

Nome: João Vitor Couto Oliveira

Primeiro Bimestre:

## Aula 1

Embora o ser humano não tenha dificuldades com enxergar coisas e saber a diferença entre elas, para os computadores é extremamente difícil de se diferenciar, fazendo com que seja necessário transformar imagens em dados utilizando uma série de processos que envolvem visão computacional, síntese de imagens, processamento de dados e, finalmente, processamento de imagens.

Historicamente imagens eram transportadas por meio de cabos contendo a informação matricial da imagem, tinham poucas tonalidades e baixa resolução.

O olho humano é capaz de observar somente um curto segmento de todo aspecto de cores existentes

O processamento digital de imagens possui muitas aplicações, dentre elas análise de imagens de raios-x, imagens com ampliação microscópica até imagens de satélite. Esse processamento leva em consideração vários aspectos de uma imagem, como o funcionamento, compressão, realce, entre outros, que serão descritos a seguir.

**Definição:** Uma imagem é definida por uma função  $f(x, y)$ , que possui uma amplitude, ou seja, é um plano em que seu relevo é descrito pela amplitude. Se os valores de  $x$ ,  $y$  e a amplitude são finitos e podem ser descritos eles são digitais, ou seja, se existe uma distinção clara entre segmentos atômicos. O processamento de imagens é denominado por uma transformação ( $t$ ) aplicado sobre a imagem  $f(x, y)$ , gerando uma imagem final representada por uma  $g(x, y)$ .

**Realce de Contraste:** A ideia geral do realce de contraste é de uma redistribuição da amplitude da imagem, em que se aplicado em um histograma (que descreve quantos pixels são relativos as tonalidades descritas) é possível aumentar de onde o pixel mais claro sai do limite máximo da imagem para um limite máximo de representação, por exemplo, uma imagem possui brilho máximo 90 enquanto o dispositivo que exibe a imagem possui um limite de 255, a ideia é de transferir o 90 para o 255, fazendo com que o histograma fique mais suave e distribuído.

**Correção de iluminação irregular:** A ideia geral é de borrar a imagem em algumas sessões menores e depois pegar o valor geral dos pixels e dividi-los pelos valores dos pixels da imagem original e depois multiplica-los pelo valor máximo de representação, afim de que as regiões mais claras, que possuem média menor, sejam divididas por valores médios ainda menores e assim fiquem mais escuras e as regiões mais escuras sejam divididos por valores médios maiores para que fiquem mais claras.

**Redução de ruído aleatório:** A ideia geral é de fazer uma média de todos os valores para que os pixels destoantes fiquem mais fracos.

Redução de ruído periódico: É feito a transformada de Fourier.

**Analisar de imagens:** é feita pela extração de informações quantitativas, fazendo medidas, desse modo sendo possível fazer várias medidas com precisão altíssima com ajuda de um computador. A visão computacional tem como objetivo criar modelos e automatizar o processo de reconhecimento visual.

A sequência de processamento e análise é dada por: Captura da imagem, digitalização, pré-processamento (limpeza e filtragem), segmentação (divisão em regiões), pós-processamento (realce), extração de atributos e a classificação e reconhecimento.

## Aula 2

O olho humano é uma ferramenta de extrema importância para nós, embora sua precisão seja altíssima, existem algumas coisas que ele não é tão bom, como por exemplo a precisão nos tons de cinza, em que são poucos comparados com outras cores. Por isso em processamento de imagens podemos reduzir a quantidade de informação a ser processada pois o olho humano mal percebe a diferença, é estimado que o olho humano consiga perceber a diferença entre 32 a 64 tons de uma cor.

Equipamentos de captura de luz são normalmente matriciais e são capaz de transformar energia da luz em sinais elétricos, mas essa transformação é dependente de dois fatores: a amostragem, que é referente ao espaço; e a quantização, que é referente a intensidade. Se pensarmos de modo gráfico, ao representar o sinal elétrico recebido pelo sensor, podemos mais facilmente observar que um sinal analógico normalmente é formado por senoides, e se formos tentar inserir a senoide com precisão analógica seria

precisão infinita, o que digitalmente é impossível, portanto, fazemos uma redução por meio da amostragem, em que são escolhidos alguns pontos específicos para serem medidos, normalmente representados pelo eixo  $x$ . Já a quantização é a quantidade de fatores possíveis no eixo  $y$ , ou seja, o eixo  $y$  é subdividido em algumas partes, de modo que juntamente com as amostras do eixo  $x$  é formada uma “escada”. Isso aplicado em imagens cria o que chamamos de resolução, ou a quantidade de *dpi* (*dots per inch*), em que uma imagem for capturada e quantificada em um número limitado de pixels, cores e suas intensidades.

Esse processo pode criar um tipo de imagem que não é interessante, que são as “escadas” nas linhas dos objetos, que são problemas de amostragem. O teorema de Nyquist diz que “uma taxa de amostragem de no mínimo duas vezes o valor da frequência máxima alcançada pelo sinal analógico é necessário para possibilitar o registro digital de todas as frequências analisadas”. Por isso, por exemplo, nos computadores Windows, é normal que os aparelhos de áudio possuam uma frequência de amostragem de 44KHz, que é o dobro do que é absorvido pela maioria dos seres humanos, já a sua quantização fica mais próximo do que é absorvido, com 16 bits, ou seja, mais de 65 mil níveis de representação. Em imagens a amostragem é conhecida como a resolução da imagem (1920 x 1080), e a quantização é a quantidade de cores representadas, essas dependem da quantidade de detalhes existentes na imagem.

Afim de ampliar a quantidade de amostragem podem ser feitos alguns processos com a imagem, em que o processo de ampliação tenta aumentar a resolução, ou seja, ampliar o tamanho da matriz de representação, esse método é chamado de interpolação de imagem. Podem ser usados os seguintes métodos:

- O vizinho mais próximo, onde só é replicado o vizinho mais próximo nos pixels novos.
- A interpolação bilinear onde é feita a média dos valores de intensidade mais próximos, em que é usado o vizinho da direita, da esquerda, de cima e de baixo.
- A interpolação bicúbica é como a bilinear porém usa 16 vizinhos mais próximos ao invés de 4.

Esses métodos usam um conceito de vizinhança que pode variar dependendo do caso, onde podem ser os vizinhos: de cima, baixo e dos lados; ou pode ser os oito vizinhos (incluindo as diagonais); ou somente as diagonais. Essas conexões podem ser representadas de algumas maneiras, de modo que para consumir o mínimo de espaço possível, é feito um método de escolha de caminho, onde o pixel que não é vizinho e possui um caminho por meio de outros pixels para ser representado é usado.

É possível realizar operações nesses elementos de imagem afim de melhorar sua representação, por exemplo, se existem um conjunto de imagens com ruídos aleatórios, é possível nivelá-los ou até removê-los a partir da média dos valores de intensidade. Aplicação frequente da subtração de imagens é no cálculo das diferenças em imagens, revelando pra o olho humano até a mais sutil diferença. Com a multiplicação é possível fazer uma correção de sombreamento, para melhorar uma imagem; ou até aplicar mascaras de áreas de interesse. De mesmo modo também é possível aplicar operações lógicas e de conjuntos.

Operações de transformações espaciais geométricas são as mais conhecidas, pois são as transformações de rotação, escala, translação (ou mudar a imagem de lugar) e esticar. Todas operações aplicadas em matrizes para gerar tais resultados. Além desses, podem ser feitos processos de expansão e contração de regiões da imagem ou até ajustes de brilho ou cor, em que um dos exemplos mais conhecidos é o *cross-dissolve*, que é capaz de mesclar uma imagem em outra por meio de filtros de amplitude. Um caso onde foram usados ambos os exemplos acima foi quando a uma, ao mostrar seu poder de processamento fez um vídeo em que uma imagem de uma mulher é transformada em onça por meio dos métodos de *cross-dissolve* e de *morphing*.

Operações sobre pixels individuais são aquelas operações que aplicam uma função em um pixel e gera um pixel de saída. Algumas das operações feitas fazem esse processo, como as operações de vizinhança, que calculam um pixel de saída baseado nos pixels vizinhos.

Também existem métodos que calculam probabilidade baseado em imagens. Um exemplo é a probabilidade da cor, em que é calculado baseado na quantidade de pixels que possuem certa cor em uma imagem. O desvio

padrão da imagem também pode ser usado para dizer o contraste da imagem, onde quanto menos desvio, menor o contraste.

Também pode se verificar a presença de objetos com conjuntos de pontos conectados entre si.

### Aula 3

**Cores:** Cores dependem do seu comprimento de onda, que pode ser representado pela combinação das cores primárias, tanto pela adição de cores quanto pela subtração das cores, utilizados em displays e impressoras, respectivamente.

Nos dias de hoje os displays possuem uma memória específica para cada pixel, que armazena 8 bits para cada cor primária (Verde, vermelho e azul), gerando uma combinação total de 16 Mega cores, chamado sistema *True Color*. Cores também são distinguidas em 3 características principais:

- **Brilho:** define a intensidade da cor, que é independente do grau de pureza;
- **Matiz:** define o comprimento de onda dominante;
- **Saturação:** define a “pureza” de uma cor, e junta com a matriz compõe a cromaticidade, em que uma cor pode ser caracterizada por seu brilho e cromaticidade. Pode ser definido pela proporção da quantidade de cor em relação à cor cinza média, em que quanto menos cinza na composição da cor, mais saturada ela é (quanto maior a saturação mais pura é a cor).

Matematicamente é utilizado um esquema de cores que utiliza desses 3 conceitos para formar um sistema de cores chamado HSI (*hue* (matiz), *saturation* (saturação) e *intensity* (brilho)).

**Modelos de cores:** sempre que se tem  $R=G=B$ , tem-se um tom de cinza. Utilizando um sistema RGB aditivo, é possível representar todas as cores visíveis por meio de um cubo, colocando cada cor sob seu eixo. É possível fazer a conversão dos tipos de cores, onde para transformar:

- **RGB em CMY:** faz o máximo de intensidade (normalmente 255) menos a intensidade da cor;
- **RGB em YIQ:** multiplica-se as cores por uma matriz A, pré-definida que armazena as constantes de relação em formato HSI.

Em questão de precisão, o modelo mais fiel seria o HSI, pois ele desacopla a informação de brilho da informação de cores (matiz e saturação). Esse sistema utiliza como objeto de representação um cone, onde quanto mais próximo da base do cone mais claro e quanto mais próximo do pico do cone mais escuro, e o ângulo em que a reta de raio do círculo da base é traçada.

### Aula 5

Realce de Imagens:

Os realces aplicados em imagens nem sempre são relacionados ao melhoramento para a visualização humana. O processamento feito pode ser para que o computador processe uma imagem de forma mais precisa, onde o realce é feito com a computação em mente.

Esse processamento pode ter operações de dois tipos:

- **Operação pontual:** altera dados pixel a pixel, onde tem um pixel de entrada e baseado somente nem gera um único pixel de saída, também podem ser usados várias imagens, porém sempre calculando baseado no mesmo pixel;
- **Operação baseada em vizinhança:** para gerar um pixel de saída utiliza de pixels vizinhos ao pixel base da operação, ou seja, uma área de pixels.

Operações pontuais aplicam uma função T sobre um pixel de coordenada (x, y). Um exemplo de uma dessas funções é no caso de uma distribuição uniforme de intensidade, onde ela é reduzida em alguns pontos e incrementada em outros, tendo um resultante pra cada valor.

Outro exemplo é uma função T de binarização um grupo de intensidades é transformado em intensidade 0 e a partir de um certo ponto se torna o valor máximo até a intensidade de entrada máxima. A função T também pode inverter as cores de entrada. Funções de potência e logarítmicas são as mais utilizadas e funcionam como foi descrito na distribuição uniforme.

Também existem funções que fazem um processo chamado fatiamento de tons de cinza, que podem fazer uma imagem ter somente alguns tons de cinza, que dependendo da intensidade se encaixa em um determinado tom de

cinza, ou também pode manter a intensidade normal em uma sessão e em outra fica em alguns tons específicos.

Fatiamento de planos de bits: como cada pixel possui um conjunto de bits para representar sua intensidade, podemos dividir a imagem bit a bit, onde pode se desenhar uma imagem binária baseado em cada bit da imagem.

O histograma é uma ferramenta muito utilizada em PDI, onde é um gráfico que representa a quantidade de cada intensidade, em que ele plota uma barra pra cada intensidade dos pixels da imagem. Pode ser utilizado para demonstrar contraste.

É possível aplicar funções T para melhorar o histograma, um exemplo é uma equalização de histograma, ela aplica uma média entre as intensidades dos pixels com intuito de aumentar o contraste da imagem. Também é possível fazer equalizações de histograma de forma especificada, onde a área de distribuição do histograma é focada, e também equalização local, onde a imagem é dividida em sessões e daí é equalizada, essa equalização também é aplicada baseado em vizinhança, em que a média é feita baseada em uma vizinhança do pixel a ser calculado.

O processamento local faz um realce baseado nas características do pixel, sendo que, se estiver em um certo intervalo, ele faz um realce, senão permanece igual. Também está relacionado com o uso de máscara, onde ela é capaz de escolher sessões da imagem.

A subtração de imagens é muito utilizada na medicina para encontrar uma diferença entre imagens, em que existindo duas imagens capturadas do mesmo lugar a diferença entre as duas depois dessa operação é só o que sobra.

Um processo de redução de ruído pontual pode ser feito baseado em algumas imagens em que, ao fazer a média de cada pixel se obtém uma imagem em que o ruído de cada uma perde sua intensidade cada vez mais, gerando uma nova imagem que possui menos ruído.

Operações baseadas em vizinhança são mais complexas, porém mais efetivas. Dentre os processos existem filtragens, dentre elas, a filtragem espacial, que utiliza máscaras de convolução para obter seu resultado, denotado por  $f(x) * g(x)$ , a máscara de convolução obtém um gráfico referente aos valores de área da interseção entre as duas funções. Em imagens o processo é caracterizado por duas matrizes, uma imagem e um filtro, onde a matriz de filtro é menor que a matriz de imagem. Ela é multiplicada em uma vizinhança de pixels referente ao tamanho da máscara, e depois que todos os valores são multiplicados sua soma é inserida (depois de normalizada) para um valor da matriz de saída, isso para todos os índices da matriz de saída.

Na prática a filtragem espacial, pode fazer um processo de filtro passa-baixa, em que a matriz máscara faz um filtro de média, que suaviza a imagem. Mas acaba borrando, pois além de suavizar ruídos, ela suaviza as bordas.

Normalmente utilizada para eliminar ruídos, quanto maior a matriz mascara, mais borrada fica a imagem final. Também pode ser usado o filtro da mediana, em que ele organiza os valores referentes aos pixels abaixo da matriz mascara, pega o valor central e o insere na imagem de saída. Esse método consegue reduzir ruídos de forma mais eficiente ainda mantém as bordas.

A suavização de imagens ainda tem utilização. Em que pode ser utilizado para obter os elementos mais claros de uma imagem com muitos ruídos, como uma imagem de estrelas. Para obter a localização das estrelas de brilho mais intenso, pode ser feito esse processo, juntamente com um processo de binarização ou limiarização, que fazem com que, a partir de um certo ponto, um certo conjunto de tonalidades seja descartado, realçando mais ainda as partes mais claras da imagem. Essas funções podem ser aplicadas de mesmo modo para verificar áreas de objetos, em pixels, mas a intenção permanece, essas alterações de redução de resolução são voltadas para o melhor processamento do computador, em que quanto mais detalhes, mais complexos ficam os cálculos.

Existem processos, como a passa alta, que preserva os valores que tem alta frequência, ou seja, um conjunto de pixels que tem uma diferença muito grande entre elas, que são as bordas e detalhes da imagem. Esse método obtém primeiramente os valores de baixa frequência, por meio da média (que é uma operação baseada em vizinhança) e subtrai a original pela matriz de baixa (que é uma operação pontual), obtendo uma imagem contendo somente as áreas de alta frequência da imagem, ou os detalhes distintos do padrão.

**Detecção de bordas:** Muito utilizado em PDI, esse processo reduz a quantidade de informação da imagem, tanto facilitando a computação de um elemento quanto tornando mais simples de ser enxergado. O processo para se obter é simples, em tese, em que é desenhado uma linha naqueles pixels que tem uma distinção muito grande de intensidade do seu vizinho, ou seja, uma região de alta frequência, representando uma borda.

**Filtro de realce de bordas:** Esses filtros tem como base encontrar a diferença entre os pixels, a ideia é aplicar uma derivada e, por definição uma derivada se trata de uma subtração de dois pontos de funções em um intervalo h, porém quando se trata de imagens ela é dada pela subtração do pixel atual pelo pixel anterior, o que mostra a diferença entre os pixels, e por isso quanto maior a derivada maior é a diferença, que significa que existe uma borda ali. Sua aplicação é feita por meio do método da convolução, onde é possível obter uma matriz máscara que gera como resultado uma matriz que faça essa derivada. Em que é feito uma iteração de derivadas por inclinação (vertical e horizontal). Para fazer com que a informação das duas ao mesmo tempo é feito a operação gradiente nas duas coordenadas relativas a cada matriz de borda (vertical, que é o GY e horizontal, que é o GX) para obter a magnitude da borda e criar uma outra imagem, essa sim seria a imagem de bordas. Existem alguns métodos de convolução que são capazes de obter esses resultados utilizando matrizes máscara de área maior, que tiram o problema de perda de resolução.

Fazendo a mesma operação, mas utilizando a segunda derivada temos os mesmos valores, porém a diferença é mais acentuada. Um método de se calcular é o Laplaciano, que obtém a intensidade referente a magnitude em uma passada só, entretanto esse método tem suas desvantagens, ela não é capaz de mostrar o ângulo da borda, como é possível de se obter por meio da magnitude do método da derivada primeira. Isso faz com que perca a possibilidade de se obter os ângulos por meio da separação das matrizes. Gradiente da imagem: Um gradiente é referente a uma derivada vertical ou horizontal. A direção do gradiente pode ser calculada pelo arco tangente do gradiente de x dividido pelo gradiente de y.

**Intensidade da borda:** A matriz intensidade (que é a imagem contendo somente as bordas ou áreas de derivada maior) é calculado pela raiz quadrada da soma dos termos gradientes (x e y) ao quadrado. Esse método gera várias bordas e, nem sempre é interessante ter toda essa informação. Por isso é feito um processo de limiarização que descarta os termos menores que são menos significativos, ficando assim com uma matriz intensidade que possui somente as maiores raízes.

Aula 6

**Domínio da frequência:** Sabendo como funcionam frequências em imagens é possível atingir resultados de redução de ruídos periódicos muito mais interessantes que nos métodos apresentados anteriormente. Funções periódicas podem ser obtidas através de uma soma de funções trigonométricas seno e cosseno, essas funções, quando somadas, se possuem frequências diferentes elas geram frequências complexas e com elas é possível representar várias coisas. Uma delas é uma onda quadrada, que pode ser obtida ao somar uma quantidade suficiente de funções senos e cossenos. Somas de funções que tem o mesmo período, mesmo que multiplicadas (antes de serem somadas) mantém seu período. Série de Fourier: É a somatória de funções seno e cosseno, que podem representar qualquer função f(x), em que é uma somatória infinita composta pela soma das funções seno e cosseno multiplicados individualmente por um cociente a ou b, que são obtidos através de um cálculo das integrais que usa como parâmetro a função desejada. Utilizando a fórmula para conseguir uma função quadrada é obtido pela somatória infinita de:

$$\frac{4k}{\pi} \left( \sin(x) + \frac{1}{3} \sin(3x) + \frac{1}{5} \sin(5x) + \dots \right).$$

**Transformada de Fourier:** Ela é utilizada para decompor um sinal e suas frequências. Utiliza como base números complexos, por conta do uso do conjugado, que representa esses números por meio de um par de eixos bidimensional, em que um eixo é referente aos Reais e o outro para os Imaginários. Isso permite que seja obtido os ângulos da mesma forma que é feito com o conceito de magnitude no método anterior. Em que:

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-j*2*\pi*u*x} dx \text{ onde } j = \sqrt{-1}$$

Em que F(u) é a função da transformada de Fourier e u é a função f(x). A f(x) pode ser recuperada por meio da transformada inversa de Fourier:

$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j*2*\pi*u*x} du \text{ onde } j = \sqrt{-1}$$

Que existem se f(x) for contínua e F(u) for derivável em todo o intervalo.

Normalmente a função F(u) é complexa e por isso podemos representa-la de outra forma. Em que:  $F(u) = R(u) + jI(u)$  onde R(u) é a parte real e I(u) é a parte imaginária. E mais:  $F(u) = |F(u)| e^{j\theta(u)}$  em que  $|F(u)|$  chamado espectro de Fourier e  $\theta(u)$  chamado ângulo de fase, são dados pelas expressões:

$$|F(u)| = \sqrt{R^2(u) + I^2(u)} \text{ e } \theta(u) = \tan^{-1} \left( \frac{I(u)}{R(u)} \right)$$

O espectro de potência de f(x) é dado por  $P(u) = |F(u)|^2$ , também chamado de densidade espectral.

Como em PDI trabalhamos com dados discretos, as equações dadas acima podem ser representadas como o somatório do limite dos termos e F(u) é composta de uma soma infinita de termos seno e cosseno em que u determina a frequência do termo. Em imagens a transformada de Fourier é estendida para uma função que admite f(x, y).

De maneira discreta, é possível construir a transformada de Fourier da seguinte forma:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-\frac{2*\pi*u*x}{N}}$$

Em que N é a quantidade de iterações que irá fazer, onde N é igual a quantidade de pontos que se deseja calcular. Isso faz com que cada ponto tenha uma coordenada específica em duas matrizes, correspondentes as partes Real e Imaginária do número. Com o resultado da função podemos calcular o seu módulo, que é o espectro de Fourier, porém seus resultados podem variar de 0 até 2,6 milhões, portanto é necessário normalizar os valores para exibi-los em uma imagem, pela fórmula:

$$D(u, v) = c * \log(1 + |F(u, v)|)$$

Que fazem os valores variar de 0 a 6,4, assim é possível transforma-los em um intervalo (0, 255) para representa-los por meio de imagens. Para que as imagens fiquem no padrão conhecido, é feito um processo de deslocamento (shift), que troca as partes da imagem de certo modo em que as frequências baixas de amplitude alta fiquem no centro e as frequências altas fiquem nos cantos. A imagem é dividida em 4 partes e que são trocados de lugar nas diagonais.

Dentre as propriedades da transformada de Fourier estão:

- **Separabilidade:** a função 2D pode ser dividida em 2 somatórias 1D;
- **Translação:** se forem somadas as coordenadas da função f(x, y) terá um impacto na transformada, assim como se forem mudadas as coordenadas da transformada;
- **Rotação:** se a imagem for rotacionada, a transformada vai ter a mesma rotação;
- **Periodicidade e simetria:** como a transformada faz uma representação por meio de funções periódicas, a partir de um período N ela se repete;
- **Mudança de escala:** transformada da soma é igual a soma das transformadas, ou seja, a transformada de Fourier e a sua inversa são distributivas em relação a adição, mas não são quanto a multiplicação.
- **Valor médio:** a média da transformada pode ser obtida por:  $\frac{1}{N} F(0,0)$ ;
- **Laplaciano:** o operador laplaciano é usado para separar as bordas de uma imagem, e para a transformada pode ser obtida no ponto por:  $-(2\pi)^2(u^2 + v^2)F(u, v)$ ;
- **Convolução:** a convolução de f(x) por uma máscara g(x) é igual ao produto da transformada de x multiplicado pela máscara em x, ou seja:  $f(x) * g(x) = F(u)G(u)$ . Isso faz com que seja possível calcular a convolução em uma única passada ao invés de fazer a convolução completa que é muito mais custoso;

**Transformada rápida de Fourier (FFT):** A transformada de Fourier é extremamente custosa, com uma operação 1D de complexidade  $N^2$ , a transformada rápida transforma isso em uma complexidade  $N \log_2 N$ , que é muito menos custoso. Para obtê-la é preciso representar a transformada de uma maneira diferente:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) W_N^{ux} \text{ onde } w_N = e^{-\frac{j2\pi}{N}} \text{ e } N = 2^n \text{ ou } N \text{ é par}$$

A função acima então é dividida em duas, uma parte par e uma parte ímpar, e daí é obtida a seguinte equação:

$$f(u + M) = \frac{1}{2} (F_{par}(u) - F_{impar}(u) W_{2M}^u)$$

Essa fórmula tem como intuito substituir o processo de convolução, em que o processo é feito da seguinte forma: faz a transformada rápida da imagem e da máscara, separadamente, depois a matriz máscara recebe o mesmo

tamanho da imagem, preenchendo os campos vazios com 0, após isso é feito a multiplicação  $F(u)$  por  $G(u)$ , que gera uma matriz e como ela é feito o processo da transformada inversa para gerar uma imagem final.

Outras transformadas: Existem outras fórmulas que fazem um processo sobre a imagem, assim como a transformada de Fourier, algumas delas são muito utilizadas, como a transformada do cosseno, que é utilizada para compactar imagens, como no formato JPEG, que é a transformada de Fourier sem a parte imaginária (parte do seno).

Filtragem no domínio da frequência: um filtro ideal é aquele que faz com que a amplitude da onda da frequência filtrada seja igual a zero, mas na prática isso é muito difícil de fazer na prática, pois um corte completamente reto é muito difícil de ser feito em hardware. Esse filtro de frequências é um aplicado sobre o espectro de frequências de Fourier: para a redução de ruído periódico é multiplicada a matriz de espectro de frequência por outra matriz que tem 0 nas partes correspondentes ao ruído e 1 nas partes correspondentes a parte importante da imagem, isso faz com que as frequências relacionadas aos ruídos desapareçam.

#### Aula 7

Resumidamente, os processos realizados em uma operação de filtro no domínio da frequência são: com base na imagem de entrada é feito um pré-processamento na imagem, depois é feita a transformada de Fourier, depois a matriz resultado é passada pelo processo de filtragem, essa matriz filtrada é convertida novamente em imagem, que é aplicado um efeito de pós-processamento, e então é obtida a imagem resultante.

As frequências centrais do espectro de Fourier são referentes as frequências baixas da imagem, em que essas são as partes em que ocorrem menos alterações, ou seja, a parte com mais informações da imagem. Enquanto as frequências mais altas estão relacionadas aos detalhes da imagem.

Nomes dos filtros na frequência:

- **Passa-alta:** quando é aplicado uma função de filtro que reduz as frequências baixas da imagem, a imagem resultante é “borrada”.
- **Passa-baixa:** quando é aplicado uma função de filtro que reduz as frequências altas da imagem, a imagem resultante contém somente as bordas dos elementos. Se a imagem conter muitos detalhes e imagem resultante também fica com um plano de fundo acinzentado.
- **Sharpening:** quando é aplicado uma função de filtro que incrementa as frequências altas e mantém as baixas, isso faz com que a imagem fique realçada, em que suas bordas dão mais destaque. Também pode ser feita pela soma entre a imagem original e a soma com a passa baixa, ou pela redução das frequências baixas da imagem.

Todos os filtros aplicados podem ser obtidos através de algumas funções diferentes, que suavizam a filtragem, ou seja, a curva de corte da frequência pode mudar de acordo com a técnica utilizada para obter a função de filtro. Os 3 principais são: filtro ideal, que zera completamente a frequência ou deixa a frequência igual a original; filtro de Butterworth, que inicia a suavização após algum valor entre a frequência de intensidade 0 e a de intensidade 1; e o filtro Gaussiano, que suaviza desde a origem da frequência.

Também existe um tipo de filtragem que leva em consideração os valores da imagem em iluminação e reflectância, com isso ele é capaz de obter uma redução no valor de arquivo e ao mesmo tempo realce em contraste.

#### Segundo Bimestre:

#### Aula 8

**Restauração de imagens:** Tendo como origem uma imagem considerada danificada, o objetivo principal desse processo de filtragem é chegar mais próximo da imagem original, embora não saiba qual é essa, tendo em vista que a imagem original é desconhecida.

Com intuito de testar os programas desenvolvidos são feitas adições de danificação da imagem, ou seja, é adicionado um “borramento” e um ruído. Em que existem várias estratégias de adicionar um ruído que mais se parece com aqueles encontrados em imagens danificadas de fato.

Existem algumas estratégias de filtragem para redução de ruído. Dentre elas algumas foram descritas anteriormente. Das estratégias mais interessantes está a de filtros adaptativos, que leva em consideração características locais da imagem, aplicando métodos e filtragem diferentes de acordo com o local de imagem e suas características. Mantendo a imagem original se o desvio

padrão local for zero e se o desvio padrão local for maior que o desvio padrão da imagem, pois é referente a uma região de bordas; e faz a média da região se os desvios padrão local e da imagem forem iguais.

Ruídos periódicos são mais facilmente removidos com processo de filtragem no domínio da frequência.

O borramento, ou espalhamento, são funções mais difíceis de prever e entender o que as criou para tomar estratégias mais eficazes. Por isso é feito algum processo para identificar o padrão de espelhamento para trata-lo. Uma maneira de realizar este processo é obtendo uma nova foto como controle, de uma região que aparece na imagem danificada, e fazer um *matching* perfeito com imagem danificada, daí pode se obter a divisão da transformada de Fourier da imagem com ruído pela transformada da imagem controle e para obter o padrão é feita a transformada inversa, que é usado para tratar a imagem danificada criando uma nova imagem restaurada. Também pode ser feito um processo de chute do padrão de espelhamento, um chute bom, tendo como base a fabricação das lentes de câmeras e a maneira em que o aparelho funciona, é uma função de distribuição normal bidimensional, que tem valores maiores mais ao centro e cada vez menores nas bordas. Para verificar se o chute feito é bom, é quando o módulo da diferença entre a imagem danificada e a convolução entre a imagem filtrada e seu padrão de espelhamento. Para obter resultados ainda mais otimizados pode ser feito um processo de reconstrução iterativa, em que cada iteração é avaliada por um operador humano. Tendo em vista que para obter a função normal existe a variável sigma, que está diretamente relacionada com a compatibilidade com o padrão de espelhamento.

Nesses processos, esse espelhamento é feito na imagem toda uniformemente, o que não é realista quando se trata de equipamentos físicos que são passíveis de irregularidades tanto em relação a sua fabricação quanto a degradação conforme o uso. Mas assumir o contrário pode simplificar e viabilizar uma solução.

Para obter uma restauração melhor é feito ambos os processos simultaneamente para criar um bom exemplo de imagem danificada. Para cada tipo de deformidade ou danificação da imagem são feitos processos diferentes para obter a melhor restauração, dentre eles está a situação em que ocorre um espelhamento por movimento, em que é feito um tratamento de espelhamento 1D ao invés de 2D; ou em uma situação de transformações geométricas é feito um processo de transformação inversa obtendo pontos de controle da imagem.

#### Aula 9

**Segmentação de Imagens:** O objetivo é separar os objetos do fundo da imagem, separando-os por classes. Pode se atingir esse objetivo por meio de duas estratégias principais:

- **Similaridade:** quando os pixels são parecidos em uma determinada área. Essa estratégia utiliza das técnicas de limiarização, crescimento de regiões e fusão de regiões;
- **Descontinuidade:** quando existem bordas, ou distinções muito grandes entre um conjunto de pixels. Essa estratégia utiliza técnicas de detecção de mudanças bruscas, com detecção de linhas e bordas e pontos isolados.

Técnicas de similaridade:

**Limiarização:** uma das abordagens mais simples da estratégia de similaridade, em que a função atribui o valor 0 (ou a cor preta) se o pixel for menor que o limiar e 1 (ou a cor branca) se for maior ou igual ao limiar, esse processo é de binarização com limiar, não necessariamente limiarização. Em que a distinção entre os objetos e o fundo fica muito mais evidente. A tarefa mais importante, entretanto, é de encontrar o limiar, ou seja, o valor que distingue na maioria dos casos, o valor de intensidade que separa o fundo dos objetos. A técnica mais comum é utilizar o histograma, que demarca a frequência das intensidades, e separar dois grupos de intensidades, daí o limiar se torna um valor entre os grupos. Existem situações onde existem mais de um limiar, ou seja, o histograma possui mais de um grupo de intensidades. A limiarização é o processo de atribuir zero até um certo valor e a partir desse ponto manter os valores da imagem, o que se torna mais interessante com valores do mundo real, por conta da complexidade do caso a ser tratado necessitam da distinção de informação, que tornam os arquivos maiores, porém nos casos onde é possível aplicar a binarização, se torna mais interessante pois apresenta uma simplificação bem maior dos dados.

**Algoritmos de Limiarização:** Embora seja possível encontrar o limiar manualmente, é possível consegui-lo por meio de algoritmos, dentre eles estão:

- **P-Tile:** é a técnica mais simples, porém requer o tamanho da área do objeto na imagem. Encontrada a porcentagem de pixels  $p\%$  do objeto na imagem, e baseado nesse valor, encontra o limiar, utilizando o grupo de intensidades no histograma referentes a  $p\%$ .
- **Método de Otsu:** é uma técnica mais interessante, que leva em consideração a estatística da imagem, em que a imagem é dividida em duas classes: fundo e objeto, e para obter o limiar, é calculado a variância entre as classes e a variância total, e o limiar se torna a divisão entre elas.
- **Método de Pun:** é uma técnica que utiliza como base em entropia, em que são feitos chutes no valor do limiar  $t$ , e é feito um cálculo com base no histograma da imagem uma somatória  $p \log(p)$ , onde  $p$  é o valor da intensidade do pixel de 0 a  $t$  e uma de  $t$  até o fim. O valor do limiar final é aquele em que a soma das duas somatórias negativas é a maior.
- Existem outros métodos que usam como base o método da entropia.

**Limiarização local:** é um processo de limiarização que usa diferentes limiares em diferentes partes da imagem e tem como intuito corrigir o seguinte problema: existem situações onde parte do objeto são tão claras quanto parte do fundo da imagem, gerando uma certa confusão em que a imagem e o fundo se mesclam. Utilizando esse método é possível comparar pequenas sessões da imagem, que não sofrem do mesmo problema. Para evitar erros de localidade, pode-se utilizar uma vizinhança da área a ser calculada. Porém durante o processo de implementação, encontra-se outro problema: em quantas partes devo dividir minha imagem para separar os objetos sem erros? Que pode ser resolvido utilizando os processos:

- **Split and Merge:** é feita da seguinte forma: primeiramente a imagem é dividida em algumas partes (normalmente quatro) e é verificado se todos os pixels de um segmento possuem a mesma tonalidade, cor ou textura, dentro de uma certa tolerância. Se o segmento não tiver, então o segmento é dividido em mais partes, da mesma quantidade do início, e a assim por diante até que não existam segmentos que tenham tonalidades, cores ou texturas diferentes. Esse é o processo é chamado split, o processo de merge ocorre em seguida, da seguinte forma: cada segmento é unido com um segmento, ou grupos de segmentos que tem tonalidade, cor ou textura iguais, dentro de uma certa tolerância.
- **Crescimento de Regiões:** outra estratégia é pelo crescimento de regiões, que é feito da seguinte forma: a partir de um ponto de origem é verificado se os pixels vizinhos tem a mesma tonalidade, se sim a área expande contendo somente os pixels que satisfazem a requisição. Isso é feito até que não tenham mais pixels similares, formando uma segmentação de um objeto. Depois são gerados mais pontos que não pertençam a outros objetos e o processo se repete até que não tenham mais regiões que não pertençam a outros grupos.

#### Técnicas de descontinuidade:

**Deteção de pontos isolados:** é aplicado um processo de convolução em que a matriz máscara 3x3 que possui valor central 8 e valores diferentes do centro -1, o pixel de resposta recebe o valor resultante normalizado em módulo, por isso fazendo com que o pixel resposta seja grande quando a máscara passa por pontos isolados.

**Deteção de linhas:** é aplicado um processo de convolução em que a matriz máscara 3x3 que possui 3 valores iguais a 2 e 6 valores iguais a -1, em que os valores iguais a 2 tem formas diferentes baseado no que deseja ser detectado, podendo ser linhas horizontais, verticais e nas diagonais. Esse método faz com que exista uma resposta grande dependendo da máscara. Esses processos são úteis, mas não são muito comuns de aparecerem em imagens, principalmente o método de detecção das linhas. Pois isso os métodos mais usados são de detecção de bordas.

**Deteção de bordas:** usam os métodos já vistos de detecção de bordas baseados em derivadas, sendo eles os métodos de: Sobel, Laplaciano e x. Porém existe um método mais interessante nos casos onde a distinção entre o fundo e os objetos da imagem não são grandes, ou seja, a distinção de tonalidades entre o fundo e o objeto é pequena, essa abordagem é a limiarização de bordas, em que para chegar nesse estágio é necessário interessante fazer os seguintes processamentos na imagem:

- **Suavização:** um método de redução de ruído que deixa as bordas menos intensas dependendo do método, então é importante escolher um método que mantém elas, como o filtro da mediana;
- **Deteção de bordas:** utilizando um método visto acima;
- **Limiarização:** para eliminar as bordas de baixa magnitude, deixando somente as mais interessantes;
- **Afinamento das bordas:** como os processos de detecção de bordas tem o problema de deixar as bordas mais grossas, esse método resolve esse problema, definindo a localização verdadeira da borda, representado por somente 1 pixel de largura;
- **Ligação:** une os pixels que compõem uma reta, pois quando é feito o processo de afinamento, por conta do ângulo da linha formada, é impossível que tenha somente um pixel, então ela detecta esses segmentos de reta e cria uma reta mais intensa.

**Afinamento de bordas:** para obter a verdadeira localização das bordas é necessário afina-las, esse processo é feito pela corrosão dos pixels mais externos de uma linha, deixando somente o pixel central de uma linha. Esse método será visto posteriormente e é chamado Zhang-Suen, e opera com imagens binárias, ou seja, somente pixels de intensidade 0 ou de intensidade máxima. Em imagens não binárias, que representa a maioria dos casos, estamos interessados em representar as bordas mais fortes, ao invés de remover os pixels ao redor da linha. Esse método é chamado de afinamento por supressão não máxima e é caracterizado por manter o pixel de tonalidade mais alta mesmo que não seja o centro da borda.

**Ligação de bordas e detecção de fronteiras:** é o processo de ligar os pontos isolados e pequenos segmentos para formar linhas retas, círculos ou elipses, tornando contornos mais regulares. Nesse processo é utilizado a magnitude da borda e seu ângulo. No método de processamento local, é feito um processo de verificação de vizinhança, para detectar pixels que apresentem informações similares. O processo ocorre da seguinte forma:

1. É feita a varredura e rotulação dos pixels;
2. Pixels de mesma direção ou magnitude parecida recebem o mesmo rótulo;
3. As informações relacionadas as retas são armazenadas (Pontos inicial e final, direção e número de pixels)
4. A imagem é percorrida de cima para baixo e da esquerda para direita, para verificar se existem pixels não rotulados, em que é verificado a sua vizinhança e se for encontrado um pixel vizinho já rotulado ele recebe o mesmo rótulo.

Para otimizar o processo, é possível ainda separar as buscas por retas anguladas em grupos de retas. Também deve-se atentar ao caso de um pixel estar a 1 pixel de distância de outro e mesmo assim ser calculado, onde as retas nem sempre estarão conectadas em um estado anterior ao processo de ligação. Por isso são usados pixels ponte, que ligam esses dois pixels com a magnitude da média dos dois, e ângulo iguala ambos os pixels, que auxiliam o processo de demarcação de retas.

Outra maneira de fazer o processo de ligação, ou conexão, é por inundação, que é muito mais simples que o processo anterior, em que é categorizado pelo seguinte processo: é demarcado com o mesmo rótulo se o pixel tem o mesmo ângulo de seu vizinho, sempre seguindo a linha ele demarca as linhas. Porém não é muito eficiente quando se trata de preenchimento de pontos não preenchidos e também no caso de identificação dos pontos inicial e final, por conta de o processo de inundação ocorrer por lados diferentes de acordo com a abordagem.

Um processo mais geral é o detector de Canny, que visa detectar bordas de qualidade, isto é, bordas finas e sem quebras, mas para isso é necessário passar por vários processos, começando com uma redução de ruído, por meio de um filtro Gaussiano, depois é aplicado o filtro de Sobel para calcular as magnitudes e os ângulos das bordas, então é feito o processo de afinamento por supressão não máxima e então é aplicado um threshold duplo, que utiliza 2 limiares, um menor e outro maior, em que acima do maior é mantido e abaixo do menor, descartado, e com os pixels restantes, ou seja, os que estão entre os limiares, é feito o processo de Histeresys, em que os pixels de menor intensidade se tornam mais intensos, por meio de uma análise de sua vizinhança, que verifica se os vizinhos são de intensidades fortes ou fracas, se são fortes ele é mantido, se eles forem de intensidade fraca ele é descartado.

**Processamento global:** neste processo, objetos como linhas, círculos, elipses, entre outros são obtidos observando as relações globais entre os pixels, dentre eles o mais conhecido, e o que iremos aplicar é o processo da transformada de Hough. É preferível utilizar métodos globais pois apresentam uma abordagem mais eficiente, não sendo necessário analisar a vizinhança de um pixel e fazer isso para todos os pixels, o que é muito custoso, mas sim analisar a imagem como um todo.

**Transformada de Hough:** para reconhecer circunferências, Hough faz um processo onde, dado um conjunto de pontos é traçado uma circunferência centrada no ponto para cada ponto. Daí é gerada uma matriz contendo os pontos das circunferências e onde foi desenhado é somado 1 e onde houve intersecção ele tem mais de um, então é obtido o maior valor, ou seja, o maior número de intersecções, e nesse ponto é desenhado uma circunferência que interpola os outros pontos, encontrando uma circunferência na imagem. Porém, neste caso é necessário testar alguns raios de circunferência até que se encontre um que possua uma acumulação grande. Embora essa estratégia seja interessante para circunferências, ela deixa de ser interessante quando aplicada para retas, pois existem infinitas retas possíveis com um único ponto, por isso Hough aplica em retas de uma maneira diferente, para cada ponto é atribuída uma reta, porém não é mais com base no y, mas sim no b, isto é, uma reta  $y = ax + b$  se torna  $b = y - ax$ , e com criando um gráfico com eixos a e b, e desta forma para criar uma reta não é necessário duas variáveis e sim uma só (no caso o eixo a). Esse gráfico com essas retas, se os pontos estabelecem um padrão como de uma reta, ele gera um ponto de intersecção entre as retas geradas. Esse ponto representa os valores de a e de b para a reta com base nos eixos x e y, isto é, representa valores a serem colocados para a equação da reta  $y = ax + b$ . Daí o método de visualização de intersecção e torna igual a do método para encontrar circunferências, ou seja, é feita uma matriz e somado 1 ao ponto correspondente na linha reta, e onde existe o número máximo de intersecções esse é o ponto referente a reta. Foi construído um algoritmo que utilizava dos ângulos gerados pelo gradiente, criando valores a baseados na divisão do gradiente  $\frac{gx}{gy}$ .

Existe um problema com essa abordagem, em que retas de ângulo de  $90^\circ$  terão uma divisão por 0, gerando  $a = +\infty$ . Para solucionar esse problema, é abordado uma nova parametrização, com:

$$\rho = x \cos \theta + y \sin \theta$$

Obtendo ao invés de retas em um gráfico de eixos a e b, são obtidas curvas em um gráfico de eixos  $\rho$  e  $\theta$ . E a reta resultante é obtida utilizando os valores de  $\rho$  e  $\theta$  obtidos pelas intersecções formadas.

O algoritmo resultante é construído da seguinte forma:

1. É feito os acumuladores, ou a matriz,  $(\theta, \rho) = 0$  (É necessário espaçar adequadamente os intervalos);
2. O acumulador  $(\theta, \rho)$  é incrementado no ponto  $(\theta, \rho)$ , que satisfaz a equação  $\rho = x \cos \theta + y \sin \theta$ , em que  $\theta$  varia e  $\rho$  é calculado;
3. O processo de 2 se repete para todos os pontos da imagem;
4. Os picos no acumulador  $(\theta, \rho)$  fornecem os valores  $\theta$  e  $\rho$  procurados.

Esse processo exige uma matriz grande para acumular os termos, e também cria equações de reta ao invés de segmentos de retas, como era no processamento local, isto é, cada borda tem uma reta infinita. Também é possível reduzir os segmentos de  $\theta$  por meio dos ângulos proporcionados por Sobel.

Outra maneira de fazer a segmentação de imagem, é por meio do uso de movimento, ou seja, dado uma sequência de imagens, a parte que está em movimento (se o ponto de captura for fixo), a parte que se move é o objeto e o que resta é o fundo. Que se torna fácil de calcular, com um simples processo da diferença, ou subtração entre elas.

Aula 10:

**Compressão de imagens:** com o desenvolvimento da tecnologia, câmeras cada vez mais baratas são capazes de tirar fotos com precisão cada vez maior, o que gera um problema: quanto maior a precisão maior se torna o arquivo e com isso traz problemas tanto de armazenamento no dispositivo quanto de transferência dessas informações na internet. Para resolver esse problema é possível reduzir o tamanho das imagens por meio da compressão. A compressão tem o objetivo de representar uma imagem com algum nível de qualidade exigido, preservando informações essenciais de forma que ela possa ser reconstruída com a precisão desejada.

Para tanto existem 3 eixos principais: a diminuição do tamanho da imagem, a distorção que pode gerar e o grau de complexidade computacional. Na maioria dos casos um depende diretamente do outro, isto é, a imagem pode ter um tamanho muito pequeno, mas gera uma distorção grande ou até demora demais para ser transformada em imagem novamente. Existem casos onde a imagem não é distorcida nem exige tempo demais a ser computada, mas seu tamanho é pouco reduzido.

Existem dois termos importantes, que são a compressão e a compactação: a compressão é um processo sem perdas, como arquivos RAR ou GIF, enquanto a compactação gera perdas como arquivos JPEG ou MP4.

O método de compactação que utiliza o domínio da frequência que tem se mostrado mais eficiente é a transformada discreta do cosseno, caracterizado pelo seu uso nos arquivos JPEG.

Fundamentos: a redundância é o termo central da compressão de dados.

Dado  $n_1$  e  $n_2$  como sendo o número de unidades de transporte de informação e dois conjuntos de dados que representam a mesma informação então a redundância dedados relativa  $R_D$  de um conjunto de dados pode ser definida como:  $R_D = 1 - \frac{1}{C_R}$  em que  $C_R$  é a taxa de compressão dada por  $n_1 / n_2$  em que quando:

- $n_1 = n_2$ , então  $C_R = 1$  e  $R_D = 0$ , ou seja, não há redundância;
- $n_1$  é muito maior que  $n_2$ , então  $C_R \rightarrow \infty$  e  $R_D = 1$ , ou seja, existe uma alta redundância;
- $n_1$  é muito menor que  $n_2$ , então  $C_R \rightarrow 0$  e  $R_D \rightarrow -\infty$ , ou seja, o conjunto de dados contém mais informação que o conjunto original;

Em resumo a taxa de redundância calcula quantos dados existem de forma desnecessária, em que podem ser representados de outra maneira sem que haja perda, ou seja, é passível de compressão.

Uma maneira de fazer um processo de compressão é pela análise de probabilidade de uma tonalidade aparecer em uma imagem, e baseando-se nela é possível representar as tonalidades de maior frequência com menos bits e os de maior frequência com mais bits, assim armazenando uma mesma imagem ocupando menos espaço. Esse processo é chamado de codificação de comprimento variável, obtido pelo método de Huffman.

O método anterior proveito da imagem ter uma distinção de tonalidades pequena, mas existem métodos que utilizam padrões distintos, onde é possível tomar proveito de uma imagem possuir pixels próximos com tonalidades parecidas, isto é, uma redundância interpixel. Uma maneira de utilizar essa redundância seria armazenando a diferença entre os pixels ao invés da informação original, e esses valores são chamados mapeamentos, que pode ser reversível se a imagem original puder ser construída a partir dos dados transformados. Essa ideia pode ser combinada com outra redundância que ocorre bastante com imagens que passam por esse processo: muitos pixels terão a mesma diferença de um para o outro, sendo assim é possível armazenar somente 1 pixel e a quantidade de pixels a partir dele que possuem a mesma diferença, realizando um processo de compactação extremamente eficiente.

Modelos de compressão de imagens são caracterizados por duas etapas: um codificador, que comprime a imagem e faz ela se tornar um código reduzido e um decodificador, que transforma o código na imagem original. Em que a imagem original é codificada, e daí é armazenada ou transmitida e ao exibi-la ela é decodificada.

Usando a teoria da informação: ela é quem faz a pergunta “Qual é a quantidade mínima de dados realmente necessária para representar uma imagem?”, e para isso utiliza métodos probabilísticos, categorizando as informações como unidades de informação que é calculado através da entropia. Isso é usado durante um processo que verifica o quanto uma imagem pode ser comprimida, em que quanto maior a sua entropia, menos bits ela precisa para ser representado, isto é, ela pode ser mais comprimida. Uma estimativa de entropia de primeira ordem é um cálculo que leva em consideração cada pixel, que pode ser utilizado para obter o limite inferior da compressão, isto é, o mínimo que a compressão garante, que quantifica a taxa de compressão quando tratado por tamanho variável. Entropias de ordem superior, ou seja, aquelas que utilizam blocos de pixels ao invés de um pixel só, são usadas para verificar redundâncias, de modo que se as entropias de ordem superior forem iguais a primeira, significa que não existe redundância interpixel e a codificação por tamanho variável é a melhor compressão possível.

Para realizar um processo de compressão com erro tolerável, é utilizado um processo de codificação por transformada, em que uma imagem original é realizado um seguinte processo: primeiramente é dividida a imagem em sub imagens mais fáceis de serem processadas por um processo de transformada, depois é feito o processo de transformada direta, ela passa por um quantificador, que é responsável por transformar os resultados que estão em ponto flutuante (ou float) de uma transformada em valores inteiros, depois disso esses valores inteiros passam por um codificador de símbolos, ou seja, um processo que comprime esses valores.

Existem vários processos de transformadas que podem ser aplicados, e após todo esse procedimento foi comparado as transformadas, e a que mostra a menor quantidade de erros ou diferenças entre a original é o processo da transformada do cosseno, por isso é utilizada em métodos de compressão. Como processos de transformada são muito custosos quanto maior é a matriz a ser processada, comumente a imagem antes de ser processada é dividida em blocos, mas existe uma quantidade de blocos certa a ser processada, em que quanto mais blocos mais rápido é a transformada mas mais transformadas são necessárias fazer e o inverso também se aplica, por isso foram testadas quantidades de subdivisões de imagem antes de ser aplicado o processo de transformada, o que depende da transformada em que foi aplicado.

A maneira em que esse processo de compressão se torna eficiente é por meio da redução da quantidade de bytes em que são armazenados, por meio do armazenamento da transformada. Após o processo da transformada e da quantização, grande parte dos valores são zerados, valores referentes as frequências mais altas do espectro da transformada, que são os detalhes da imagem. A compressão é feita quando a matriz com a maioria dos valores sendo 0, é armazenada como uma lista, de maneira organizada, para que ela possa ser recuperada. Então a partir dessa linha de informações é possível gerar uma matriz, e dela transformar de volta em uma transformada do cosseno e aplicando o processo de transformada inversa obtém se a imagem original.

Aula 11

**Representação e descrição de imagens:** Em vários casos é interessante tomar informações de imagens, isto é, a partir de uma imagem, obter informações referentes a objetos ou até sobre o fundo da imagem, e a partir deles obter informações como sua área, seu comprimento, desta forma descrevendo imagens. Normalmente a imagem é subdividida e classificada, essa classificação é tradicionalmente feita utilizando classificador Bayesiano ou uma Rede Neural.

**Códigos da cadeia:** é uma maneira de simplificar as informações referentes a borda de uma imagem. O processo é feito ao criar pontos de amostragem nas bordas da imagem, após isso esses pontos são sobrepostos a uma grade larga, depois os pontos são descritos sobre os pontos mais próximos referentes a grade. Com esses pontos formados, é feito o processo da cadeia em que cada ponto é ligado no adjacente com uma direção (normalmente são 4 ou 8 direções, chamados viz-4 e viz-8), o código referente a direção é armazenado em uma lista referente aos códigos da cadeia. Esse código depende do ponto de partida, é convencionado de se rotacionar o ponto de início da lista até chegar no menor número gerado pela cadeia para que seja um único número.

O mesmo processo de simplificação de uma forma pode ser obtido pelo perímetro mínimo do polígono, em que é feita uma grade sobrepondo um conjunto de pontos mais denso, em que formam uma forma sem espaços. A partir dessa forma são traçadas linhas retas que são as arestas da grade ou uma linha diagonal que passa pela grade. O resultado final é um polígono que é capaz de representar as bordas da imagem.

Também é possível chegar em uma aproximação polinomial por meio de retas, em que é traçada uma reta entre os 2 pontos mais distantes da forma, e depois são encontradas as linhas, perpendiculares a esta linha traçada, as que têm maior comprimento, com os pontos mais distantes é formado um polígono. A partir das arestas desse polígono é possível repetir esse processo até que aproximação fique satisfatória. É possível fazer o mesmo processo com linhas, sem a parte de descrição por polígonos.

Outro tema em descrição e representação de imagens são as assinaturas. Uma assinatura é a representação unidimensional da fronteira, ou borda, de um objeto e para obtê-la é necessário obter o centro de um objeto, que pode ser calculado a partir da média das coordenadas x e y de um objeto pertencente a uma imagem binária. A partir dele pode ser obtido o raio e o ângulo de uma reta r. Em que se o objeto for circular, o raio da reta será

constante, e se for um quadrado o raio varia. Essa é a assinatura de um objeto, um gráfico representado pelo tamanho do raio de um objeto com retas sendo traçadas em todos os ângulos  $\theta$ . Em que se o objeto for diferente de um círculo e for rotacionado, sua assinatura irá mudar. Assinaturas são muito úteis quando se trata de reconhecimento de objetos, pois elas permanecem iguais conforme o objeto muda de posição, mas são diferentes, porém similares, quando mudam de escala ou rotacionam (se for diferente de um círculo). Essas diferenças podem ser tratadas, em que a escala pode ser tratada ao multiplicar todos os valores da assinatura por uma constante k, isto é, se uma assinatura dividida por outra todos os valores resultam em uma variável k, se as assinaturas forem iguais então é o mesmo objeto. Similarmente a rotação pode ser tratada, ao mudar a posição dos elementos da assinatura, atribuindo os primeiros pontos, em ordem aos últimos e se em algum ponto as assinaturas forem iguais, então é o mesmo objeto. Outra maneira de tratar a rotação é, ao fazer o processo da assinatura, iniciar pelo ponto mais distante, se não for um círculo, tirando a variável do ponto inicial.

**Fecho convexo:** é o menor polígono convexo que contém objeto, em que um polígono convexo é aquele que se ao criar retas com dois pontos pertencentes ao objeto e não existir uma reta que passa por pontos que estão fora do objeto, o objeto é convexo. O fecho convexo é uma ótima maneira de representar um objeto, pois a razão entre o tamanho de suas arestas pode ser utilizada como assinaturas que não variam por rotação nem escala. O esqueleto de um objeto é outra maneira de simplifica-lo, utilizando uma função que resume ela em um gráfico. Ela é a estrutura de uma forma, ou linhas centrais, que pode ser obtido por meio de um processo de afinamento. Outra maneira de obter o esqueleto de um objeto é por meio do vizinho mais próximo, em que todo ponto cujo existam mais de um vizinho mais próximo faz parte do esqueleto, porém cálculos de vizinho mais próximo são muito custosos ao serem feitos com uma imagem, por isso o processo de afinamento é mais utilizado.

O processo de afinamento de Zang Suen é feito da seguinte forma:

Dado um ponto e sua vizinhança, e tomando como base que a matriz contém somente elementos 0 e 1;

A matriz de vizinhança é representada da seguinte forma:

Passo 1:

- Se a soma dos valores referentes a vizinhança estiver entre 2 e 6. Isso verifica se o ponto é uma extremidade de segmento de esqueleto;
- Se a quantidade de transições de 0 para 1 para o próximo vizinho ( $p_2, p_3, p_4, p_5, \dots$ ) é igual a 1. Isso verifica se é uma linha de esqueleto;
- São feitos os seguintes produtos e se ambos forem iguais a zero então ele prossegue:  $P_2 * P_4 * P_6$  e  $P_4 * P_6 * P_8$ .

Os pontos que passaram todas as verificações anteriores são marcados para serem apagados. O teste é feito pela imagem toda. Um pixel só é apagado se o pixel  $P_1$  tem 8 vizinhos iguais ao seu valor (0 ou 1).

Passo 2:

Então o processo se repete mais uma vez, só que o último passo é diferente: em que os pontos são multiplicados  $P_2 * P_4 * P_8$  e  $P_2 * P_6 * P_8$ .

Os pontos que passaram todas as verificações anteriores são marcados para serem apagados. O teste é feito pela imagem toda. Um pixel só é apagado se ele passou por essas condições.

Os passos 1 e 2 são repetidos até que não haja mudança na matriz final entre um passo de repetição e outro.

Com as bordas da imagem podem ser obtidas várias informações interessantes sobre o objeto sem uso de muito processamento, são chamados descritores simples: O comprimento do contorno ou o perímetro do objeto, calculado a partir do código da cadeia, em que quando são diagonais seu tamanho é  $\sqrt{2}$  e quando não é o tamanho é 1. O diâmetro da fronteira que é obtida através da maior distância entre pixels da borda. A curvatura ou o ângulo entre retas, que pode ser aproximada pela primeira diferença. Os números de formas, que é obtido após fazer o código da cadeia, e obter a diferença, o valor da diferença é rotacionado até obter o menor número, ou magnitude, tendo assim o número da forma.

Descritor de Fourier pode ser usado para representar uma simplificação das informações referentes a uma matriz que contenha os valores de bordas de um objeto, em que, por exemplo, as bordas de um quadrado podem ser representadas como uma forma circular quando obtido somente os primeiros termos da transformada, mas se forem obtidos quase todos os pontos da transformada a matriz é idêntica a original. Para a discretização, todas as



operações vistas anteriormente com relação a transformada, ainda permanecem. Isto é operações de rotação, translação, entre outras, também fazem efeito quando o objeto passa pela transformada.

Existem vários possíveis descritores de formas, que são utilizados para identificar objetos em outras imagens, dentre eles, além dos vistos, existem descritores de mais completos, que levam em consideração o objeto como um todo, esses são descritores regionais. Que é o caso dos descritores de área e perímetro, que são muito eficientes em verificar diferenças maiores entre objetos. Com elas pode ser obtido a compacidade, que é definida por  $\frac{Perímetro^2}{Área}$ , que é ainda mais eficiente e invariante a rotação, utilizada para fazer distinção entre classes de objetos. Outros descritores regionais são: média e mediana de tonalidades, valores mínimos e máximos, número de pixels com valores acima e abaixo da média. Existem vários possíveis descritores de um objeto.

Descritores topológicos, utilizam de uma base de topologia, que descreve um objeto por meio de valores que são independentes de deformações, desde que não ocorra divisão ou fusão da forma. Dentre eles está o número de componentes conexos, bem como o número de buracos em uma forma. Esses descrevem objetos que são formados de um conjunto de componentes (C) ou um componente que tem buracos (H). Um descritor pode ser feito pelo número de Euler  $E = C - H$ . O número de Euler pode ser utilizado para descrever objetos mais complexos, que possuam vértices (W), arestas (Q) e faces (F), em que  $E = W - Q + F = C - H$ .

Descritores baseados em textura, em que algumas de suas unidades de medida são: suavidade, rugosidade, regularidade e momentos. Momentos são os mais utilizados, momentos centrais podem ser obtidos pela integral:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - (x - \bar{x})^p)(y - \bar{y})^q f(x, y) dx dy$$

A partir deles podem ser definidos informações mais interessantes ao serem normalizados:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q}{2}}} \text{ e } \gamma = \frac{p+q}{2} + 1$$

E com esses valores se obtém os 7 momentos de Hu. Eles são os descritores de momentos, em que não são muito alterados ao deformar a imagem base. É possível obter outros momentos a partir de uma abordagem estatística, em que é calculado, a partir do histograma da imagem, a probabilidade de ocorrência de certas tonalidades. Que podem ser utilizados para descrever características de uma textura. Essa forma possui uma falha, que ela não leva em consideração a posição dos pixels, que quando se trata de textura é importante.

Outra abordagem estatística que leva em consideração a localização dos pixels é o uso de matrizes de co-ocorrência, determinado as ocorrências de pares de tonalidades em certas disposições, em que normalmente é verificado a probabilidade de ocorrência dos pares de pixel da tonalidade atual e do valor abaixo e a direita, em que é feito uma matriz quadrada de tamanho igual a quantidade de tonalidades, para obter uma matriz, chamada matriz de co-ocorrência, com todas as ocorrências de pares de tonalidades e o valor final é dado pela divisão de todos os elementos pela quantidade de tonalidades. E com essa matriz é possível obter discretizadores de probabilidade máxima, que é o par que possui a maior probabilidade de ocorrer, também é possível obter o momento de ordem k, bem como a sua inversa, uniformidade e entropia. Em que esses descritores são interessantes quando as imagens utilizadas apresentam repetição.

A abordagem espectral leva em consideração o espectro de Fourier, pois como visto, é ideal para descrever padrões periódicos. Em que após feito a transformada pode ser obtidas informações referentes aos padrões da imagem e até remove-los, para obter informações referentes a outras partes da imagem. Que utiliza como localização as coordenadas polares dos elementos da imagem, onde uma coordenada polar representa coordenadas (x, y) como coordenadas (r, θ), referentes ao comprimento da reta e ao ângulo da reta, respectivamente. E com eles podem ser construídos gráficos com esses valores, em que podem ser construídas funções para cada direção uma função S(r) e para cada r uma função S(θ). Esses gráficos representam bem texturas, principalmente quando levado em consideração os picos dos gráficos e suas informações. Que podem ser discretizados por meio de suas somatórias.

Em suma, para realizar um processo de reconhecimento de objetos completo, é feito na seguinte ordem:

- As imagens capturadas são digitalizadas e binarizadas

- São encontrados descritores para as imagens
- Para que possa ser classificado pode ser utilizado uma rede neural
- Algumas imagens são utilizadas para treinar a rede neural enquanto outras são utilizadas para testa-la
- Pode se utilizar um sistema que representa as classificações em que cada classe é uma dimensão de representação

Abordagens alternativas levam em consideração Redes Neurais Convolucionais, em que não é necessário discretizar a imagem, mas o próprio algoritmo, por meio do treino da rede neural, encontra as diferenças e características relacionadas as imagens.

## Aula 12

**Morfologia:** em processamento digital de imagens é um conjunto de técnicas que visa alterar uma imagem, de modo a remover objetos, inserir novos objetos, realizar processos de filtragem, em que todos estão inclusos durante os processos de pré e pós processamento, tais como: a filtragem morfológica, um afinamento, operações de poda.

Um das operações básicas de morfologia são a dilatação e a erosão. Elas utilizam como base a teoria dos conjuntos. A dilatação ocorre da seguinte forma: dado um elemento B, chamado elemento estruturante e uma imagem A, o elemento estruturante passa por toda a imagem, em que quando o ponto central, que pode ser redefinido para qualquer ponto do elemento B, em que ele passa pelas bordas da imagem, isto é, quando partes de B não estão inclusas no elemento A, ela cria novos elementos, baseados no elemento B, que são inseridos na imagem final. Em outras palavras, tudo o que B sobrepor será anexado em A, gerando uma imagem final dilatada. O processo realizado é similar a uma convolução, onde normalmente a dilatação é aplicada em imagens binárias para que os objetos aumentem de tamanho. Outro processo é a erosão, em que o objetivo agora não é criar elementos novos e sim remover elementos, de modo que dessa vez o elemento B sempre deve ficar dentro da imagem. A imagem resultado diminui o tamanho da imagem, em que o tamanho final é o tamanho inicial menos o a largura do elemento B. A erosão tira pixels das bordas dos objetos da imagem.

**Dilatação e erosão:** essas operações são complementares, em que utilizando uma combinação dessas duas operações é possível realizar um conjunto muito grande de processamentos de imagens, desde remoção de ruído até detecção de bordas.

**Abertura e fechamento:** a abertura faz uma suavização de contorno da imagem, quebrando ligações estreitas, enquanto o fechamento suaviza os contornos fundindo ligações estreitas e deixa bordas mais grossas. A abertura é obtida pela dilatação da erosão de uma imagem, enquanto o fechamento é o inverso, ao fazer a erosão da dilatação. Também podem ser calculadas por meio de interpolações geométricas, em que supondo que o elemento estruturante B seja uma bola, para obter a abertura é feito um processo onde a bola rola pela imagem, sempre dentro do objeto, os pixels resultantes são os pixels da imagem em que B passou por cima. De maneira similar, quando o elemento A sofre um fechamento, supondo a mesma situação anterior, a imagem resultante é aquela em que, quando B roda no contorno exterior da imagem, com B completamente fora da imagem, a imagem resultante será onde B alcança, em que os pixels entre a imagem e o elemento B são preenchidos. Esses métodos podem ser aplicados para fazer um processo de redução de ruído eficiente.

**Transformada hit-or-Miss:** é um processo de reconhecimento de objetos simples, e é caracterizado pela intersecção de dois elementos, um elemento é A erosão de X, em que X é o objeto, e o outro elemento é a erosão entre o A complementar e um objeto W – X, em que W é um objeto qualquer, que é um objeto um pouco maior que X, para gerar uma janela que contenha um objeto de “folga”.

**Extração de fronteiras:** para obter as bordas de um objeto, simplesmente é feita a subtração da matriz imagem pela sua erosão.

Preenchimento de regiões: para preencher certo segmento de uma imagem, pode ser feito esse processo, que é X da iteração k que é dado pela dilatação entre X da iteração anterior e B, intersecção com A complementar, a ideia é que se a X da iteração k preenche valores que não estão na imagem anterior ele salva essa nova imagem.

**Fecho convexo:** para obter o fecho convexo é feita a operação de transformada hit-or-miss, utilizando como elemento B os seguintes valores, em que é feita uma iteração para cada B:



O fecho convexo é obtido pela união entre os termos de convergência dos 4 elementos, em que esse termo de convergência é aquele X resultante em que a iteração atual é igual o elemento da iteração anterior.

O afinamento da imagem pode ser feito por meio da imagem original menos a sua transformada *hit-or-miss*, que também pode ser obtido por meio de um processo por um processo iterativo de *hit-or-miss* onde o B da operação vai rotacionando, e o resultado final é uma espécie de esqueleto da imagem original.

A operação contrária é o espessamento, em que é obtida pela união entre a transformada *hit-or-miss* e a imagem original e os valores das matrizes referente aos B são invertidos.

Existe, também uma maneira de se obter esqueletos de um conjunto A. Por meio de um processo de erosão, em que ele é dado pela união de todos os elementos do seguinte processo iterativo: k erosões entre A e B, menos abertura entre k erosões entre A e B e B. Em que k é a última iteração antes que A se torne um conjunto vazio. Outra maneira de se criar um esqueleto para um objeto é utilizando discos, e que é inserido um disco de tamanho máximo dentro de um objeto, em que o tamanho do disco é alterado para que ele consiga passar pelo objeto completo, em que o ponto central do disco máximo é o que cria o esqueleto, ao passar por cima da imagem, isso significa que o raio do disco é exatamente a distância entre os vizinhos mais próximos do seu ponto central.

O processo de encontrar o esqueleto de um objeto pode gerar pontos isolados ao redor do objeto, então para resolver este problema e complementar o procedimento de obtenção de esqueletos, existe uma técnica chamada Poda (ou Pruning). Este processo consiste em eliminar somente os pontos extremos da imagem, afim de que os pontos isolados, que tem um conjunto de pontos pequeno desaparecerá. Porém, isso faz com que os pontos extremos do objeto também sejam perdidos. Que é obtido por meio de um processo de afinamento, porém os elementos estruturantes são diferentes dos vistos anteriormente, e são inseridos 8 elementos ao invés de 4.

Até então foi visto somente a morfologia aplicada em imagens binárias, porém é possível aplica-la em níveis de cinza, o que significa que é possível também aplicar em uma imagem colorida. O processo de dilatação fica da seguinte forma:

Considerando que uma imagem e o elemento estruturante são funções bidimensionais, a dilatação pode ser expressa pela seguinte equação:

$$(f \text{ dilatação } b)(s, t) =$$

$$\max\{f(s - x, t - y) + b(x, y) \mid (s - x), (t - y) \in D_f: (x, y) \in D_b\}$$

Em que  $D_f$  e  $D_b$  são os domínios da função f (ou a imagem) e da função b (ou o elemento estruturante). Em que a expressão acima pode ser dita como o maior valor da função f nas coordenadas  $s - x$  e  $t - y$ , somado a b nas coordenadas (x, y) tal que todas as coordenadas pertençam ao domínio da outra função.

De maneira similar, a erosão pode ser representada por:

$$(f \text{ erosão } b)(s, t) =$$

$$\min\{f(s + x, t + y) - b(x, y) \mid (s + x), (t + y) \in D_f: (x, y) \in D_b\}$$

Na prática, em tonalidades de cinza a dilatação interpreta as tonalidades mais claras como o objeto e as mais escuras como o fundo, por isso, como resultado final as tonalidades mais claras sobrepõem as tonalidades mais escuras. Já na erosão acontece o contrário, em que o processo insere as partes mais escuras nas partes mais claras.

Como visto anteriormente, as operações de abertura e fechamento podem ser obtidas através das de dilatação e erosão. Em que é possível obter um borramento satisfatório de uma imagem a partir do processo de abertura seguido pelo processo de fechamento, que é útil para calcular bordas de uma imagem. De maneira similar, como visto anteriormente, é possível reconhecer as bordas de uma imagem ao fazer um processo dilatação subtraído por um processo de erosão em uma imagem.

Existem alguns processos morfológicos que podem ser feitos em imagens, de maneira similar aos vistos anteriormente, em que um deles é a transformada *top-hat* (ou transformada da cartola), que enfatiza os detalhes da imagem, escurecendo e aumentando o contraste entre as tonalidades, que pode ser obtido por meio da subtração entre a imagem original e a sua abertura.

É possível fazer também uma segmentação de textura, utilizando fechamentos sucessivos até que o elemento estruturante seja tão grande quanto os objetos pequenos, fazendo com que sejam eliminado e só sobrem os elementos maiores, sendo possível assim fazer uma segmentação.