

# Introdução ao Processamento Digital de Imagem MC920 / MO443

Prof. Hélio Pedrini

Instituto de Computação

UNICAMP

<http://www.ic.unicamp.br/~helio>

1º Semestre de 2020

## 1 Fundamentos

- Tipos de Sinais
- Modelo de Imagens
- Digitalização
- Resolução x Profundidade de Imagens
- Representação de Imagens Digitais
- Imagem Multibanda ou Multiespectral
- Imagem Multidimensional
- Sistema Visual Humano
- Ruído e Entropia
- Relacionamento entre Elementos de Imagens
- Métricas de Qualidade de Imagens
- Limites da Imagem

- Conceito de *sinal* varia com respeito ao contexto no qual ele está sendo utilizado.
- De um ponto de vista geral, um sinal é a manifestação de um fenômeno que pode ser expresso de forma quantitativa.
- Um sinal pode ser representado como uma função de uma ou mais variáveis independentes e, tipicamente, contém informação acerca da natureza ou comportamento do fenômeno físico sob consideração. Por exemplo, o sinal de voz pode ser definido por uma função de uma variável (tempo), enquanto o sinal correspondente a uma imagem pode ser definido por uma função de duas variáveis (espaço).

- No domínio temporal, pode-se analisar como as variações do sinal evoluem com o decorrer do tempo.
- Um sinal pode ser *contínuo* ou *discreto*:
  - ▶ Em um sinal contínuo, seus estados podem ser definidos em qualquer instante de tempo, ou seja, sem interrupção.
  - ▶ Um sinal discreto é definido por um conjunto de valores enumeráveis ou inteiros, cujo intervalo depende da natureza do sinal.
- Sinais podem ainda ser classificados como *analógicos* ou *digitais*:
  - ▶ Sinais analógicos podem variar continuamente no tempo.
  - ▶ Um sinal digital pode assumir apenas valores discretos.

### Exemplos:

- a) Uma onda sonora é um exemplo de sinal analógico.
- b) O código Morse, utilizado em telegrafia, é um exemplo de sinal digital.

- Circuitos eletrônicos podem converter um tipo de sinal em outro.

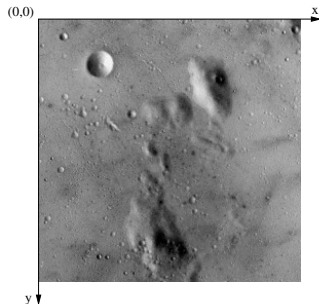
### Exemplos:

- a) Uma onda sonora capturada por um microfone deve ser transformada em sinal digital para manipulação em um computador ou para transmissão como qualquer outro tipo de dado.
- b) Em contrapartida, para reproduzir o som em um alto-falante, o sinal de áudio digital deve ser convertido em sinal analógico.

# Fundamentos

## Modelo de Imagens

- A representação e manipulação de uma imagem em computador requer a definição de um modelo matemático adequado da imagem.
- Uma *imagem* pode ser definida como uma função de intensidade luminosa, denotada  $f(x, y)$ , cujo valor ou amplitude nas coordenadas espaciais  $(x, y)$  fornece a intensidade ou o brilho da imagem naquele ponto.
- Convenção do sistema de coordenadas: a origem está localizada no canto superior esquerdo da imagem.



- Um modelo físico para a intensidade de uma cena sob observação pode ser expresso em termos do produto entre dois componentes, a quantidade de luz incidente na cena e a quantidade de luz refletida pelos objetos presentes na cena.
- Esses componentes são chamados de *iluminância* e *reflectância*, respectivamente, e são representados por  $i(x, y)$  e  $r(x, y)$ .
- Assim, a função  $f(x, y)$  pode ser representada como

$$f(x, y) = i(x, y) r(x, y)$$

para  $0 < i(x, y) < \infty$  e  $0 < r(x, y) < 1$

- A natureza de  $i(x, y)$  é determinada pela fonte de luz, enquanto  $r(x, y)$  é determinada pelas características dos objetos na cena.
- Os valores para os componentes  $i(x, y)$  e  $r(x, y)$  nas equações acima são limites teóricos.

- A iluminância é medida em  $\text{lúmen/m}^2$  ou lux.
- A reflectância é medida em valores percentuais ou no intervalo entre 0 e 1.

### Exemplos: Valores médios ilustram alguns intervalos típicos de $i(x, y)$ :

- Em um dia claro, o Sol pode produzir aproximadamente  $900000 \text{ lúmen/m}^2$  de iluminância na superfície da Terra. Esse valor decresce para menos de  $10000 \text{ lúmen/m}^2$  em um dia nublado.
- O nível de iluminância típico em um escritório é de aproximadamente  $1000 \text{ lúmen/m}^2$ .
- Em uma noite clara, a lua cheia gera aproximadamente  $0.1 \text{ lúmen/m}^2$ .

### Exemplos: Valores típicos de $r(x, y)$ :

- 0.93 para a neve.
- 0.80 para parede branca.
- 0.65 para aço inoxidável.
- 0.01 para veludo preto.



- A maioria das técnicas de análise de imagens é realizada por meio de processamento computacional, então a função  $f(x, y)$  deve ser convertida para a forma discreta.
- Uma *imagem digital* pode ser obtida por um processo denominado *digitalização*, o qual envolve dois passos, a amostragem e a quantização:
  - ▶ A *amostragem* consiste em discretizar o domínio de definição da imagem nas direções  $x$  e  $y$ , gerando uma matriz de  $M \times N$  amostras, respectivamente.
  - ▶ A *quantização* consiste em escolher o número inteiro  $L$  de níveis de cinza (em uma imagem monocromática) permitidos para cada ponto da imagem.

- Cada elemento  $f(x, y)$  dessa matriz de amostras é chamado pixel (acrônimo do inglês *picture element*), com  $0 \leq x \leq M - 1$  e  $0 \leq y \leq N - 1$ .
- A imagem contínua  $f(x, y)$  é aproximada, portanto, por uma matriz de dimensão  $M$  pixels na horizontal e  $N$  pixels na vertical:

$$f(x, y) \approx \begin{bmatrix} f(0, 0) & f(1, 0) & \cdots & f(M - 1, 0) \\ f(0, 1) & f(1, 1) & \cdots & f(M - 1, 1) \\ \vdots & \vdots & \ddots & \vdots \\ f(0, N - 1) & f(1, N - 1) & \cdots & f(M - 1, N - 1) \end{bmatrix}$$

- O conceito de dimensão de um pixel ao longo do eixo  $x$ , ou do eixo  $y$ , está relacionado com o espaçamento físico entre as amostras.
- Cada pixel tem associado um valor  $L_{min} \leq f(x, y) \leq L_{max}$ , tal que o intervalo  $[L_{min}, L_{max}]$  é denominado *escala de cinza*.
- A intensidade  $f$  de uma *imagem monocromática* nas coordenadas  $(x, y)$  é chamada de *nível de cinza* da imagem naquele ponto.
- Uma convenção comum é atribuir a cor preta ao nível de cinza mais escuro (por exemplo, valor 0) e atribuir a cor branca ao nível de cinza mais claro (por exemplo, valor 255).

- A digitalização adequada de uma imagem (ou outro sinal) requer alguns cuidados para que nenhuma informação seja perdida no processo de amostragem.
- Um desses cuidados inclui a escolha correta do espaçamento entre as amostras tomadas da imagem contínua.
- Reformulando o problema de maneira diferente, o objetivo é encontrar condições de amostragem sob as quais a imagem contínua possa ser completamente recuperada a partir de um conjunto de valores amostrados.
- Inicialmente, a análise do problema será realizada por meio de um sinal unidimensional e, posteriormente, estendido para o caso bidimensional.

- Considere um sinal  $f(x, y)$  com banda limitada<sup>1</sup> no domínio  $[-B, B]$  do espaço de frequências, sendo  $B$  um número real.
- A frequência de amostragem,  $F_a$ , é a frequência espacial com que as amostras do sinal são tomadas no processo de amostragem e está relacionada com o intervalo de amostragem  $\Delta x$ , na direção  $x$ :

$$F_a = \frac{1}{\Delta x}$$

- A escolha da frequência de amostragem adequada é determinada pelo *teorema da amostragem de Whittaker-Shannon*, o qual estabelece que um sinal contínuo pode ser completamente reconstruído a partir de um conjunto de amostras, se

$$\Delta x \leq \frac{1}{2B}$$

ou, de forma similar, se

$$F_a \geq 2B$$

---

<sup>1</sup>Um sinal  $f(x)$  com banda limitada possui a propriedade de que sua transformada de Fourier  $F(u)$  possui valores muito baixos para  $u$  fora do intervalo  $[-B, B]$ .

- Em outras palavras, o teorema afirma que pelo menos uma amostra a cada meio período do sinal deve ser tomada para que o sinal possa ser completamente reconstruído ou, de forma similar, que a frequência de amostragem seja no mínimo duas vezes a frequência máxima do sinal a ser amostrado.
- O limite de frequência de amostragem  $\frac{1}{2B}$  é conhecido como *limite de Nyquist*, em homenagem a Harry Nyquist (1928), que demonstrou a importância do limite nas áreas de telefonia e telegrafia.
- Seus experimentos mostraram que não era necessário transmitir o sinal de voz completo para que a conversação fosse compreendida, bastando enviar pequenas amostras do sinal elétrico correspondente à voz, tomadas a intervalos regulares.

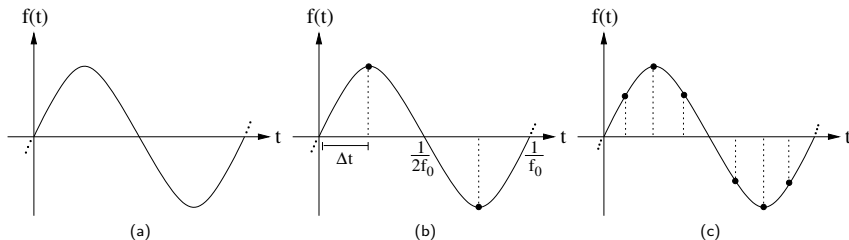
- Se a equação da frequência de amostragem não for satisfeita, um fenômeno denominado *aliasing* ocorrerá, comprometendo a completa recuperação do sinal.
- O fenômeno de *aliasing* pode ser observado por meio de um sinal periódico simples.

### Exemplo:

Seja o sinal  $f(t) = a \sin(2\pi f_0 t)$ , com frequência  $f_0$  e amplitude  $a$  que varia no tempo  $t$ . Um gráfico do sinal  $f(t)$  é mostrado na figura (a) a seguir. Neste caso, a banda limitada de  $f(t)$  é  $[-f_0, f_0]$ . O limite de Nyquist é, portanto, dado por

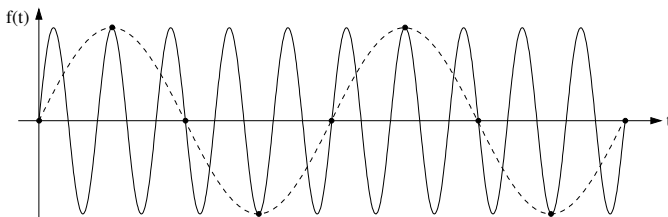
$$\Delta t \leq \frac{1}{2f_0}$$

- Assim, pelo menos uma amostra a cada meio período do sinal deve ser tomada para que o sinal possa ser completamente reconstruído (figura(b)).
- Na figura(c), a taxa de amostragem é quatro vezes superior ao limite de Nyquist. Pode-se notar que as amostras representam corretamente o sinal.





- A figura abaixo mostra um sinal periódico (curva contínua).
- Uma taxa de amostragem inferior ao limite de Nyquist é utilizada para reconstruir o sinal.
- Como resultado, um sinal completamente distinto (mostrado na curva tracejada) do sinal original é obtido.
- Na reconstrução do sinal, obtém-se um sinal de frequência muito inferior à frequência do sinal original.
- As altas frequências do sinal original aparecem como baixas frequências no sinal reconstruído, caracterizando o fenômeno de *aliasing*.



### Exemplo:

Sabe-se que o ser humano é capaz de ouvir sons cujas frequências variam entre 20 Hz e 20 kHz. Portanto, segundo o teorema da amostragem, para que todas as frequências audíveis possam ser registradas, deve-se utilizar uma taxa de amostragem mínima de 40 kHz. Sinais de áudio são tipicamente amostrados à taxa de 44.1 kHz, um pouco superior ao limite de Nyquist para assegurar a recuperação de frequências próximas do limite da audição.

- A extensão do teorema de Whittaker-Shannon pode ser realizada para sinais  $n$ -dimensionais.
- Seja um sinal  $f(x_1, x_2, \dots, x_n)$ . O limite de Nyquist deve ser satisfeito considerando-se cada coordenada  $x_i$ ,  $i = 1, 2, \dots, n$ , ou seja, existe um vetor  $B = (B_1, B_2, \dots, B_n)$  de forma que

$$\Delta x_1 \leq \frac{1}{2B_1}, \dots, \Delta x_n \leq \frac{1}{2B_n}$$

- Para o caso bidimensional, supondo um sinal  $f(x, y)$  com banda limitada  $2W_x$  e  $2W_y$  nas direções  $x$  e  $y$ , respectivamente, o sinal pode ser completamente reconstruído se

$$\Delta x \leq \frac{1}{2W_x} \quad \text{e} \quad \Delta y \leq \frac{1}{2W_y}$$

- É usual em processamento digital de imagens assumir que as dimensões da imagem e o número de níveis de cinza sejam potências inteiras de 2.
- No caso em que o número de níveis de cinza é igual a 2, a imagem é chamada *binária*.
- Imagens binárias possuem grande importância prática, pois ocupam menos espaço de armazenamento e podem ser manipuladas por meio de operadores lógicos que estão disponíveis diretamente nas instruções dos computadores.

- Considerando que o processo de digitalização envolve parâmetros de amostragem e quantização, uma questão é saber quantas amostras  $N \times M$  e níveis de cinza  $L$  são necessários para gerar uma boa imagem digital.
- Isso depende, fundamentalmente, da quantidade de informação contida na imagem e do grau de detalhes dessa informação que é perceptível ao olho humano.
- Tais parâmetros levam aos conceitos de resolução espacial e profundidade da imagem.

### Resolução Espacial

- A *resolução espacial* está associada à densidade de pixels da imagem. Quanto menor o intervalo de amostragem entre os pixels da imagem, ou seja, quanto maior a densidade de pixels em uma imagem, maior será a resolução da imagem.
- É importante notar que uma imagem contendo um grande número de pixels não necessariamente possui resolução maior do que outra contendo menor número de pixels.
- A resolução de uma imagem deve ser escolhida de modo a atender ao grau de detalhes que devem ser discerníveis na imagem.

### Resolução Espacial

#### Exemplo:

Seja, por exemplo, uma imagem  $f(x, y)$  representando uma região de  $400\text{cm}^2$ , consistindo em 20 amostras uniformemente espaçadas na direção  $x$  e 20 amostras uniformemente espaçadas na direção  $y$ .

- Cada pixel da imagem possui dimensão de  $1\text{cm} \times 1\text{cm}$ .
- Uma resolução maior para a mesma região poderia consistir em 40 amostras na direção  $x$  e 40 amostras na direção  $y$ , cada pixel agora correspondendo a  $0.5\text{cm} \times 0.5\text{cm}$ .
- Uma imagem de resolução menor poderia ter 10 amostras na direção  $x$  e 10 amostras na direção  $y$ , em que cada pixel corresponderia a  $2\text{cm} \times 2\text{cm}$ .

### Profundidade da Imagem

- Como mencionado anteriormente, o número de níveis de quantização da imagem  $f(x, y)$  é normalmente uma potência de 2, ou seja,  $L = 2^b$ , em que  $L$  é o número de níveis de cinza da imagem e  $b$  é chamado de *profundidade da imagem*.
- Assim, a profundidade de uma imagem corresponde ao número de bits necessários para armazenar a imagem digitalizada.

#### Exemplo:

Seja  $L = 256$ . Isso significa que cada pixel pode ter associado um valor de cinza entre 0 e 255. A profundidade da imagem, neste caso, é de 8 bits por pixel.



### Profundidade da Imagem

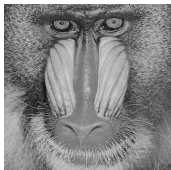
- A tabela a seguir mostra o número de bytes empregados na representação de uma imagem monocromática para diferentes dimensões  $M \times N$  pixels, com 2, 8, 32, 128 e 512 níveis de cinza.

$M$	$N$	Número de bytes				
		$L = 2$	$L = 8$	$L = 32$	$L = 128$	$L = 512$
320	256	10240	30720	51200	71680	92160
480	320	19200	57600	96000	134400	172800
640	400	32000	96000	160000	224000	288000
800	600	60000	180000	300000	420000	540000
1024	720	92160	276480	460800	645120	829440
1280	800	128000	384000	640000	896000	1152000

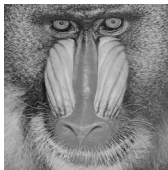
# Fundamentos

## Resolução Espacial

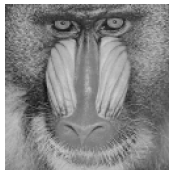
- As figuras (a)-(f) a seguir mostram os resultados da redução da resolução espacial de uma imagem em seis resoluções diferentes.
- Todas as imagens são apresentadas com as mesmas dimensões, ampliando-se o tamanho do pixel de forma a tornar mais evidente a perda de detalhes nas imagens de baixa resolução.



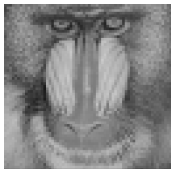
(a)  $512 \times 512$



(b)  $256 \times 256$



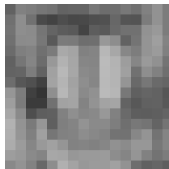
(c)  $128 \times 128$



(d)  $64 \times 64$

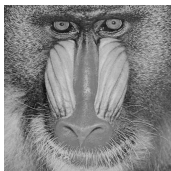


(e)  $32 \times 32$

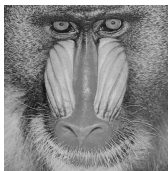


(f)  $16 \times 16$

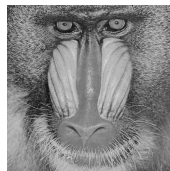
- A figura (a) representa uma imagem de  $512 \times 512$  pixels com 64 níveis de cinza ( $b = 6$ ).
- As figuras (b)-(f) foram obtidas reduzindo-se o número de bits de  $b = 5$  até  $b = 1$  e mantendo as dimensões das imagens com  $512 \times 512$  pixels.



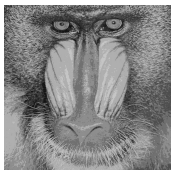
(a) 64



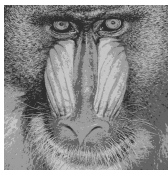
(b) 32



(c) 16



(d) 8



(e) 4



(f) 2

# Fundamentos

## Representação de Imagens Digitais

- Uma imagem digital pode ser representada por meio de uma matriz bidimensional, na qual cada elemento da matriz corresponde a um pixel da imagem.
- A figura a seguir mostra a representação matricial de uma imagem. Uma pequena região da imagem é destacada, sendo formada por números inteiros correspondendo aos níveis de cinza dos pixels da imagem.



(a) representação matricial

120	138	120	151	139
110	129	129	139	146
150	138	137	138	129
137	129	129	128	137
146	145	131	132	145

(b) região da imagem

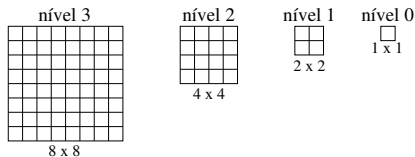
- Há várias vantagens associadas ao uso de matrizes para representar imagens.
- Matrizes são estruturas simples para armazenar, manipular e visualizar dados.
- Uma desvantagem da matriz é sua inerente invariabilidade espacial, já que a estrutura não é adaptativa a eventuais irregularidades que possam existir na imagem. Isso pode produzir uma grande quantidade de redundância de dados.
- Métodos de compressão podem fornecer ganhos significativos em termos de espaço de armazenamento e tempo para transmissão de imagens.

- Imagens podem ser representadas em múltiplas resoluções por meio de representações hierárquicas.
- Uma estrutura muito utilizada é a *pirâmide*. A representação piramidal de uma imagem com  $N \times N$  pixels contém a imagem e  $k$  versões reduzidas da imagem. Normalmente,  $N$  é uma potência de 2 e as outras imagens possuem dimensões  $N/2 \times N/2$ ,  $N/4 \times N/4$ , ...,  $1 \times 1$ .
- Nessa representação, o pixel no nível  $l$  é obtido pela combinação de informação de vários pixels na imagem no nível  $l + 1$ .
- A imagem inteira é representada como um único pixel no nível superior, o nível 0, e o nível inferior é a imagem original (não reduzida). Um pixel em um nível representa informação agregada de vários pixels no nível seguinte.

# Fundamentos

## Representação de Imagens Digitais

- A figura (a) ilustra uma sequência de imagens representadas em diferentes resoluções.
- Uma imagem e suas versões reduzidas obtidas pela média dos valores de cinza em vizinhanças  $2 \times 2$  pixels e dispostas em uma estrutura piramidal são mostradas na figura (b).



(a) sequência de imagens



(b) pirâmide

# Imagem Multibanda ou Multiespectral

- Imagens *multibandas* ou *multiespectrais* podem ser vistas como imagens nas quais cada pixel tem associado um valor vetorial  $f(x, y) = (L_1, L_2, \dots, L_n)$ , em que  $L_{min} \leq L_i \leq L_{max}$  e  $i = 1, 2, \dots, n$ .
- Em geral,  $L_i$  pode representar grandezas diferentes, tais como temperatura, pressão ou frequência, amostradas em pontos  $(x, y)$  e com intervalos de valores distintos.
- Uma imagem multiespectral também pode ser representada como uma sequência de  $n$  imagens monocromáticas, tal que cada imagem é conhecida como *banda*, em que  $f_i(x, y) = L_i$ ,  $i = 1, 2, \dots, n$ .
- Imagens multiespectrais são muito utilizadas em sensoriamento remoto, no qual sensores operam em diferentes faixas do espectro eletromagnético, denominadas bandas espectrais.
  - ▶ Dependendo do alvo, tal como vegetação, água ou solo, a interação da radiação eletromagnética produz menor ou maior resposta espectral, cujo valor está associado à posição espacial de um pixel da banda em particular.

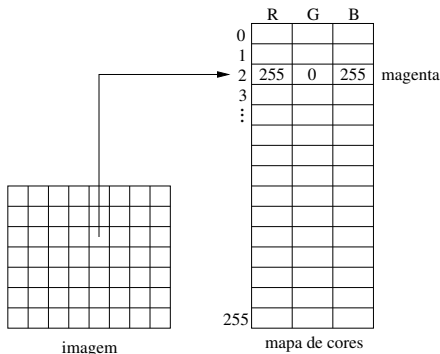


# Imagem Multibanda ou Multiespectral

- Uma imagem colorida é uma imagem multibanda ou multiespectral, em que a cor em cada ponto  $(x, y)$  é definida por meio de três grandezas:
  - ▶ luminância: associada com o brilho da luz.
  - ▶ matiz: comprimento de onda dominante.
  - ▶ saturação: grau de pureza do matiz.
- Uma representação comum para uma imagem colorida utiliza três bandas das cores primárias vermelha ( $R$ ), verde ( $G$ ) e azul ( $B$ ) com profundidade de 1 byte por pixel para cada banda, ou seja, profundidade de 24 bits por pixel.

# Imagem Multibanda ou Multiespectral

- Uma imagem colorida também pode ser armazenada por meio de uma imagem monocromática e um mapa de cores.
- O valor de cinza de cada pixel na imagem torna-se um índice para uma entrada do mapa de cores, enquanto a entrada do mapa de cores contém o valor das componentes  $R$ ,  $G$  e  $B$  referentes à cor do pixel.
- A quantidade de entradas do mapa de cores determina o número de cores utilizadas para representar a imagem.

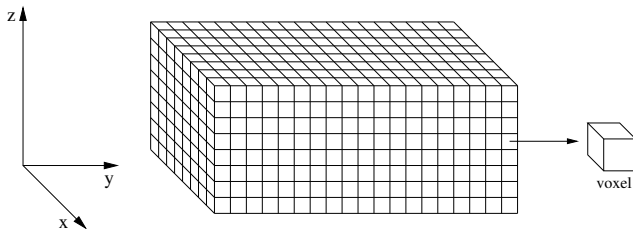


# Imagem Multidimensional

- Há situações em que é necessária uma extensão dos conceitos de amostragem e quantização para uma terceira dimensão, a qual representa, em geral, o espaço ou o tempo.
- Assim, uma imagem digital 3D pode ser representada como uma sequência de imagens monocromáticas ou multibandas ao longo do eixo espacial  $z$  ou do eixo temporal  $t$ , conhecida como *imagem multidimensional*.
- Equipamentos tomográficos geram imagens monocromáticas de cortes (ou fatias) normalmente paralelas e uniformemente espaçadas em uma dada região 3D.
- Considerando as dimensões  $p \times p$  de um pixel nessas imagens e o espaçamento  $d$  entre os cortes, a extensão do pixel em 3D forma um pequeno paralelepípedo de dimensões  $p \times p \times d$ , que é chamado *voxel* (acrônimo do inglês *volume element*).

# Imagem Multidimensional

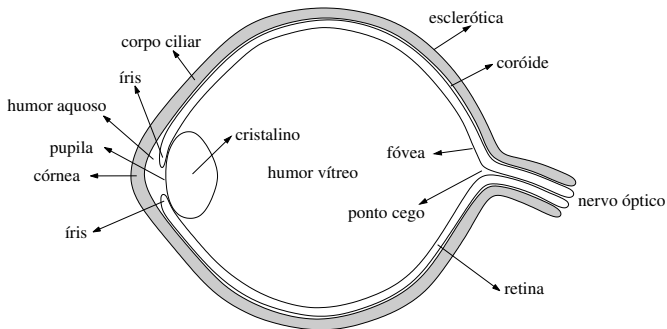
- Os voxels representam pontos de amostragem e são usados para reconstruir no computador a forma ou a função de estruturas tridimensionais.
- Imagens tomográficas possuem tipicamente  $512 \times 512$  ou  $256 \times 256$  pixels e profundidade de 1 ou 2 bytes por pixel.



- Dentre as principais capacidades sensoriais dos seres humanos que permitem adequada percepção do ambiente que os cerca, a visão é uma das mais importantes.
- A visão envolve diversas funções complexas, tais como detecção, localização, reconhecimento e interpretação de objetos no ambiente.
- Uma vez que a área de visão computacional procura dotar as máquinas com capacidades visuais, torna-se fundamental compreender o funcionamento do sistema visual humano sob os diversos aspectos psicofísicos e neurofisiológicos.
- A compreensão do sistema visual humano pode auxiliar o desenvolvimento de sistemas capazes de adquirir, analisar e interpretar informações visuais, com o objetivo de ampliar o número de tarefas que as máquinas podem realizar.

# Sistema Visual Humano

- Corte transversal do olho humano.



- O globo ocular possui formato aproximadamente esférico, com diâmetro de cerca de 20 mm, situado no interior de uma cavidade óssea, chamada órbita, sendo protegido pelas pálpebras e cílios.
- O globo ocular é aderido à órbita pelos músculos extrínsecos, os quais lhe dão capacidade de movimentação.

- O globo é envolvido por três membranas:
  - ▶ uma camada externa formada pela esclerótica e pela córnea.
  - ▶ uma camada intermediária formada pela coróide, íris e corpo ciliar.
  - ▶ uma camada interna formada pela retina.
- A esclerótica é uma camada resistente e opaca que envolve o globo ocular, protegendo as estruturas internas do globo.
- Na parte frontal do globo, localiza-se a córnea, que é um tecido transparente por onde a luz penetra no olho.
- A córnea funciona como uma lente, cujo poder de refração deve permitir que a imagem se forme em uma camada neurosensorial, a retina.
- A coróide está localizada abaixo da esclerótica.
  - ▶ Essa membrana contém uma rede de vasos sanguíneos, os quais nutrem as sensíveis estruturas oculares.
  - ▶ O revestimento da coróide é fortemente pigmentado, ajudando a reduzir a quantidade de luz que entra no olho.

- A íris é responsável por controlar a quantidade de luz que penetra no olho.
- A abertura central da íris é conhecida como pupila, cujo diâmetro varia de 2 a 8 mm.
- A pupila expande ou contrai seu tamanho de acordo com a luminosidade do ambiente, regulando assim a entrada de luz no olho.
- Exatamente atrás da íris está o cristalino, que é uma lente gelatinosa e elástica, cuja função é auxiliar a córnea a focalizar a luz que entra no olho para formar a imagem na retina.
- A distância focal do cristalino é modificada por movimentos de um anel de músculos, os músculos ciliares, permitindo ajustar a visão para objetos próximos ou distantes.



- Imediatamente atrás das lentes, localiza-se a maior câmara do olho, a qual está preenchida por um fluido viscoso chamado humor vítreo, produzido pelo corpo ciliar.
- O humor aquoso é um líquido incolor existente entre a córnea e o cristalino que, juntamente com o humor vítreo, é responsável pela manutenção do volume e da pressão intra-ocular.
- A membrana mais interna do olho é revestida por uma camada de tecidos nervosos, chamada retina.
- A retina é responsável pela sensação da imagem visual projetada pelas estruturas da parte frontal do olho, além de codificar essas informações com sinais nervosos e transmiti-las para o cérebro.
- Cada olho recebe e envia ao cérebro uma imagem, no entanto, os objetos são vistos como um só, devido à capacidade de fusão das imagens.

- A visão binocular (com os dois olhos) proporciona maior campo visual e a noção de profundidade.
- O ponto cego, uma pequena região da retina onde está localizado o nervo óptico, não possui fotorreceptores.
- A retina é composta por células sensíveis à luz, os cones e os bastonetes. Essas células transformam a energia luminosa das imagens em impulsos elétricos que são transmitidos ao cérebro pelo nervo óptico.
- Alterações químicas ocorrem nos bastonetes e cones quando a luz atinge a retina.
- A vitamina A é o composto químico utilizado tanto pelos cones quanto pelos bastonetes para a síntese de substâncias fotossensíveis.

- Cones:

- ▶ São altamente sensíveis à cor e responsáveis pela capacidade do olho em discernir detalhes nas imagens.
- ▶ São em número de 6 a 7 milhões em cada olho e estão localizados na porção central da retina, chamada mácula lútea.
- ▶ Ao centro da mácula lútea está a fóvea, uma região onde as células nervosas estão afastadas para o lado, permitindo que a luz atinja diretamente os receptores.
- ▶ Na fóvea, portanto, a acuidade visual é máxima.

- Bastonetes:

- ▶ O número de bastonetes é muito maior, cerca de 75 a 150 milhões, distribuídos na superfície periférica da retina.
- ▶ Os bastonetes são mais sensíveis à baixa intensidade de luz e permitem uma percepção geral da imagem captada no campo de visão.

- Várias semelhanças podem ser destacadas entre o sistema visual humano e um sistema de sensores, tal como uma câmera fotográfica.
- O obturador da câmera possui função similar à da pálpebra do olho.
- O diafragma de uma câmera controla a quantidade de luz que atravessa as lentes, similar à íris no olho humano.
- As lentes da câmera são análogas ao conjunto formado pelo cristalino e córnea, cujo objetivo comum é focalizar a luz para tornar nítidas as imagens que serão formadas em uma superfície sensível. No olho humano, esta superfície sensível é a retina.
- Nas câmeras fotográficas, películas fotossensíveis ou filmes fotográficos são utilizados para registrar as imagens em câmeras analógicas, enquanto cartões de memória são utilizados em câmeras digitais.

- Imagens reais frequentemente sofrem degradações durante seu processo de aquisição, transmissão ou processamento.
- Essa degradação é normalmente chamada de *ruído*. O ruído pode ser considerado uma variável aleatória  $z$ , caracterizada por uma função densidade de probabilidade  $p(z)$ .
- Os tipos de ruído mais comumente modelados são o ruído:
  - ▶ *impulsivo*.
  - ▶ *Gaussiano*.
  - ▶ *uniforme*.
  - ▶ *Erlang*.
  - ▶ *exponencial*.
  - ▶ *Rayleigh*.
  - ▶ *Poisson*.

- O ruído impulsivo é caracterizado pela ocorrência aleatória de pixels cujos valores de luminosidade diferem significativamente dos valores de seus pixels vizinhos.
- Um tipo de ruído impulsivo em que a imagem é degradada pela ocorrência de pixels brancos e pretos é conhecido como ruído *sal-e-pimenta*.
- A função densidade de probabilidade do ruído impulsivo é dada por

$$p(z) = \begin{cases} P_a, & \text{para } z = a \\ P_b, & \text{para } z = b \\ 0, & \text{caso contrário} \end{cases}$$

- O ruído Gaussiano é caracterizado pela ocorrência de pixels com valores de intensidade que variam conforme a distribuição Gaussiana.
- Esse ruído é uma boa aproximação da degradação que ocorre em muitas aplicações práticas, sendo utilizado, por exemplo, para modelar ruído gerado por componentes eletrônicos de um sistema de aquisição digital de imagens.
- Uma variável aleatória com distribuição Gaussiana possui sua densidade de probabilidade dada pela curva Gaussiana.
- No caso unidimensional, a função densidade de probabilidade é

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z - \mu)^2}{2\sigma^2}}$$

em que  $\mu$  é a média e  $\sigma^2$  é a variância da variável aleatória  $z$ .

- O ruído uniforme segue a função densidade de probabilidade dada por

$$p(z) = \begin{cases} \frac{1}{b-a}, & \text{para } a \leq z \leq b \\ 0, & \text{caso contrário} \end{cases}$$

- A média  $\mu$  e a variância  $\sigma^2$  da função de probabilidade uniforme são definidas como

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$



- A função densidade de probabilidade do ruído Erlang (ou gama) é dada por

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} e^{-az}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases}$$

em que  $a > 0$  e  $b$  é um valor inteiro positivo. A média  $\mu$  e a variância  $\sigma^2$  da função de probabilidade de Erlang são definidas como

$$\mu = \frac{b}{a} \qquad \sigma^2 = \frac{b}{a^2}$$

- O ruído exponencial segue a função densidade de probabilidade dada por

$$p(z) = \begin{cases} a e^{-az}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases}$$

em que  $a > 0$ . A média  $\mu$  e a variância  $\sigma^2$  da função de probabilidade exponencial são definidas como

$$\mu = \frac{1}{a} \qquad \sigma^2 = \frac{1}{a^2}$$

- O ruído exponencial segue a função densidade de probabilidade do ruído Erlang para o caso particular em que  $b = 1$ .

- A função densidade de probabilidade do ruído Rayleigh é dada por

$$p(z) = \begin{cases} \frac{2}{b}(z - a) e^{-(z - a)^2/b}, & \text{para } z \geq a \\ 0, & \text{para } z < a \end{cases}$$

- A média  $\mu$  e a variância  $\sigma^2$  da função densidade de Rayleigh são definidas como

$$\mu = a + \sqrt{\pi b/4} \qquad \sigma^2 = \frac{b(4 - \pi)}{4}$$

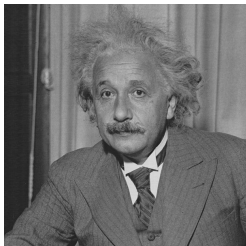
- O ruído Poisson segue a função densidade de probabilidade dada por

$$p(z) = \begin{cases} \frac{e^{-\mu} \mu^z}{z!}, & \text{para } z \geq 0 \\ 0, & \text{para } z < 0 \end{cases}$$

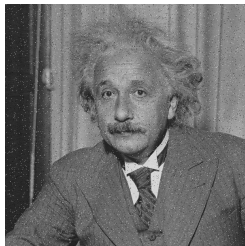
tal que a média da função de probabilidade é dada por  $\mu$  e variância  $\sigma^2 = \mu$ .

## Ruído em Imagens

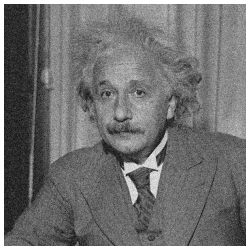
- Exemplos de imagens corrompidas por ruído.



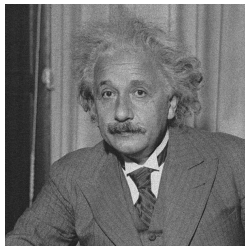
(a) imagem original



(b) com ruído impulsivo



(c) com ruído Gaussiano



(d) com ruído Poisson

# Entropia em Imagens

- O conceito de *entropia* ou *incerteza* foi introduzido por Shannon (1948) para medir a quantidade de informação transferida por um canal ou gerada por uma fonte.
- Quanto maior for o valor de entropia, mais incerteza e, portanto, mais informação estará associada ao canal.
- O princípio fundamental da teoria de informação estabelece que a geração de informação pode ser modelada como um processo probabilístico.
- Uma imagem pode ser considerada como o resultado de um processo aleatório, no qual a probabilidade  $p_i$  corresponde à probabilidade de um pixel em uma imagem digital assumir um valor de intensidade  $i$ ,  $i = 0, 1, \dots, L_{\max}$ .

# Entropia em Imagens

- A distribuição dos níveis de intensidade da imagem pode ser transformada em uma função densidade de probabilidade, dividindo-se o número de pixels de intensidade  $i$ , denotado  $n_i$ , pelo número total  $n$  de pixels na imagem, ou seja

$$p_i = \frac{n_i}{n}$$

em que  $\sum_{i=0}^{L_{\max}} p_i = 1$ .

- A entropia  $H$  de uma imagem pode ser calculada como

$$H = - \sum_{i=0}^{L_{\max}} p_i \log p_i$$

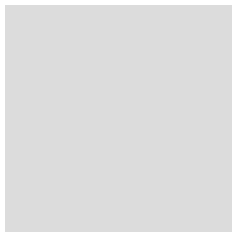
# Entropia em Imagens

- A entropia de uma imagem é uma medida positiva e, quando a base do logaritmo for dois, a unidade resultante é dada em bits.
- O menor valor para a entropia é zero, ocorrendo quando todos os pixels possuem uma mesma intensidade de cinza.
- Por outro lado, a máxima entropia ocorre quando uma imagem contém a mesma quantidade de pixels para todas as intensidades.

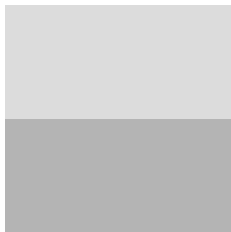


# Entropia em Imagens

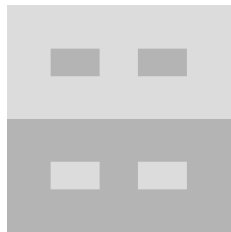
- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



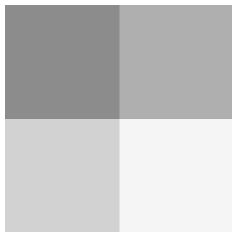
(a)  $H=0$



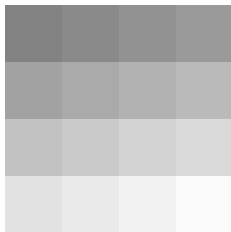
(b)  $H=1$



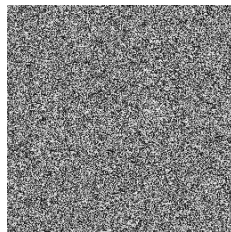
(c)  $H=1$



(d)  $H=2$



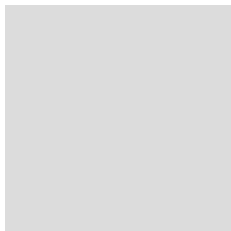
(e)  $H=4$



(f)  $H=8$

# Entropia em Imagens

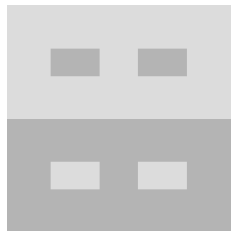
- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



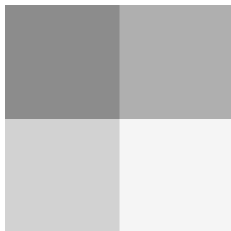
(a)  $H=0$



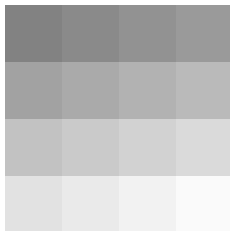
(b)  $H=1$



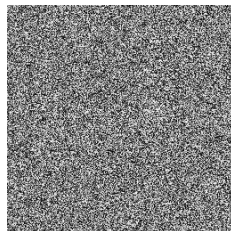
(c)  $H=1$



(d)  $H=2$



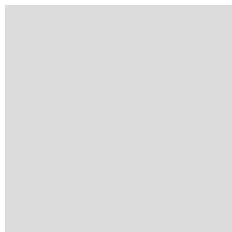
(e)  $H=4$



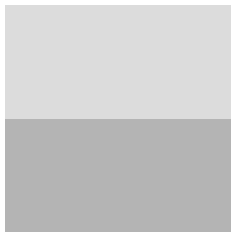
(f)  $H=8$

# Entropia em Imagens

- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



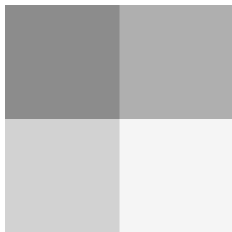
(a)  $H=0$



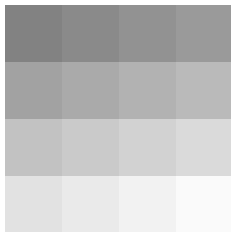
(b)  $H=1$



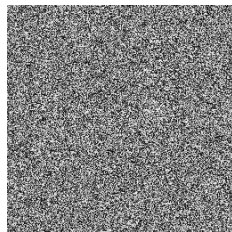
(c)  $H=1$



(d)  $H=2$



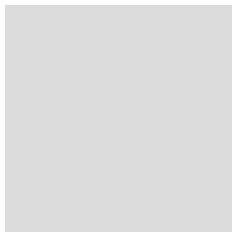
(e)  $H=4$



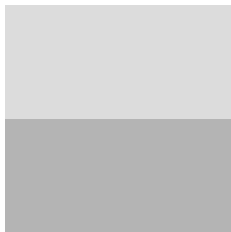
(f)  $H=8$

# Entropia em Imagens

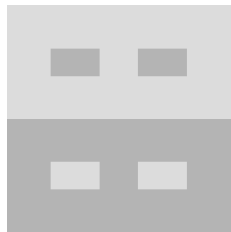
- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



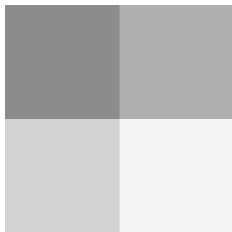
(a)  $H=0$



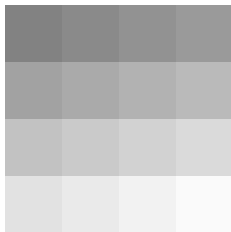
(b)  $H=1$



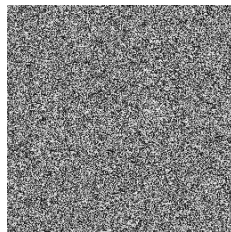
(c)  $H=1$



(d)  $H=2$



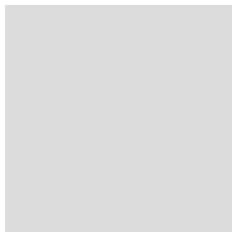
(e)  $H=4$



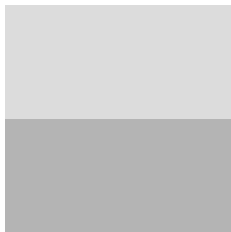
(f)  $H=8$

# Entropia em Imagens

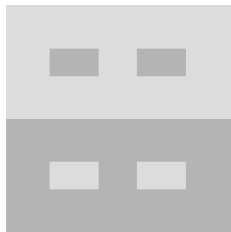
- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



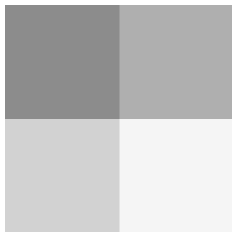
(a)  $H=0$



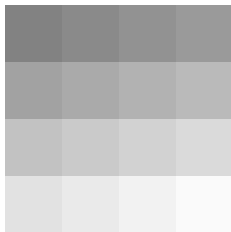
(b)  $H=1$



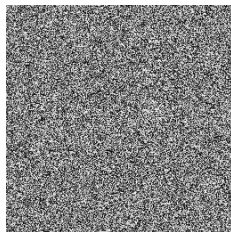
(c)  $H=1$



(d)  $H=2$



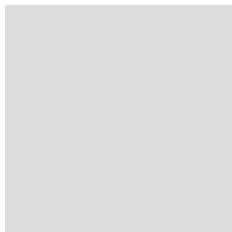
(e)  $H=4$



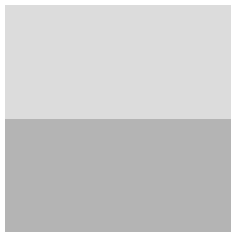
(f)  $H=8$

# Entropia em Imagens

- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



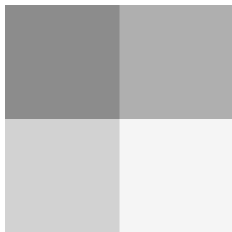
(a)  $H=0$



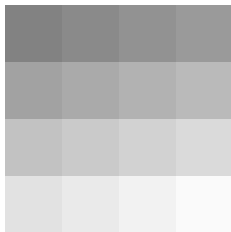
(b)  $H=1$



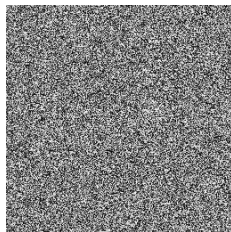
(c)  $H=1$



(d)  $H=2$



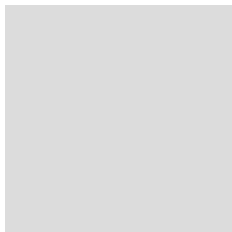
(e)  $H=4$



(f)  $H=8$

# Entropia em Imagens

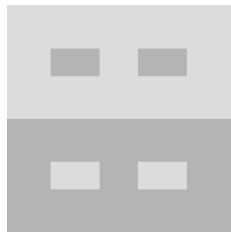
- Exemplos de imagens com dimensões  $256 \times 256$  pixels apresentando diferentes valores de entropia.



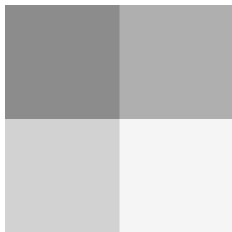
(a)  $H=0$



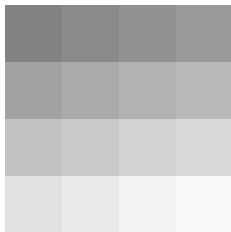
(b)  $H=1$



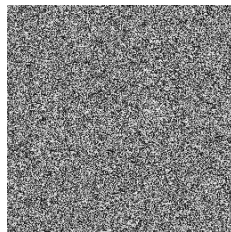
(c)  $H=1$



(d)  $H=2$



(e)  $H=4$



(f)  $H=8$

## Observações:

- Valores mínimo e máximo de entropia são ilustrados nas figuras (a) e (f), respectivamente.
- A imagem da figura (f) possui todos os 256 níveis de cinza possíveis distribuídos com a mesma quantidade de pixels.
- Valores intermediários de entropia são mostrados nas figuras (b) a (e).
- A entropia não está relacionada com a disposição espacial da informação, conforme ilustrado nas figuras (b) e (c), em que as duas imagens possuem igual quantidade de pixels com as mesmas intensidades, porém, distribuídos espacialmente de maneira diferente.

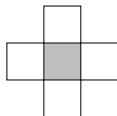


# Relacionamentos Básicos entre Elementos de Imagem

- Um elemento  $f$  em uma matriz bidimensional é denotado pelo pixel  $f(x, y)$ , enquanto em uma matriz tridimensional é denotado pelo voxel  $f(x, y, z)$ .
- Relacionamentos entre elementos:
  - ▶ Vizinhança.
  - ▶ Conectividade.
  - ▶ Adjacência.
  - ▶ Caminho.
  - ▶ Componentes Conexos.
  - ▶ Borda e Interior.

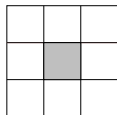
- Vizinhança-4: quatro pixels vizinhos horizontais e verticais do pixel  $f(x, y)$ , cujas coordenadas são:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

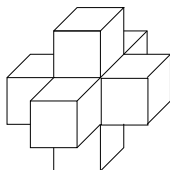


- Vizinhança-8: quatro pixels vizinhos horizontal, verticais e diagonais do pixel  $f(x, y)$ , cujas coordenadas são:

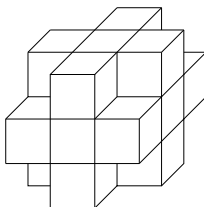
$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1), \\ (x - 1, y - 1), (x - 1, y + 1), (x + 1, y - 1), (x + 1, y + 1)$$



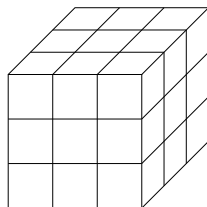
- Extensão do conceito de vizinhança para imagens tridimensionais.
- Vizinhos podem ser definidos de acordo com o número de voxels compartilhando faces, arestas ou vértices em comum.



(a) vizinhança-6



(b) vizinhança-18



(c) vizinhança-26

- Para a vizinhança-6, por exemplo, os vizinhos do voxel  $(x, y, z)$  são:

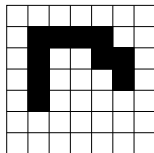
$$(x - 1, y, z), (x + 1, y, z), (x, y - 1, z), (x, y + 1, z), (x, y, z - 1), (x, y, z + 1)$$

- A *conectividade* entre elementos é um conceito importante utilizado para estabelecer limites de objetos e componentes de regiões em uma imagem.
- Para verificar se dois elementos são conexos é necessário determinar se eles são vizinhos segundo o tipo de vizinhança adotado e se os elementos satisfazem determinados critérios de similaridade, tais como intensidade de cinza, cor ou textura.
- Por exemplo, em uma imagem binária, em que os pixels podem assumir os valores 0 ou 1, dois pixels podem ter vizinhança-4, mas somente serão considerados conexos se possuírem o mesmo valor.

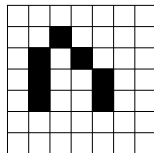
- Um elemento  $f_1$  é *adjacente* a um elemento  $f_2$  se eles forem conexos de acordo com o tipo de vizinhança adotado.
- Dois subconjuntos de pixels,  $S_1$  e  $S_2$ , são adjacentes se pelo menos um elemento em  $S_1$  for adjacente a algum elemento em  $S_2$ .

# Caminho

- Um *caminho* na imagem do pixel  $(x_1, y_1)$  a um pixel  $(x_n, y_n)$  é uma sequência de pixels distintos com coordenadas  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , em que  $n$  é o comprimento do caminho, e  $(x_i, y_i)$  e  $(x_{i+1}, y_{i+1})$  são adjacentes, tal que  $i = 1, 2, \dots, n - 1$ .
- Se a relação de conectividade considerar vizinhança-4, então existe um caminho-4; para vizinhança-8, tem-se um caminho-8.
- Exemplos de caminhos:
  - ▶ o caminho-4 possui comprimento 10
  - ▶ o caminho-8 possui comprimento 7



(a) caminho-4

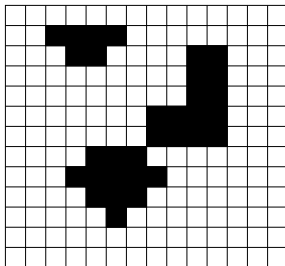


(b) caminho-8

- O conceito de caminho também pode ser estendido para imagens 3D.

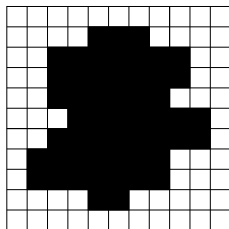
## Componentes Conexos

- Um subconjunto de elementos  $C$  da imagem que são conexos entre si é chamado de *componente conexo*.
- Dois elementos  $f_1$  e  $f_2$  são conexos se existir um caminho de  $f_1$  a  $f_2$  contido em  $C$ .
- Exemplo de imagem bidimensional contendo:
  - ▶ três componentes conexos caso seja considerada a vizinhança-4.
  - ▶ dois componentes conexos se considerada a vizinhança-8.

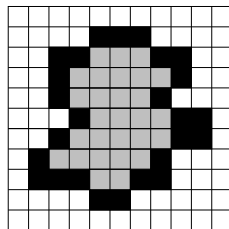


## Borda e Interior

- A *borda* de um componente conexo  $S$  em uma imagem bidimensional é o conjunto de pixels pertencentes ao componente e que possuem vizinhança-4 com um ou mais pixels externos a  $S$ .
- Intuitivamente, a borda corresponde ao conjunto de pontos no contorno do componente conexo.
- O *interior* é o conjunto de pixels de  $S$  que não estão em sua borda.
- Exemplo de uma imagem binária com sua borda e interior.



(a) imagem original



(b) pixels da borda e interior





- Muitas aplicações requerem o cálculo da distância entre dois pixels ou dois componentes de uma imagem.
- Não há uma única forma para se definir distância em imagens digitais.
- Dados os pixels  $f_1$ ,  $f_2$  e  $f_3$ , com coordenadas  $(x_1, y_1)$ ,  $(x_2, y_2)$  e  $(x_3, y_3)$ , respectivamente, qualquer métrica de distância  $D$  deve satisfazer todas as seguintes propriedades:
  - (i)  $D(f_1, f_2) \geq 0$       ( $D(f_1, f_2) = 0$  se, e somente se,  $f_1 = f_2$ )
  - (ii)  $D(f_1, f_2) = D(f_2, f_1)$
  - (iii)  $D(f_1, f_3) \leq D(f_1, f_2) + D(f_2, f_3)$

## Distância Euclidiana

- A *distância Euclidiana* entre  $f_1$  e  $f_2$  é definida como

$$D_E(f_1, f_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- Os pixels com uma distância menor ou igual a algum valor  $d$  formam um disco de raio  $d$  centrado em  $f_1$ .
- A figura a seguir mostra o conjunto formado por pontos com distância  $D_E \leq 3$  de um ponto central  $(x, y)$ .

			3			
	$2\sqrt{2}$	$\sqrt{5}$	2	$\sqrt{5}$	$2\sqrt{2}$	
	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	
3	2	1	0	1	2	3
	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	
	$2\sqrt{2}$	$\sqrt{5}$	2	$\sqrt{5}$	$2\sqrt{2}$	
			3			

- A distância Euclidiana está mais próxima do caso contínuo, entretanto, requer mais esforço computacional e pode produzir valores fracionários.

## Distância $D_4$

- A distância  $D_4$  entre  $f_1$  e  $f_2$ , também denominada *city-block*, é definida como

$$D_4(f_1, f_2) = |x_1 - x_2| + |y_1 - y_2|$$

- Os pixels com uma distância  $D_4$  de  $f_1$  menor ou igual a algum valor  $d$  formam um losango centrado em  $f_1$ .
- Em particular, os pontos com distância 1 são os pixels com vizinhança-4 do ponto central.
- A figura a seguir mostra o conjunto formado por pontos com distância  $D_4 \leq 3$  de um ponto central  $(x, y)$ .



## Distância $D_8$

- A distância  $D_8$  entre  $f_1$  e  $f_2$ , também denominada *chessboard*, é definida como

$$D_8(f_1, f_2) = \max(|x_1 - x_2|, |y_1 - y_2|)$$

- Os pixels com uma distância  $D_8$  de  $f_1$  menor ou igual a algum valor  $d$  formam um quadrado centrado em  $f_1$ .
- Os pontos com distância 1 são os pixels com vizinhança-8 do ponto central.
- A figura a seguir mostra o conjunto formado por pontos com distância  $D_8 \leq 3$  de um ponto central  $(x, y)$ .

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	0	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

## Observações:

- A distância  $D_4$  entre dois pixels  $f_1$  e  $f_2$  é igual ao comprimento do caminho mais curto entre esses pixels, considerando-se a vizinhança-4.
- Do mesmo modo, a distância  $D_8$  corresponde ao caminho-8 mais curto entre esses pontos.

- Operações lógicas e aritméticas podem ser utilizadas para modificar imagens.
- Embora essas operações permitam uma forma simples de processamento, há uma grande variedade de aplicações em que tais operações podem produzir resultados de interesse prático.

- Dadas duas imagens,  $f_1$  e  $f_2$ , as operações aritméticas mais comuns entre dois pixels  $f_1(x, y)$  e  $f_2(x, y)$  são a adição, subtração, multiplicação e divisão, definidas de acordo com a tabela a seguir.

Adição	$f_1(x, y) + f_2(x, y)$
Subtração	$f_1(x, y) - f_2(x, y)$
Multiplicação	$f_1(x, y) \cdot f_2(x, y)$
Divisão	$f_1(x, y) / f_2(x, y)$

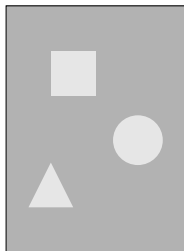
- Como as operações aritméticas podem produzir imagens com valores fora do intervalo de níveis de cinza das imagens originais, alguns cuidados devem ser tomados para contornar essa situação.

- A adição de duas imagens com 256 níveis de cinza, por exemplo, pode resultar em número maior que o valor 255 para determinados pixels. Por outro lado, a subtração de duas imagens pode resultar em valores negativos para alguns pixels.
  - ▶ Uma maneira de resolver esse problema é, após a aplicação do operador aritmético, realizar uma transformação da escala de cinza na imagem resultante para manter seus valores dentro do intervalo adequado.
- A divisão de imagens pode produzir valores fracionários, os quais devem ser convertidos para valores inteiros. Além disso, divisão por zero deve ser evitada.
  - ▶ Uma maneira simples de evitar esse problema é adicionar o valor 1 a todos os valores de intensidade dos pixels, tal que o intervalo de níveis de cinza passa a ser interpretado de 1 a 256, ao invés de 0 a 255.

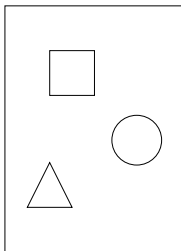


## Operações Aritméticas

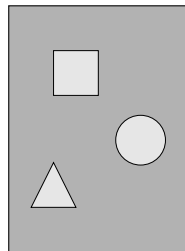
- A adição de imagens pode ser utilizada para sobrepor o conteúdo de uma imagem em outra.
- A figura a seguir mostra a combinação de uma imagem contendo três objetos com o mapa de bordas extraído a partir dos objetos.



(a) imagem original



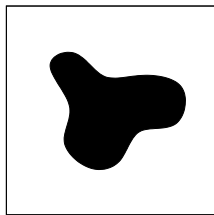
(b) mapa de bordas



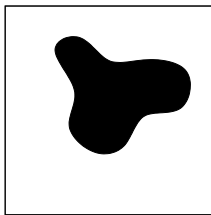
(c) sobreposição do mapa de bordas à imagem original

- Uma outra aplicação do uso da adição é a remoção de ruídos pelo cálculo da média das imagens.

- A subtração de imagens possui vários usos interessantes, sendo uma maneira de identificar diferenças entre imagens.
- A figura a seguir mostra duas imagens e o resultado da subtração entre elas.



(a) imagem  $f_1$



(b) imagem  $f_2$



(c) diferença  $|f_1 - f_2|$

- As regiões dos objetos que permanecem inalteradas nas imagens são eliminadas.
- Os objetos que se moveram são claramente mostrados.

- Uma das principais aplicações da multiplicação ou divisão de imagens é no ajuste de brilho, eventualmente necessário para corrigir problemas que possam surgir durante o processo de aquisição de imagens.
- Outras utilizações desses operadores incluem a filtragem de imagens no domínio de frequência e na modelagem de ruído.

## Operações Lógicas

- As principais operações lógicas utilizadas em processamento de imagens são mostradas na tabela a seguir.

AND	$f_1(x, y) \text{ AND } f_2(x, y)$
OR	$f_1(x, y) \text{ OR } f_2(x, y)$
XOR	$f_1(x, y) \text{ XOR } f_2(x, y)$
NOT	$NOT(f_1(x, y))$

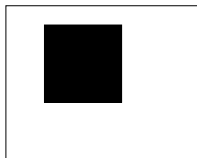
- Essas operações podem ser combinadas para formar expressões lógicas mais complexas.
- Operações lógicas podem ser aplicadas apenas a imagens binárias, enquanto operações aritméticas podem ser usadas em pixels com valores diversos.
- A terminologia adotada é que pixels com valores iguais a 1 (preto) pertencem aos objetos e pixels com valores iguais a 0 (branco) correspondem ao fundo.

# Operações Lógicas

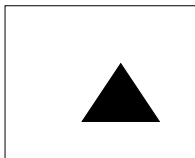
- A operação AND produz o valor 1 na imagem resultante quando os pixels correspondentes nas duas imagens de entrada possuem valor igual a 1.
- A operação XOR produz 1 quando apenas um dos pixels (mas não ambos) possui valor 1, caso contrário, produz 0.
- O resultado da operação OR é 1 quando pelo menos um dos pixels das imagens é igual a 1.
- A operação NOT inverte o valor do pixel na imagem.

# Operações Lógicas

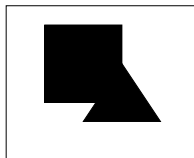
- As operações lógicas podem ser utilizadas para combinar informação entre as imagens ou extrair regiões de interesse.
- Alguns exemplos de aplicação de operadores lógicos são mostrados na figura a seguir.



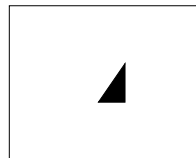
A



B



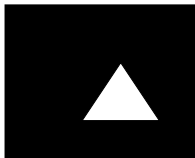
A OR B



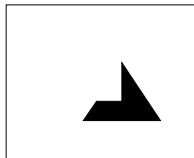
A AND B



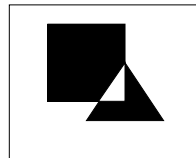
NOT(A)



NOT(B)



NOT(A) AND B



A XOR B

- Além de processamentos pixel a pixel, as operações lógicas e aritméticas podem ser utilizadas em processamentos orientados à vizinhança.
- Tipicamente, esse tipo de processamento utiliza as operações com máscaras, em que os termos janela e filtros são frequentemente utilizados como sinônimos de máscara.

## Operações Aritméticas e Lógicas

- Para a região da imagem mostrada abaixo, a substituição do valor de  $f(x, y)$  pela média aritmética  $\bar{f}(x, y)$  dos valores dos pixels adjacentes em uma vizinhança de  $3 \times 3$  pixels pode ser realizada pela equação:

$$\begin{aligned}\bar{f}(x, y) &= \frac{1}{9} \left[ f(x-1, y-1) + f(x, y-1) + f(x+1, y-1) + f(x-1, y) + \right. \\ &\quad \left. f(x, y) + f(x+1, y) + f(x-1, y+1) + f(x, y+1) + f(x+1, y+1) \right] = \\ &= \frac{1}{9} \sum_{m=-1}^1 \sum_{n=-1}^1 f(x+m, y+n)\end{aligned}$$

		$f(x-1, y-1)$	$f(x, y-1)$	$f(x+1, y-1)$		
		$f(x-1, y)$	$f(x, y)$	$f(x+1, y)$		
		$f(x-1, y+1)$	$f(x, y+1)$	$f(x+1, y+1)$		



- A aplicação de uma máscara em cada pixel da imagem é uma tarefa de alto custo computacional.
- Por exemplo, a aplicação de uma máscara de dimensões  $3 \times 3$  em uma imagem de  $256 \times 256$  pixels requer nove multiplicações e oito adições para cada pixel, resultando em um total de 589 824 multiplicações e 524 288 adições (desconsiderando efeitos de borda da imagem).

# Métricas de Qualidade em Imagens

- A qualidade de uma imagem é bastante dependente da aplicação na qual a imagem é utilizada.
- Imagens podem ser utilizadas, por exemplo, para entretenimento em jogos eletrônicos ou em televisão, assim como em aplicações que requerem grande precisão, tais como na medicina ou automação industrial.
- Nesses exemplos, os tipos e graus de degradação que uma imagem pode sofrer, em geral, são bem distintos.
- Uma imagem pode sofrer degradações durante o processo de aquisição, transmissão ou processamento.
- Métricas de qualidade ou fidelidade podem ser utilizadas para avaliar a similaridade de uma imagem transformada  $g$  em relação à original  $f$ .

# Métricas de Qualidade em Imagens

- Algumas medidas são voltadas a avaliações subjetivas, as quais se baseiam em análises realizadas por observadores humanos.
- Por outro lado, avaliações objetivas procuram medir a qualidade das imagens por meio de funções entre a imagem original e a imagem transformada.
- Os métodos de avaliação objetiva mais comuns são baseados em medidas de similaridade ou diferenças entre as imagens.
- Considerando duas imagens,  $f$  e  $g$ , ambas com dimensões  $M \times N$  pixels, algumas métricas para medir a similaridade entre as imagens são listadas a seguir.

- O erro máximo (ME, do inglês, *Maximum Error*) é a maior diferença absoluta entre cada par de pontos na imagem original e na imagem aproximada, expresso como

$$ME = \max |f(x, y) - g(x, y)|$$

- Quanto menor essa métrica, melhor a nova imagem se aproxima da original.
- O erro máximo, por depender apenas dos valores entre dois pixels nas imagens, é bastante sensível a ruído ou a variações locais nas imagens.

# Erro Médio Absoluto

- O erro médio absoluto (MAE, do inglês, *Mean Absolute Error*) é a soma da diferença absoluta de cada ponto da imagem original e da imagem aproximada, dividido pela multiplicação das dimensões da imagem, expresso como

$$\text{MAE} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |f(x, y) - g(x, y)|$$

- Quanto menor essa métrica, melhor a nova imagem se aproxima da original.

## Erro Médio Quadrático

- O erro médio quadrático (MSE, do inglês, *Mean Square Error*) é a soma do quadrado das diferenças de cada ponto da imagem original e da imagem aproximada, dividido pela multiplicação das dimensões da imagem, expresso como

$$\text{MSE} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2$$

- Quanto menor essa métrica, melhor a nova imagem se aproxima da original.

## Erro Médio Quadrático

- Uma variação muito utilizada dessa métrica é conhecida como raiz do erro médio quadrático (RMSE, do inglês, *Root Mean Square Error*), expressa como

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2}$$

- Outra variação dessa métrica é o erro médio quadrático normalizado (NMSE, do inglês, *Normalized Mean Square Error*), definido como

$$\text{NMSE} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y)]^2}$$

cujos valores variam entre 0 e 1, o que torna possível a avaliação de imagens com dimensões diferentes.

## Relação Sinal-Ruído de Pico

- A relação sinal-ruído de pico (PSNR, do inglês, *Peak Signal to Noise Ratio*) é utilizada para avaliar a diferença global entre duas imagens, expressa como

$$\begin{aligned}\text{PSNR} &= 20 \log_{10} \frac{L_{\max}}{\text{RMSE}} = 10 \log_{10} \frac{(L_{\max})^2}{\text{MSE}} = \\ &= 10 \log_{10} \frac{MNL_{\max}^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2}\end{aligned}$$

sendo  $L_{\max}$  o valor máximo de intensidade de cinza.

- $L_{\max} = 255$  para imagens representadas por 8 bits de profundidade.
- A métrica PSNR é expressa em *decibel* (dB), unidade originalmente definida para medir intensidade sonora em escala logarítmica.
- Valores típicos de PSNR variam entre 20 (para RMSE = 25.5) e 40 (para RMSE = 2.55).
- Quanto maior essa métrica, melhor a nova imagem se aproxima da original.



- Uma métrica relacionada é a relação sinal-ruído (SNR, do inglês, *Signal to Noise Ratio*), definida como

$$\text{SNR} = 10 \log_{10} \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y)]^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - g(x, y)]^2}$$

- A covariância entre duas imagens  $f$  e  $g$  é dada por

$$\sigma_{fg} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f][g(x, y) - \mu_g]}{MN}$$

em que  $\mu_f$  e  $\mu_g$  representam o nível de cinza médio nas imagens  $f$  e  $g$ , respectivamente.

## Coeficiente de Correlação

- O coeficiente de correlação entre duas imagens  $f$  e  $g$  é dado por

$$r = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f][g(x, y) - \mu_g]}{\sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - \mu_f]^2 \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [g(x, y) - \mu_g]^2}}$$

- Observando-se a equação acima, a correlação pode ser vista como a covariância dividida por um fator que depende da distribuição dos níveis de cinza (variância da imagem) de cada uma das imagens.
- O valor de  $r$  varia entre  $-1$  a  $1$ , sendo que valores próximos de zero representam um relacionamento linear mais fraco entre as duas imagens.

## Coeficiente de Jaccard

- O coeficiente de Jaccard entre duas imagens  $f$  e  $g$  é dado por

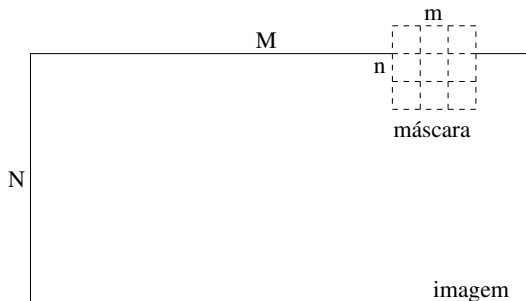
$$J = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \begin{cases} = 1, & \text{se } f(x, y) = g(x, y) \\ = 0, & \text{caso contrário} \end{cases}}{MN}$$

em que a igualdade  $f(x, y) = g(x, y)$  é obtida considerando-se um determinado valor de tolerância  $\delta$ , ou seja,  $|f - g| \leq \delta$ .

- O coeficiente de Jaccard é igual a 0 para duas imagens que não apresentam nenhuma similaridade.
- Duas imagens que apresentam todos os elementos idênticos possuem coeficiente de Jaccard igual a 1.

## Limites da Imagem

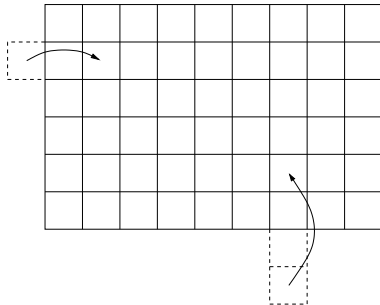
- Durante algumas operações, como a filtragem espacial descrita anteriormente, parte da máscara pode se localizar fora das dimensões  $M \times N$  da imagem, como ilustrado na figura abaixo.



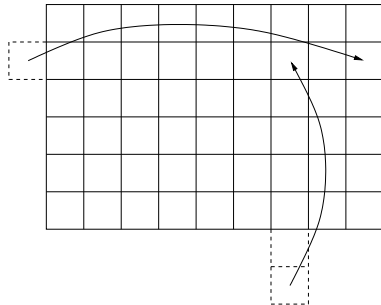
- Quando isso ocorrer, a operação requer cuidados especiais.

- Há várias estratégias diferentes para tratar desse problema:
  - ▶ Simplesmente ignorar os pixels para os casos em que a operação em questão não possa ser realizada.
  - ▶ Copiar o valor do pixel correspondente da imagem original, caso não seja possível realizar a operação. A imagem terá uma borda de pixels não processados.
  - ▶ Utilizar uma máscara modificada para realizar a operação, tratando os pixels da borda de maneira especial. Isso pode aumentar consideravelmente a complexidade da operação.
  - ▶ Verificar se o pixel possui coordenadas dentro das dimensões da imagem. Se não possuir, as coordenadas podem ser refletidas na imagem original ou repetidas de forma circular.

# Limites da Imagem



(a) indexação por reflexão



(b) indexação circular