

Fundamentos de Imagem Digital

Prof. Alexandre Xavier Falcão

Aula 01

1 Introdução

Atualmente, imagens digitais podem ser capturadas por uma variedade de sensores e o processamento dessas imagens possibilita diversas aplicações: transmissão de vídeo, diagnóstico médico, controle de qualidade de processos industriais, vigilância, biometria, etc.

O processamento de imagem digital consiste de um conjunto de operações matemáticas (algoritmos) aplicadas sobre a imagem para gerar uma nova imagem, representação ou descrição de conteúdo (Figura 1). Essas transformações são necessárias em outras disciplinas, tais como computação gráfica, visualização científica, recuperação de imagens por conteúdo e visão computacional. Em análise de imagens, por exemplo, a descrição do conteúdo é normalmente submetida ao reconhecimento de padrões. Assim, a imagem de um veículo pode ser analisada para fins de reconhecer o número da placa.

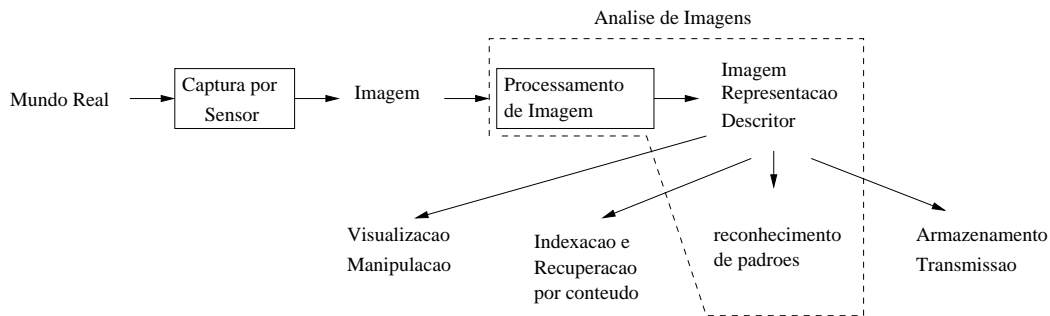


Figura 1: Sistema de processamento de imagens.

Essas disciplinas se unem em uma área maior, que denominamos *Computação Visual*. Essencialmente, elas têm em comum a síntese, manipulação, processamento, e visualização de *objetos gráficos* (imagens ou outros modelos geométricos com propriedades associadas). A Figura 2 ilustra a iteração entre computação gráfica, modelagem geométrica ou estatística, processamento de imagens e visão computacional em um sistema de computação visual. Este sistema mostra como informações visuais do mundo real são processadas no computador para a resolução de problemas.

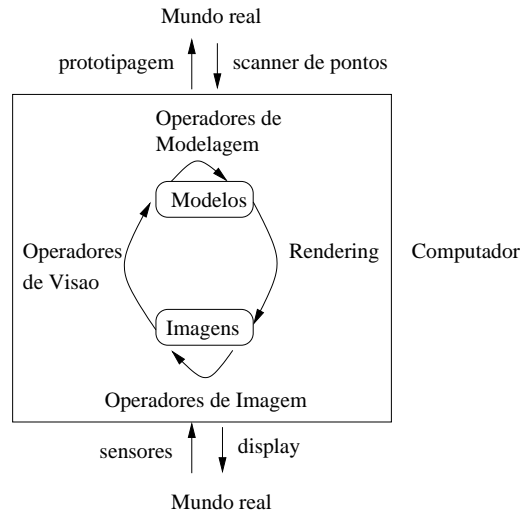


Figura 2: Sistema de computação visual.

2 O que é imagem digital?

O brilho refletido em pontos $(x, y) \in \mathbb{R}^2$ de uma dada região do espaço (cena) pode ser *amostrado* por um arranjo de sensores com espaçamentos (dx, dy) (e.g., $dx = dy = 1mm$) e *quantizado* em valores inteiros $I(x, y)$, formando uma imagem cinza (Figura 3).

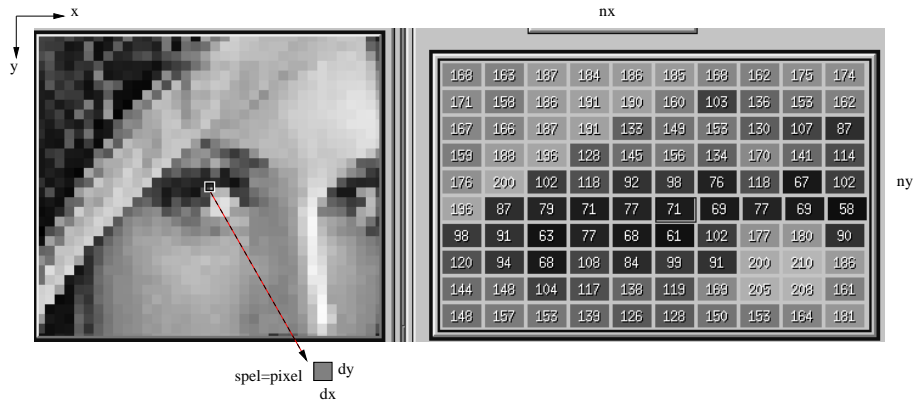


Figura 3: Imagem cinza ($n = 2$ e $k = 1$).

O brilho $I(x, y)$ está normalmente no intervalo $[0, 2^b - 1]$, onde b é o número de bits por pixel (*profundidade* da imagem) necessário para seu armazenamento. Para $b = 8$, por exemplo, a imagem pode armazenar até 256 níveis de cinza, onde o preto corresponde ao 0 e o branco ao 255.

A imagem cinza é vista como um par $\hat{I} = (D_I, I)$, cujo domínio $D_I \subset Z^2$ é definido pelas coordenadas $x = 0, 1, \dots, nx - 1$ e $y = 0, 1, \dots, ny - 1$. Cada elemento $p = (x, y) \in D_I$ é denominado pixel (*picture element*) ou, mais genericamente, spel (*space element*). As coordenadas do mundo real podem ser recuperadas se soubermos a coordenada real de ao

menos um pixel e o vetor (dx, dy) usado na aquisição.

A imagem é dita *tridimensional* se a amostragem cobrir uma região volumétrica. Neste caso, $D_I \subset Z^3$ é definido pelas coordenadas $x = 0, 1, \dots, nx - 1$, $y = 0, 1, \dots, ny - 1$ e $z = 0, 1, \dots, nz - 1$. Este é o caso de imagens médicas, onde a propriedade física quantizada, por exemplo, pode ser a densidade dos tecidos na tomografia de raios-X ou o tempo de relaxação dos spins na ressonância magnética (Figura 4). O spel $p = (x, y, z)$ é denominado voxel (*volume element*) e a profundidade $b = 12$ (e.g., $-1024 \leq I(p) \leq 3071$ na escala Hounsfield da tomografia de raios-X e $0 \leq I(p) \leq 4095$ na ressonância magnética).

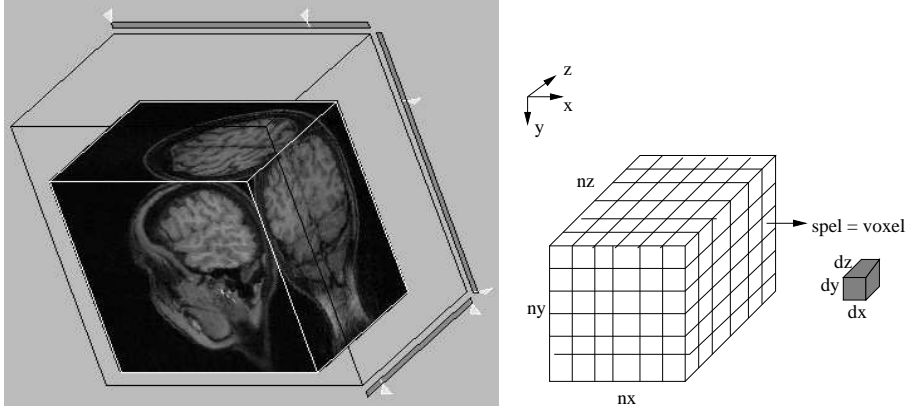


Figura 4: Imagem de ressonância da cabeça ($n = 3$ e $k = 1$).

Genericamente, uma imagem pode ser multi-dimensional e multi-paramétrica (multi-espectral ou multi-bandas) tendo dimensão maior que 2 e número de propriedades físicas quantizadas maior que 1. Portanto, uma *imagem digital* pode ser vista como um par $\hat{I} = (D_I, \vec{I})$, onde $D_I \subset Z^n$ e $\vec{I}(p) = (I_1(p), I_2(p), \dots, I_k(p))$ é um mapeamento vetorial que associa k valores inteiros para cada spel $p \in D_I$. No caso de uma foto colorida, por exemplo, $n = 2$ e $k = 3$, onde $(I_1(p), I_2(p), I_3(p))$ são os valores de luz refletida em p nos comprimentos de onda do vermelho, verde, e azul, respectivamente. A profundidade é $b = 24$ bits por pixel, sendo 8 bits por pixel a profundidade de cada banda. Imagens de satélite com o sensor thematic mapper, por exemplo, têm $n = 2$ e $k = 7$ (azul, verde, vermelho, infra-vermelho próximo, infra-vermelho médio, infra-vermelho termal, e infra-vermelho médio), e cada banda tem profundidade $b = 10$ bits por pixel. Um vídeo digital, no entanto, mistura espaço e tempo para formar 3 dimensões ($n = 3$) e cada quadro possui 3 bandas ($k = 3$), como ilustra a Figura 5.

Muito embora os valores originais dos spels sejam inteiros, as transformações matemáticas podem gerar valores reais e até complexos. Uma transformada de Fourier $\mathcal{F}(\hat{I})$ de uma imagem cinza $\hat{I} = (D_I, I)$, por exemplo, gera uma imagem complexa $\hat{F} = (D_I, \vec{F})$, onde $\vec{F} = (F_1, F_2)$ contém a parte real e a imaginária. Esses valores devem ser armazenados desta forma para preservar a informação. Para fins de visualização da imagem da tela do computador, entretanto, os valores dos pixels devem ser requantizados em valores inteiros compatíveis com a profundidade do monitor.

Observe então que um spel p tem uma coordenada espacial em D_I e uma coordenada paramétrica $(I_1(p), I_2(p), \dots, I_k(p)) \subset Z^k$. Estas coordenadas definem um espaço de parâmetros

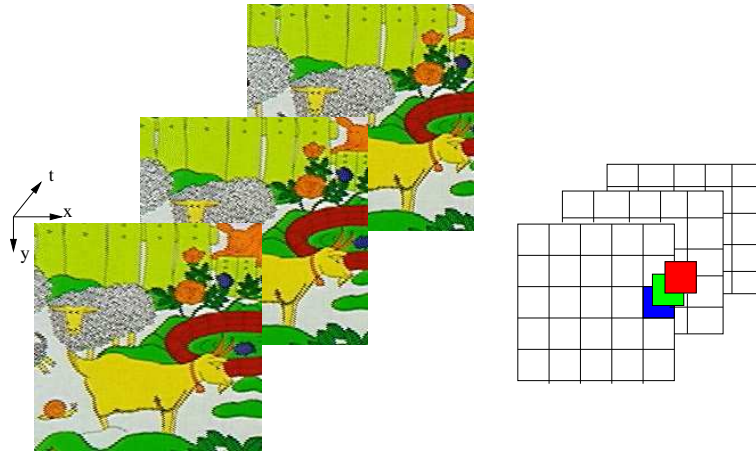


Figura 5: Três quadros de um vídeo digital ($n = 3$ e $k = 3$).

(características ou atributos) da imagem \hat{I} (Figura 6).

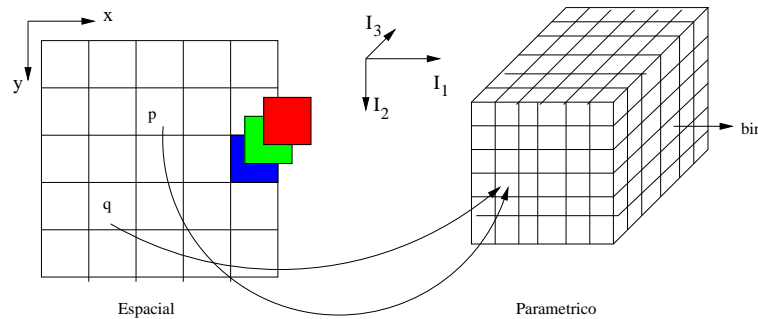


Figura 6: Mapeamento entre as coordenadas espaciais e paramétricas para uma imagem colorida. Pixels de mesma cor caem no mesmo *bin* (similar ao spel) no espaço paramétrico.

3 Resoluções de imagem

A amostragem das propriedades físicas para uma região fixa do espaço e/ou intervalo de tempo pode gerar maior ou menor número de amostras dependendo do espaçamento entre elas. Em um vídeo digital, por exemplo, quanto menor for (dx, dy) e o intervalo de tempo dt entre quadros consecutivos, maior serão as resoluções *espacial* e *temporal*, respectivamente.

A quantização de uma propriedade física também é feita em um intervalo fixo de valores reais. Quanto menor for a diferença entre os valores consecutivos, maior será a profundidade b e, conseqüentemente, maior será a *resolução radiométrica*.

Similarmente, a amostragem de um intervalo fixo espectral pode gerar mais ou menos bandas, dependendo da distância espectral entre bandas consecutivas. Quanto menor for esta distância, maior será a *resolução espectral* e, se a amostragem for uniforme, maior será o número k de bandas. Exemplos de resoluções espaciais e espectrais são mostrados na Figura 7.



Figura 7: (a) 256x256-24bits (b) 64x64 (c) 12bits.

4 Exercícios

1. Como ficaria a definição matemática de uma imagem resultante da transformada de Fourier aplicada a cada banda de uma imagem multi-espectral?
2. Considere a ressonância magnética um pé para diferentes posições em instantes consecutivos de tempo. Como ficaria a definição matemática da imagem resultante?
3. Considere duas imagens cinzas, \hat{A} e \hat{B} , com $|D_A| = 100 \times 200$ e $|D_B| = 200 \times 400$ pixels respectivamente. A primeira cobre uma área de $10 \times 20 \text{ cm}^2$ e a segunda uma área de $100 \times 200 \text{ cm}^2$.
 - (a) Qual imagem possui maior resolução espacial?
 - (b) Qual deveria ser o tamanho da imagem \hat{A} para possuir a mesma resolução espacial da imagem \hat{B} ?
 - (c) Quais são as dimensões em cm^2 dos pixels nas imagens \hat{A} e \hat{B} ?
 - (d) Qual é o comprimento de um segmento de reta em \hat{A} , que vai do pixel $p = (0, 0)$ ao pixel $q = (50, 100)$?
 - (e) Quantos bytes são necessários para armazenar \hat{A} , se os valores dos pixels vão de 0 a 2000? Qual é a profundidade de \hat{A} neste caso?
4. Qual é o tamanho de uma imagem gerada pela amostragem de uma região de $100 \times 200 \text{ cm}^2$ a intervalos $dx = 0,1 \text{ mm}$, $dy = 0,2 \text{ mm}$? Se esta imagem fosse a foto de um objeto, que tipo de distorção você veria? Por quê?
5. Qual a profundidade de uma imagem com 65536 níveis de cinza?