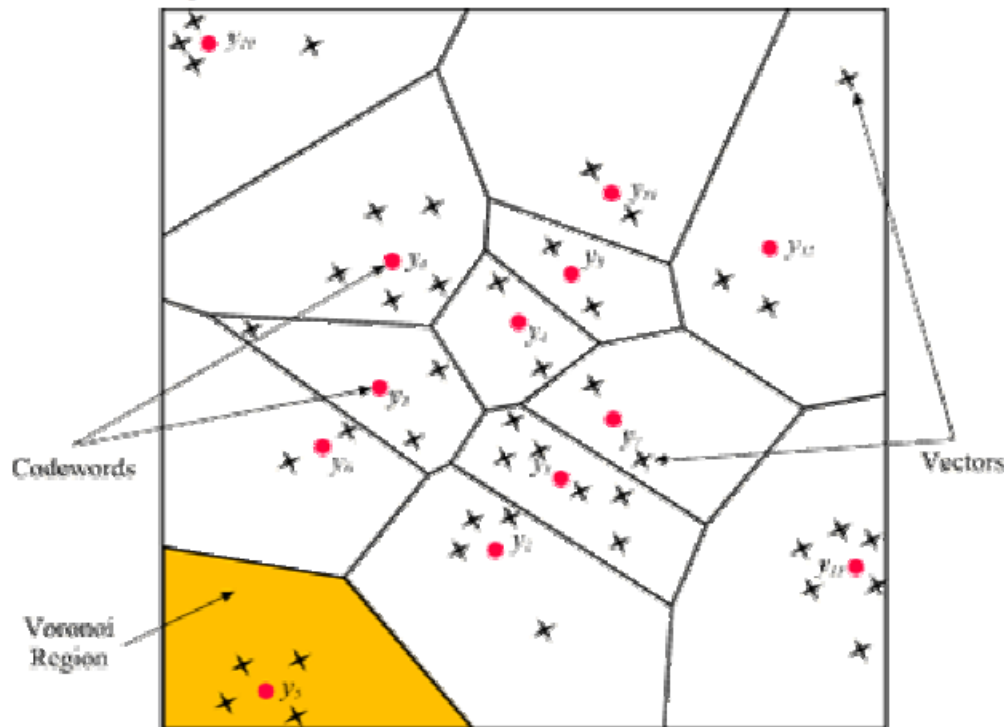


Termos Técnicos

- ***Codebook*** (livro de código)
 - Em criptografia, um *codebook* é um documento usado para implementar um código.
 - Um *codebook* contém uma tabela de pesquisa para codificação e decodificação; cada palavra ou frase tem um ou mais vetores que o substituem a *codeword*
- ***Codeword*** (palavra-código) = ***Codevector***
 - é um elemento de um código
- ***Codeindex*** (índice do vetor-código)
 - É o inteiro que rotula uma *codeword*

Blocos de uma VQ

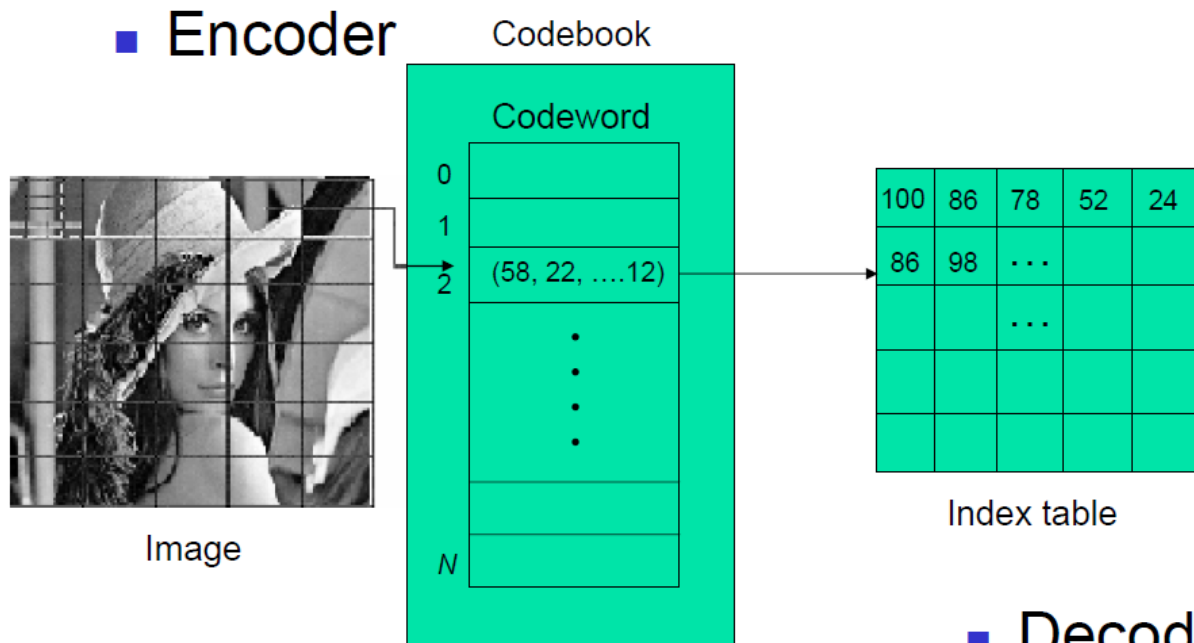
■ 2D Space



Formado pelas codewords

Algoritmo LBG: Projeto do Codebook
Y. Linde, A. Buzo, and R. M. Grey, **An algorithm for vector quantizer design**, IEEE Transactions on Communication, 28(1), 84–95 (1980)

■ Encoder

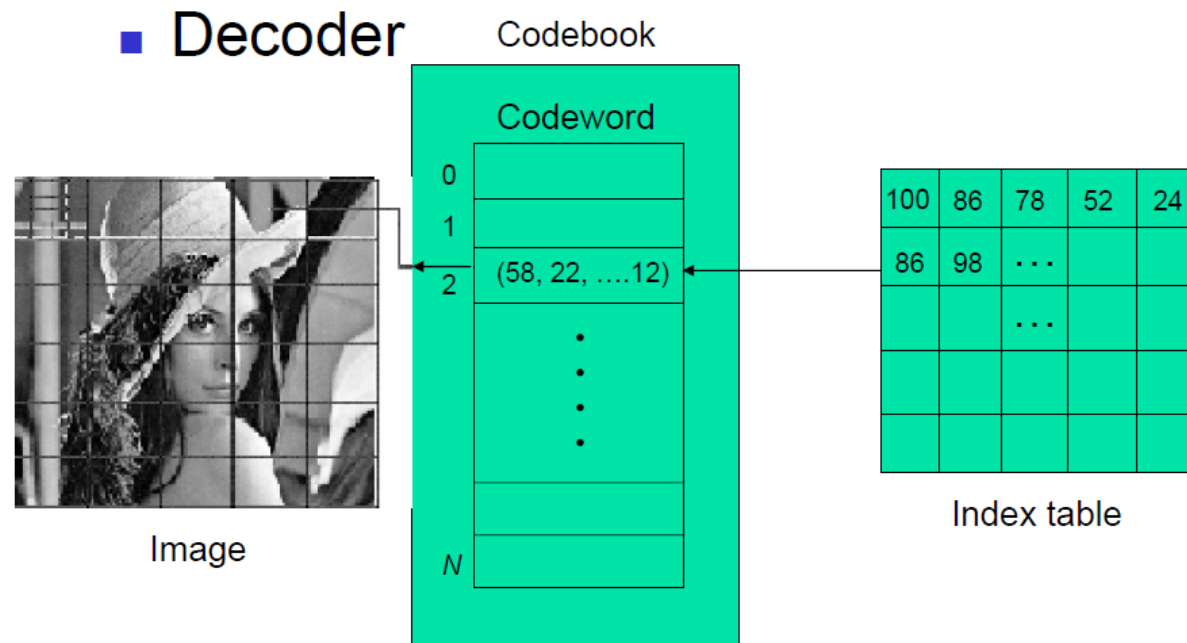


VQ para compressão de Imagem

Distância Euclidiana (L2)

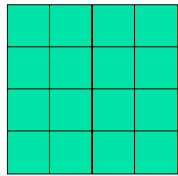
$$d(x, y_i) = \|x - y_i\|^2 = \sum_{j=1}^k (x_j - y_{ij})^2$$

■ Decoder

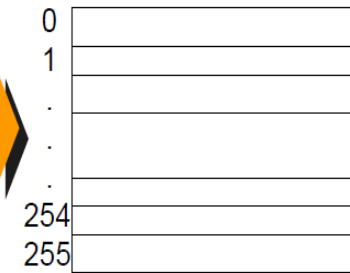
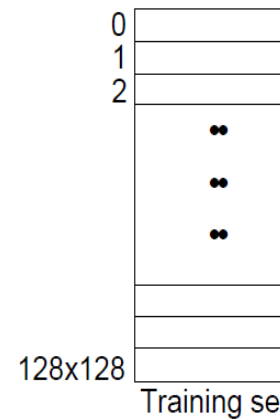
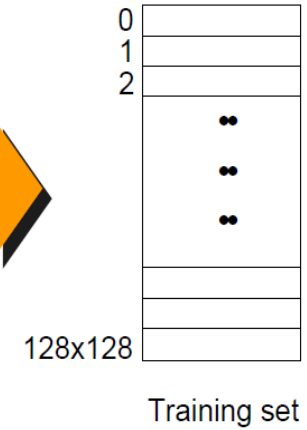
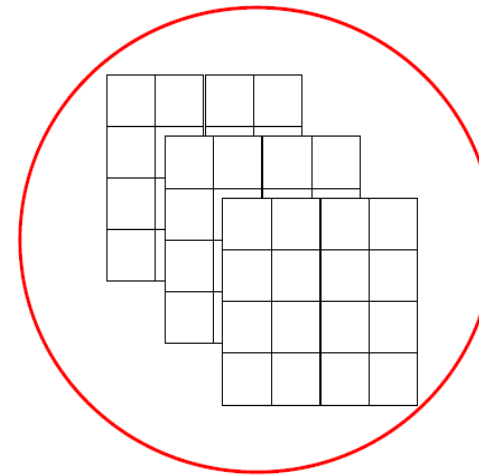
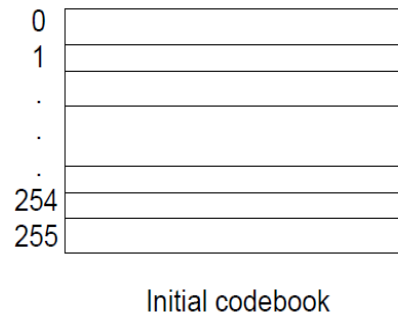
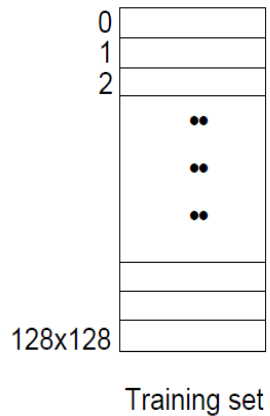


Treinamento do Codebook

- Image size is 4x4
- 128x128 vectors for 512x512 image
- Convenience
- Without loss of generality

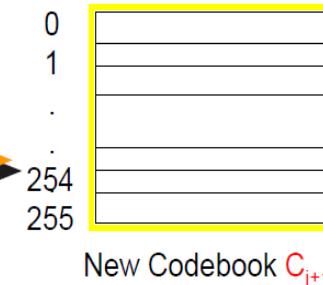


$$d(x, y_i) = \sqrt{\sum_{j=1}^k (x_j - y_{ij})^2}$$



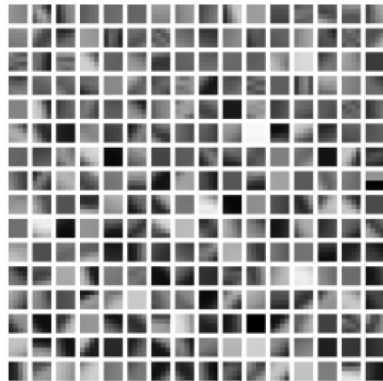
Compute mean values

Replace the old vectors



Exemplo de Geração de Codebook

- A codebook
 - Training set
 - Lena & Boots
- Codebook size
 - 256
- Block size
 - 4x4



- Number of codevectors : N_C
- Input vector dimension: N
- $(\log_2 N_C)/N$ bits/pixel
 - Example: 4×4 blocks, $N_C = 128$, $\log N_C = 7$
 - bit rate = $7/(4 \times 4)$
- Two work in VQ
 - Codebook generation
 - Speedup search



Original



0.5 bpp, 30.2 dB



Original



0.5 bpp, 30.65 dB

Clusterização K-Means início:

- **K-means** (MacQueen, 1967) é um dos algoritmos de aprendizagem não supervisionados mais simples que resolvem este problema de agrupamento;
- O procedimento segue uma maneira simples e fácil de classificar um determinado conjunto de dados por meio de um determinado número de clusters (suponha k clusters) fixo a priori;
- A ideia principal é definir k centróides, um para cada cluster. Esses centróides devem ser colocados de forma adequada por causa de diferentes locais iniciais causam resultados diferentes. Portanto, a melhor escolha é colocá-los como tanto quanto possível longe um do outro;
- O próximo passo é levar cada ponto pertencente a um determinado conjunto de dados e associá-lo ao centróide mais próximo;
- Quando nenhum ponto está pendente, a primeira etapa é concluída e uma primeira clusterização é feita

MACQUEEN, J., *SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS*, Berkeley Symp. on Math. Statist. and Prob., Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press, 1967), 281-297.

K-Means Loop e Parada

- Neste ponto, precisamos recalcular k-novos centróides como baricentros dos clusters resultantes da etapa anterior
- Depois de termos esses k-novos centróides, uma nova iteração deve ser feita entre os mesmos pontos de ajuste de dados e os novos centróides mais próximos;
- Um loop foi gerado.
 - Como resultado deste loop, os k-centróides mudam sua localização passo a passo até que não ocorram mais mudanças são feitas.
 - Quando os centróides não se moverem mais, chegou-se na convergência;
- Este algoritmo visa minimizar uma função objetiva, em geral a função de erro quadrático médio.

Questões para Entregar:

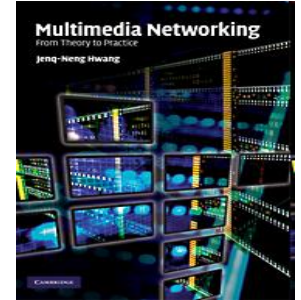
1. Um projetor multimídia de resolução 1920x1080 pixels, instalado a 3.5 metros da tela de projeção resulta numa largura total de 3 metros de imagem projetada na tela. Assinale qual é a altura em metros da imagem projetada na tela, e qual a razão de tiro deste projetor, respectivamente.
2. Considerando que visão humana tem uma resolução espacial de 50 CPD:
 - a) Qual a distância que um projetor multimídia deve ser posicionado numa sala de aula para que a plateia da primeira fila obtenha resolução retina das imagens projetadas?
Dados do projetor: razão de tiro 1.2 e resolução 1280x768 pixels.
Dados da sala: distancia da primeira fila para a tela: 2.5m
 - b) No seu uso diário de um monitor de computador qual tamanho de pixel seria suficiente para você obter resolução retina? Justifique sua resposta através dos cálculos necessários e apresente as considerações de distâncias típicas do seu ambiente de estudo ou trabalho no computador.
3. Explique como é feita uma quantização dos coeficientes DCT-2D em imagem e ilustre graficamente por que a características dos coeficientes da transformada DCT-2D permite a compactação de energia do sinal, concentrando as informações nos coeficientes de baixa frequência.

Questões (cont).

4. Seja um sinal de vídeo que transporta imagens a cores com resolução 1920x1080 pixels, amostradas com taxa de 50 quadros por segundo e 8-bits. Este vídeo é comprimido por um codificador com sub-amostragem 4:2:2, resultando numa taxa média de compressão de 1: 791. Indique qual é a taxa média de transmissão deste video em Mbit/s, antes e depois da compressão, respectivamente:
5. Descreva as duas técnicas abaixo:
 - a. Predição espacial intra-quadro, que analisa blocos de pixels da vizinhança dentro da mesma imagem;
 - b. Predição de compensação de movimento, que analisa blocos de pixels em posições próximas em imagem adjacentes.
6. Pesquise diferenças na escolha dos valores limiares num quantizador escalar.
7. Pesquise como é a implementação de um algoritmo LBG de projeto de codebook num VQ.
8. Cite semelhanças e diferenças entre os algoritmos LBG e K-Means de VQ.

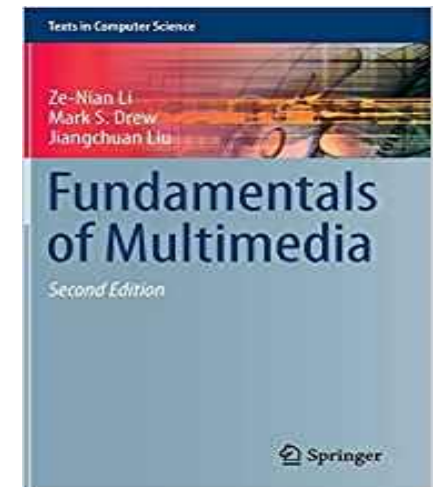
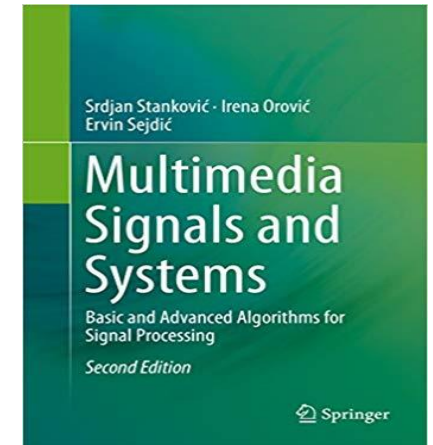
Bibliografia

- HWANG, Jenq-Neng. **Multimedia Networking: From Theory to Practice**, Cambridge University Press, 2009.
- Tokumichi Murakami .The Development and Standardization of Ultra High Definition Video Technology. *In: **High-Quality Visual Experience***. Mrak, Marta; Grgic, Mislav; Kunt, Murat (Eds.). Springer, 2010.



Outras Referências

- STANKOVIC, S.; OROVIC, I.; SEJDIC, E. **Multimedia Signal and Systems**. Springer, 2012.
- LI, Z-N.; DREW, M.; LIU, J. **Fundamentals of Multimedia**. 2. ed. Springer, 2014.



Referencias em Video e Cores

- Andreas Koschan and Mongi Abidi. ***Digital Color Image Processing***. John Wiley & Sons, Inc., 2008. Chapter 3, pp.37-70.
- K. JACK, ***Video Demystified***. 5. ed. Newnes, 2007. Chapter 3, pp.15-36.
- Chen, Janglin; Cranton, Wayne; Fihn, Mark (Eds.) ***Handbook of Visual Display Technology***. Springer, 2012. Section 2.2, pp.131-178.

