

PRINCÍPIOS BÁSICOS
DA VISÃO POR COMPUTADOR
E
DO PROCESSAMENTO
DE IMAGENS

Jacques Facon

O autor agradece:

- a Alessandra, Andréia de Jesus, Carla, Fabrício, Francisco, Joaquim, Julius Cesar, Lorimar, Miguel, Paulo, Luciana, Liuz Antonio, Paulo Cesar, Paulo Roberto, Raquel, Stefano que participaram de forma ativa a essa compilação.

Sumário

1	INTRODUÇÃO	1
2	VISÃO COMPUTACIONAL SEGUNDO FISCHLER	3
2.1	Introdução	3
2.2	O Paradigma Sinais-Símbolos	5
2.3	Análise de Imagem de Baixo Nível	6
2.4	Aquisição de imagem (escanear e quantizar)	7
2.5	Preprocessamento de imagem (binarização e suavização)	8
2.5.1	Binarização de imagem	9
2.5.2	Suavização de imagem	9
2.6	Detecção de descontinuidades e homogeneidades locais (Borda, textura, cor)	10
2.6.1	Detecção da Borda Local	11
2.6.2	Detecção de Borda Local Baseada em Inibição Lateral	13
2.6.3	Análise da Homogeneidade local	14
2.6.4	Análise de texturas micro e macro estruturadas	15
2.6.5	Geometria de uma cena local a partir de uma única imagem (forma a partir do sombreamento e textura)	16
2.6.6	Geometria de uma cena local a partir de múltiplas imagens (estereoscopia e fluxo ótico)	17
3	RECONHECIMENTO DE PADRÕES SEGUNDO ROSENFELD	19
3.1	Preambulo	19
3.2	Introdução	19
3.3	Pré-processamento	20
3.3.1	Modificação de níveis de cinza	20
3.3.2	Suavização	21
3.3.3	Agudização	22
3.4	Segmentação	23
3.4.1	Subpopulações homogêneas	23
3.4.2	Bordas	25
3.4.3	Características locais	26
3.4.4	Segmentação Sequencial e Iterativa	26
4	SEGMENTAÇÃO POR BINARIZAÇÃO SEGUNDO O’GORMAN	29
4.1	Introdução	29
4.2	Limiarização	29
4.2.1	Limiarização global	31
4.2.2	Limiarização local adaptativa	33
4.2.3	A escolha de um método de limiarização	33
5	GLOSSÁRIO DE HARALICK	35

Lista de Figuras

2.1	Paradigma Sinais-Símbolos para a visão computacional: são transformados dados brutos adquiridos em uma descrição da imagem por uma série de passos indutivos.	5
2.2	Processo de aquisição de uma imagem	7
2.3	O Detector de Sobel de Bordas a) Janela do Sobel; b) Aplicando janela do Sobel para uma imagem; c) Exemplo do detector de bordas Sobel usado para encontrar a direção da borda no elemento circulado no array da imagem	11
2.4	O operador de Duda	13
2.5	Inibição Lateral usando o detector "passagem por zero" baseado em diferenças de Gaussianas	14
4.1	Binarização da imagem de um documento com diferentes valores de limiar.	30
4.2	Binarização a partir de funções de densidade de probabilidade Gaussianas.	32

List of Tables

2.1	Requisitos Funcionais para um Sistema de Visão Genérico	4
-----	---	---

Capítulo 1

INTRODUÇÃO

Este documento articula-se em duas partes distintas. Na primeira, apresenta-se uma coletânea de descrições de pesquisadores de fama internacional a respeito das áreas de Visão Computacional e de Processamento de Imagens. A segunda parte é constituída de um glossário dos termos técnicos encontrados nessas duas áreas acima citadas. Nesta parte, as palavras chaves apareceram em português e em inglês para facilitar a busca. Propositamente, alguns termos não foram traduzidos na íntegra por não ter tradução fácil na língua portuguesa.

Capítulo 2

VISÃO COMPUTACIONAL SEGUNDO FISCHLER

2.1 Introdução

Segundo [FF87], a visão computacional (CV) é um sub-campo da inteligência artificial e se preocupa em desenvolver uma compreensão dos princípios que dão sustentação a aptidão visual em sistemas naturais e artificiais, e em proporcionar uma máquina com algumas capacidades do sistema visual humano. Tais capacidades incluem a habilidade para descrever uma cena baseada em dados fornecidos por sensores de imagens, e produzir entendimento da função, objetivo, e intenção dos objetos reconhecidos. A tabela 2.1 resume os requisitos funcionais de um sistema de visão genérico.

Os desafios para a visão computacional são dois:

- O dispositivo de computação deveria ser capaz de simular experiências físicas, como imaginar o movimento ou a reestruturação e a distorção de objetos na imagem para resolver qualquer problema ou ainda comparar a imagem com imagens de referências armazenadas na memória;
- O computador deveria ter algum modo de interagir fisicamente e sentir o mundo externo para construir um banco de dados de conhecimento e experiências. Sem interação física, não há nenhum modo razoável para capturar e armazenar na memória do computador um modelo completo apropriado que reflete toda a complexidade e os detalhes de uma imagem do mundo real.

Em outras áreas da Inteligência Artificial (IA), já observou-se a conveniência da declaração, *"se você é um martelo, tudo se parece com um prego"*. Para visão computacional, isto pode ser parafraseado como, *"se você é um computador digital, então tudo se parece com um número ou um símbolo"*. Então, para um computador digital tratar com o mundo percebido visualmente, os sinais adquiridos por sensores de imagens devem primeiro ser convertidos em números e finalmente em símbolos. Portanto, conduz-se ao paradigma de sinal-símbolos (signal-to-symbols) descrito na próxima seção 2.2. O resto da abordagem de Fischler discute algumas técnicas ¹ envolvidas nas descrições simbólicas obtidas dos sinais. As seções são sequenciais para refletir a complexidade crescente e abstração das técnicas correspondentes, iniciando com as representações de baixo nível e algoritmos, e procedendo pelos níveis

¹Tentou-se selecionar técnicas que são representativas e podem ser entendidas sem a necessidade da apresentação matemática envolvida.

intermediário e mais alto.

Uma diferença notável entre muitas representações de visão computacional e a da Inteligência Artificial geral é que em visão, usa-se freqüentemente listas de elementos de figura (pixels) ou outras representações icônicas que refletem a imagem sentida e assim retêm uma correspondência mais direta para o mundo real. A precisão e adequacidade de quaisquer das representações na hierarquia de sinais-símbolos são julgadas por, o quanto fielmente ela retrata a imagem do mundo real que era sentida originalmente, isto é, a preocupação primária é com a modelagem física do mundo. Isto está em contraste com os sistemas de inteligência artificial convencionais que tipicamente não têm uma componente perceptiva, e assim trabalha dentro de um modelo completo, consistente, e fechado da realidade.

Modelagem	Descrição
Modelagem geométrica	Determinar a configuração tridimensional das superfícies e objetos em uma cena, incluindo a localização do espectador (sensor) com respeito à cena observada.
Modelagem fotométrica	Determinar a localização e a natureza das fontes de iluminação e os correspondentes sombreamento e efeitos de reflectância induzidas em uma imagem da cena.
Segmentação de imagem	Dividir a imagem em sub unidades significativas ou coerentes que podem ser analisadas e identificadas independentemente.
Rtulação	Identificar os objetos visíveis em uma imagem como seja partes das classes de objetos conhecidas, ou seja indivíduos conhecidos. Determina os atributos físicos (tamanho, composição material, etc.) de objetos reconhecidos.
Descrição relacional e argumentação	Determina as relações entre os objetos em uma imagem, por exemplo, a aparência da cena instantes antes que a imagem era adquirida, e como a imagem aparecerá imediatamente em seguida. Como os objetos em uma imagem podem ser reorganizados para alcançar um determinado propósito?
Interpretação semântica	Determinar a função, objetivo, intenção, etc., de objetos em uma imagem.

Tabela 2.1: Requisitos Funcionais para um Sistema de Visão Genérico

As perguntas básicas são:

- Qual é a natureza da descrição simbólica do computador do mundo visual, e como é obtida?
- Quais são as representações usadas no paradigma sinais-símbolos?
- Que algoritmos existem para obter estas representações e extrair suas informações? Como pode-se construir uma máquina que pode reconhecer obje-

tos e pode recriar a geometria da imagem dos dados fornecidos por imagens bidimensionais?

- Quais são as poucas idéias chaves e suposições principais que estão subtendidas na maioria dos atuais algoritmos de visão computacional?

Conclue-se com um olhar crítico ao paradigma de sinais-símbolos, e indicar as exigências que devem ser satisfeitas por um computador se é para alcançar a competência do nível humano na percepção visual.

2.2 O Paradigma Sinais-Símbolos

A Visão Computacional (CV) é o conjunto de teoria e técnicas que representa nosso atual entendimento de como a competência da percepção visual pode ser implementada num hardware de computador. O paradigma dominante, sinais-símbolos, é um dos quais o dado bruto é transformado em uma descrição significativa e explícita da imagem correspondente por uma série de passos indutivos usando progressivamente as representações mais abstratas (figura 2.1). Estes passos podem ser divididos em três categorias, baseado na natureza da modelagem requerida para levar a cabo a análise:

- a análise de baixo nível da cena está baseado nas propriedades de imagens locais,
- a análise de nível intermediário da cena usa modelos geométricos e fotométricos genéricos, e
- a análise de "alto nível" da cena está baseada nas metas orientada dos modelos semânticos e suas relações.

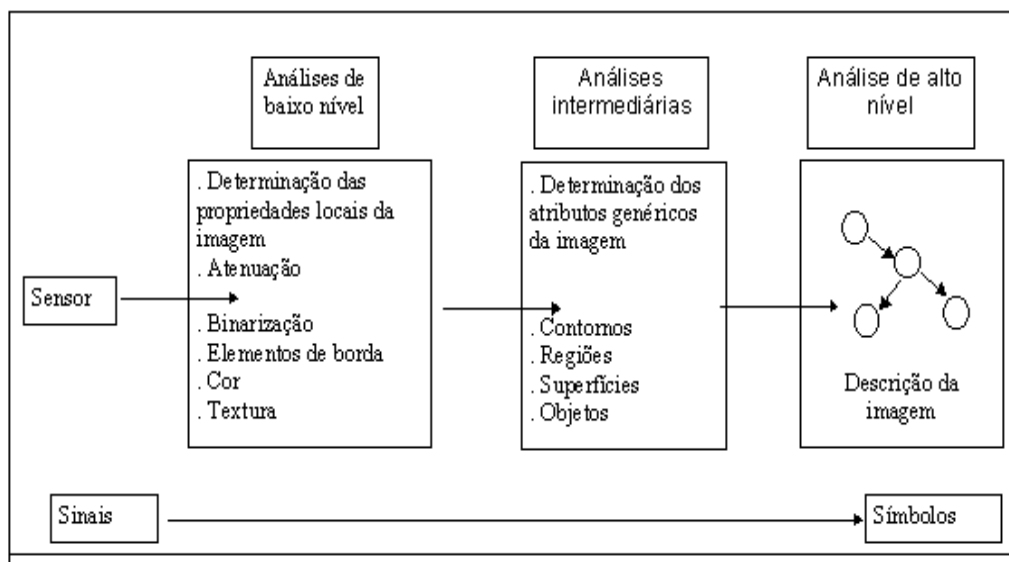


figura 2.1: Paradigma Sinais-Símbolos para a visão computacional: são transformados dados brutos adquiridos em uma descrição da imagem por uma série de passos indutivos.

2.3 Análise de Imagem de Baixo Nível

Nos mais baixos níveis da hierarquia de processamento, as representações e técnicas de transformação tendem a ser independentes de qualquer propósito final.

Preocupa-se aqui com a modelagem física e estatística das propriedades locais genéricas das superfícies visíveis na imagem e sua aparência na imagem. A análise de baixo nível de imagem também se preocupa com os processos de transformar sinais contínuos derivados do sensor em representações digitais discretas, e a redução de ruído e distorção introduzida pelo processo sensor. As entradas para esta fase do processamento são sinais brutos de um ou mais sensor, e a saída é tipicamente um conjunto de listas registradas, com cada lista correspondendo a um atributo particular da imagem tais como a orientação da superfície local, reflectância da superfície, localização da borda, etc...

As técnicas de análise de baixo nível da cena foram desenvolvidas para:

- Reduzir ruído da imagem e detalhe de imagem não desejado sem degradação séria da informação necessária para recuperação de descrição de "alto nível" da imagem .
- Eliminar ambigüidade, isto é, separar as contribuições da iluminação, reflectância da superfície, e orientação da superfície para o brilho de um ponto na imagem (de certo modo, inverter o processo de imagem).
- Detectar homogeneidades locais e discontinuidades que podem ser usados para dividir a imagem em regiões que correspondem a objetos coerentes na imagem.
- Detectar características de imagem locais distintas que são marcadores importantes e delimitadores de características da imagem.
- Deduzir a geometria da superfície local (profundidade tridimensional e orientação) a partir de sombra, textura, análise de estéreo, e a análise de uma sucessão contínua de imagens ("fluxo ótico").

A análise normalmente se trata de fenômeno local e usa modelos baseado em tais conceitos gerais como continuidade (ou discontinuidade) de intensidade, textura, ou cor. Enquanto a análise de baixo nível da cena provê uma interpretação do fenômeno físico do mundo real para uso pelo mecanismos de integração e raciocínio que incluem a análise intermediária e de alto nível da imagem, as técnicas de análise de baixo nível da cena têm em si poucas (se houver alguma) da característica de atributos de raciocínio. As técnicas de análise de baixo nível da cena quase sempre estão baseadas na suposição de continuidade:

- (1) que a imagem foi obtida de uma forma essencialmente randômica no espaço, e nenhuma tentativa deliberada foi feita para fazer coisas separadas e descontínuas, parecerem contínuas na imagem, ou alinhar sombra e boras oclusas, ou fazer uma linha curva parecer reta, etc.; e
- (2) que devido a imagem é composta amplamente de superfícies contínuas, a geometria, luminosidade, e propriedades físicas medidas em qualquer ponto na imagem são bons preditores dos valores que aparecem na vizinhança daquele ponto.

Primeiro, descrever-se-á o processo de aquisição de imagem, e as operações de pré-processamento usadas para melhorar a qualidade e utilidade da imagem. Então, descrever-se-ão os vários métodos para detecção de discontinuidade e de homogeneidade local. A determinação da geometria de uma única ou de múltiplas imagens completará a discussão da análise de baixo nível.

2.4 Aquisição de imagem (escanear e quantizar)

O processo da visão computacional começa com uma imagem já existente (uma foto ou figura) tirada anteriormente, ou com uma cena sendo gravada por uma câmara de tipo TV. Para traduzir a informação de uma cena iluminada numa forma adequada para o computador processar, ela deve primeiro ser convertida numa tabela de números que representa a intensidade de luz refletida em cada ponto da cena.

Se começar com uma imagem, em detrimento da cena atual, pode-se obter os dados de intensidade pela movimentação de um pequeno orifício (ou janela) sobre a imagem, de tal maneira que o nível médio de intensidade de luz da imagem dentro da janela é captado por um dispositivo foto-sensitivo. Este processo de transmissão de imagem pode ser realizado usando um dispositivo mecânico que fisicamente move um sensor sobre a imagem de forma regular e exaustivo, ou pelo menos um Flying Spot Scanner que move um feixe de luz seqüencialmente sobre a imagem e lê a luz refletida (ou transmitida).

O sinal elétrico contínuo que resulta do processo mecânico de transmissão da mensagem deve ainda ser convertido numa matriz de números por amostragem do sinal por intervalos regulares de tempo (correspondendo aos intervalos espaciais regulares sobre a imagem) e por aproximação da voltagem medida pelo inteiro mais próximo num intervalo pré-definido de números, como mostra a figura 2.2. Com a tecnologia atual para armazenamento, processamento e exibição de dados pictoriais, um conjunto de valores para amostragem e digitalização da imagem é tipicamente uma grade de 512×512 pixels com 256 níveis de intensidade possíveis para cada pixel da grade.

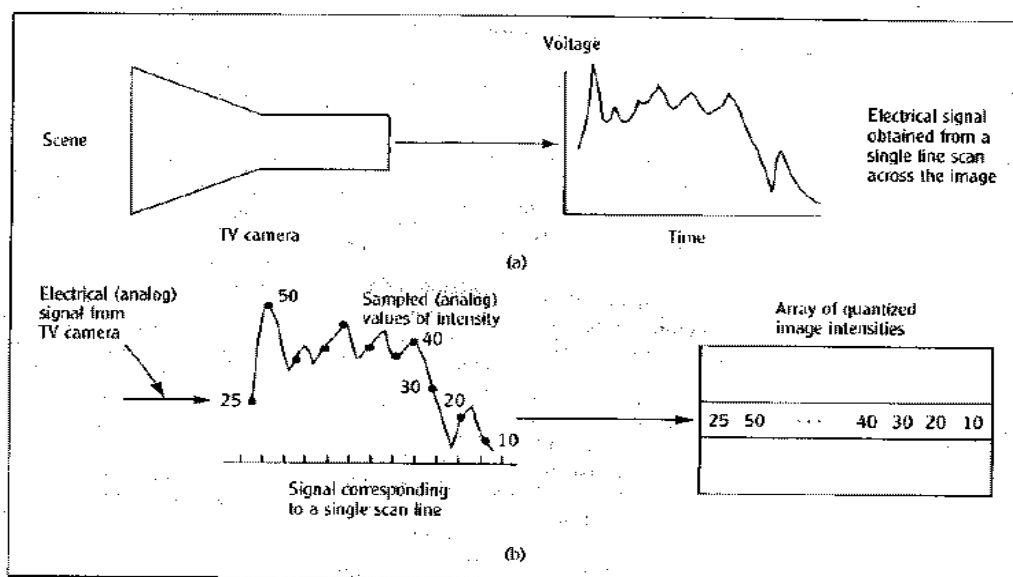


figura 2.2: Processo de aquisição de uma imagem

Um parâmetro importante é a fidelidade com que a aparência da cena atual é capturada pela matriz de números inteiros que é extraída para descrevê-la. Em todas as abordagens de digitalização, a amostragem da intensidade da cena é realizada de forma espacial. A menor distância da imagem que pode ser tolerada entre as amostras de intensidade depende das características do sistema de lentes,

da superfície sensível do sensor e do sistema usado para converter os sinais a partir das quantidades contínuas de forma espacial para quantidades discretas. As intensidades amostradas são no futuro quantizadas em valores digitais para satisfazer parâmetros práticos sobre a memória ou o tamanho do registro da palavra no computador.

Quantas informações são perdidas neste processo de digitalização?, de que modo pode-se reproduzir mais fielmente a imagem original a partir de um número finito de matrizes de números? Espantosamente suficiente, a imagem contínua pode ser exatamente recriada a partir da representação de amostragem (porém não quantizada) se o espaço de amostragem é menor que um valor determinado pela taxa espacial máxima de mudanças das intensidades na imagem.

Quando os valores do sinal analógico amostrados são convertidos para quantidades inteiras, alguma quantidade do ruído quantizada é introduzida na conversão digital-analógico que geralmente não pode ser completamente removido. Porém, um número suficiente de níveis de quantização pode ser escolhido para assegurar que a amplitude do ruído já introduzido pelo processo sensorial não cresce significativamente. Deste modo, conclui-se que o processo de conversão de uma imagem contínua dentro de uma matriz de inteiros não causa alguma perda fundamental de informação.

2.5 Pré-processamento de imagem (binarização e suavização)

O Processamento de imagens trabalha com operações que são relativamente independentes do conteúdo da cena para alterar os valores armazenados na matriz digital representando a cena. Operações de pré-processamento são geralmente projetadas para remover ruídos, aumentar certos aspectos da imagem (margem/ borda) e introduzir outras mudanças que simplificarão os passos de processamento de nível mais alto. Supõe-se que a intensidade das imagens é especialmente contínua sobre toda a imagem, e que esta continuidade pode ser aproximada, por exemplo, por um polinômio de baixa ordem. Uma importante meta em pré-processamento é evitar a eliminação das margens existentes ou introduzir falsas bordas.

As operações típicas de pré-processamento são a binarização, que reduz a imagem quantizada digitalmente contendo muitos valores inteiros de intensidade numa imagem binária contendo somente dois valores possíveis e a suavização que representa o uso das várias operações de filtragem para aumentar ou, certamente, suprimir certos aspectos de uma cena:

- Binarização ou limiarização : a operação de binarização permite particionar a imagem em um prévio estágio na análise, e mais tarde simplificar as etapas posteriores de processamento. O conceito de binarização é simplesmente um: pressupõe-se que todos os pixels em uma região coerente da imagem tem uma intensidade maior que (ou menor que) um certo valor. Um valor de limiar é escolhido, e todos os pixels cujo nível de intensidade está abaixo desse limiar são setados para o valor "preto", e todos aqueles acima desse limiar são setados para o valor "branco". Técnicas de seleção automática de limiar são discutidas no parágrafo 2.5.1.
- Suavização (filtragem) : Operações de suavização
 - (1) removem ruídos e artefatos de iluminação que foram introduzidos na imagem durante o processo de aquisição da imagem;

- (2) realçam as bordas e outras características da imagem; e
- (3) degradam detalhes indesejados abaixo do nível de resolução no qual a interpretação da imagem está sendo realizada.

A suavização é usualmente executada pela substituição de valores de intensidade de cada pixel por um novo valor baseado nos valores de intensidade dos pixels na vizinhança do pixel dado. O problema que ocorre na suavização é como evitar *atirar o bebe fora com a banheira com a água*; isto é, como evitar a eliminação de dados essenciais em experimentos para executar metas mencionadas acima. O procedimento básico é como selecionar subconjuntos apropriados da imagem para processar coerentemente sem cruzar bordas que separam diferentes entidades de cenas. Por exemplo, como a suavização consiste em operações em uma janela centrada em volta do pixel sendo modificado?, como pode-se mantê-lo em bordas borradas quando pixels interiores e exteriores do objeto encontram-se na mesma janela? Não existe nenhum algoritmo perfeito de suavização. Algumas das abordagens comuns para a suavização e as suposições implícitas feita por cada uma estão sendo descritas no parágrafo 2.5.2.

2.5.1 Binarização de imagem

A binarização de imagem transforma uma imagem em níveis de cinza, cujos pixels podem ter qualquer valor de intensidade, em uma imagem binária na qual cada pixel é ou preto ou branco.

A binarização é uma simples forma de divisão da imagem, aumenta certas características da imagem, e simplifica as etapas posteriores de processamento. Em algumas situações, diferentes limiares podem ser usados em diferentes porções da imagem para compensar certas variações de iluminação conhecidas ou deduzidas ou mudar o contraste local da cena. Uma pessoa, interativamente ajustando o valor do limiar e visionando o efeito de cada valor de limiar, pode escolher um valor de limiar que melhor atinja o efeito desejado. A seleção automática do limiar é normalmente baseada nas seguintes técnicas:

- Efeito na imagem: um procedimento iterativo para a seleção do limiar pode ser baseada no número, na área e na estabilidade das regiões geradas por diferentes limiares - um bom valor de limiar deverá produzir principalmente regiões largas e bem separadas que conservam suas formas mesmo com pequenas variações do valor de limiar. Um mapa de contorno pode ajudar a visualizar os efeitos dos diferentes limiares no particionamento final da imagem.
- Análise de histograma : A intensidade do histograma é um gráfico cujo eixo x mostra todos os valores possíveis de níveis de cinza, e cujo eixo y mostra o número de pixels na imagem que tem cada uma destas intensidades. Um limiar apropriado para uma imagem pode automaticamente ser determinado pela análise da forma de um histograma . Isto freqüentemente supondo que picos individuais do histograma correspondem a regiões coerentes (intensidade relativamente constante) e que o fundo (segundo plano) é a região mais clara ou mais escura de todas elas.

2.5.2 Suavização de imagem

A suavização de imagens é empregada para reduzir o ruído, para realçar as características da imagem e para reduzir detalhes indesejados. A maioria das técnicas de suavização pode ser classificada em três famílias: mediação local, suavização baseada em modelos, suavização geométrica.

- Suavização baseada em mediação local : A suavização baseada nas operações de mediação local pressupõe que a superfície de intensidade sobre toda a imagem. Essa hipótese é um caso especial da suposição mais geral que a maioria das imagens descreve a superfície de cenas contínuas numa resolução de imagem adequada para a interpretação ser possível. Em uma abordagem de suavização baseada nessa concepção, troca-se o centro de uma pequena região em volta de um pixel (tipicamente um janela quadrada) pela média ponderada dos valores encontrados nessa janela; essa operação é idêntica para cada pixel da imagem. Outra abordagem usada para evitar o efeito de pixels indesejados e para reter bordas consiste em usar o filtro da mediana, no lugar da média, dos valores de intensidade da janela.
- Suavização Baseado em Modelos A Priori : A técnica seguinte é típica de uma abordagem global para suavizar. Suponha que tenha-se um modelo de como a iluminação varia em uma imagem. Por exemplo, se sabe-se que a iluminação pode ser modelada como uma função quadrática, pode-se ajustar uma superfície quadrática nos valores de intensidade dos pixels em uma porção estendida da imagem. A intensidade de cada pixel é então subtraída do valor correspondente da superfície ajustada, deixando somente as variações de grande ordem do sinal subjacente. Assim impede-se que um artefato conhecido interfira na análise da informação intrínseca que reside na imagem.
- Suavização Geométrica : A suavização geométrica de uma imagem pode ser realizada assumindo-se que pequenas regiões isoladas consistem em ruídos e podem ser eliminadas e que pequenos furos (intervalos) entre regiões são artefatos da imagem que podem ser preenchidos. Tal suavização pode ser prontamente realizadas em imagem binárias usando seqüências de operações de emagrecimento ("shrinking") e crescimento. Nesta abordagem de emagrecimento e crescimento para eliminação de pequenas regiões ruidosas, usa-se em primeiro a operação de emagrecimento na qual pixels pretos que não estejam completamente cercados de pretos são setados como brancos. A operação de emagrecimento pode ser realizada de forma interativa muitas vezes. Uma operação de crescimento pode então ser usada na qual todos os pixels pretos que não são completamente cercados por pretos são providos com pixels pretos ao redor. Qualquer pequena região de ruído preta terão sido eliminadas pelas operações de emagrecimento, e as maiores regiões serão mantidas inalteradas se os números de operações de emagrecimento e crescimento forem iguais. Se uma seqüência de operações de crescimento for aplicada primeiramente (seguida por operações de emagrecimento) pequenos intervalos nos objetos pretos ou entre objetos adjacentes serão preenchidos, mas as formas dos objetos em geral permanecerão inalteradas.

2.6 Detecção de descontinuidades e homogeneidades locais (Borda, textura, cor)

Nenhuma descrição ou interpretação semântica razoável é possível se cada pixel da cena não está relacionado com seus vizinhos. Porém, a maioria das cenas que encontra-se, podem ser decompostas em objetos coerentes ou regiões que são relativamente homogêneas com respeito a um ou mais dos seguintes atributos: intensidade, cor, textura, distância, movimento, composição material, coesão física, etc. Correspondentemente, há lugares em uma imagem onde as descontinuidades são nítidas em alguns dos atributos acima; por exemplo, através da borda que sobrepõe outra parte da cena. Identificando regiões homogêneas e vastas descontinuidades

2.6. DETECÇÃO DE DESCONTINUIDADES E HOMOGENEIDADES LOCAIS (BORDA, TEXTURA, COR,

simplifica o problema de análise da imagem em dois pontos importantes:

- descontinuidades são associadas tipicamente com as bordas dos objetos, e a posição das bordas fazem explicitamente o objeto nítido.
- identificando porções da imagem correspondente para objetos coerentes permite analisar estas porções isoladamente, se for necessário.

Alguns métodos semelhantes para encontrar descontinuidades e homogeneidades são dados abaixo. Algumas das técnicas são independentes da situação dos objetos dentro da cena, enquanto outros assumem certas características específicas dos objetos do mundo real.

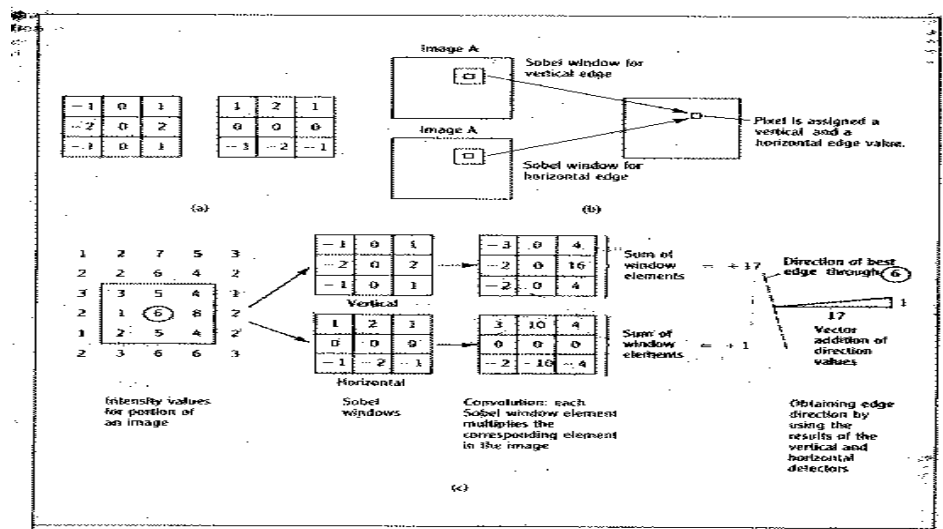


figura 2.3: O Detecção de Bordas de Sobel a) Janela do Sobel; b) Aplicando a janela do Sobel para uma imagem; c) Exemplo do detector de bordas Sobel usado para encontrar a direção da borda no elemento circulado no array da imagem

2.6.1 Detecção da Borda Local

Há muito tempo que a detecção de bordas dos objetos que aparecem em uma imagem é um passo essencial na análise da cena, e para esta razão é preciso considerar dedicado esforço para o desenvolvimento de algoritmos eficazes de detecção de bordas. Uma classe desse tipo de algoritmos, detector de bordas locais, determina um valor de borda para os pixels individuais, mas não ligam pixels juntos para formar um segmento de borda estendido. Então, uma associação adicional ou passo unindo ainda deve ser levado em consideração no computador para obter uma representação interna do segmento de borda conectada. Unindo bordas é uma operação de nível imediato, e será descrita posteriormente. Algoritmos desse tipo podem ser agrupados dentro das seguintes categorias:

Operadores de gradiente locais: Estes algoritmos estão baseados na caracterização de uma borda como uma descontinuidade de intensidade local. São examinadas as intensidades em uma região local da imagem e um valor da borda (e algumas vezes a orientação da borda) é associado a um elemento de imagem baseado na mudança de intensidade dentro daquela região local.

O gradiente mais simples e mais comumente usado tem as seguintes características (por exemplo, o detector de bordas Sobel: ver figura 2.3): uma convolução entre a imagem e um conjunto de máscaras digitais (por exemplo, um quadrado de 3×3 pixels) é efetuada. Desta maneira cada máscara do operador avalia a intensidade do gradiente em uma direção particular na posição da imagem correspondente ao centro da máscara do operador (figura 2.3). Um gradiente aproximado pode ser calculado por usar o operador em duas direções ortogonais e empregando a adição de vetor.

Resultados melhores podem ser obtidos executando a operação de um grande número de orientações angulares e seleccionando o valor máximo obtido a magnitude do gradiente numa posição dada, e a direção correspondente como a direção de gradiente. Pixels da borda são determinados por limiarização da imagem do gradiente.

O gradiente mais simples, como descrito acima, ignora certo número de considerações relevantes a respeito das imagens reais. Primeiramente, a maioria das imagens reais tem gradientes suaves que são artefatos da iluminação e dos processos de imagem. Não é possível definir um limiar absoluto em um gradiente local (por exemplo, como requerido pelo gradiente da Sobel) para distinguir as descontinuidades intensas (bordas) desses artefatos dos gradientes suaves. Deste modo depara-se com o problema de como detectar bordas de baixo contraste (usando um operador de gradiente com limiar fixo) sem sermos inundados por falsos alarmes que surgem dos gradientes suaves. Há também o problema na qual descontinuidades de intensidades correspondendo de tamanho à bordas diferentes em uma imagem podem variar (resoluções); Escolher uma máscara de tamanho simples (mesmo com múltiplas orientações) não é adequado. O parágrafo 2.6.2 apresenta uma abordagem para detectar bordas locais que é mais ágil para tratar desses problemas.

Ajustamento Genérico de Modelos: Estes algoritmos são baseados no modelo de uma borda como um perfil específico de intensidades. Dentro de alguma área de busca local, uma escolha mais adequada é efetuada para contornos específicos. A abordagem do modelo genérico de ajustamento é muito específico do tipo de descontinuidades processadas, em contraste com a aproximação de gradiente local que satisfaz uma grande variedade de tipos de descontinuidades de intensidade.

É freqüentemente aceitável descrever uma borda como sendo uma descontinuidade de intensidades de tipo degrau geometricamente reto sobre alguma extensão local da imagem. Neste caso, pode-se mover uma janela pequena sobre a imagem e encontrar os melhores ajustes em cima do modelo para o padrão de intensidade visto através da janela a cada uma de suas paradas. O detector de bordas Hueckel é o mais comumente empregado operador deste tipo. Aceita a digitalização por intensas luzes em uma pequena subárea em forma de disco (contendo no mínimo 32 pixels) e produz uma descrição da borda mais provável (em termos de descontinuidade de brilho) encontrada dentro do disco.

Ajustamento Semântico de Modelos: Os algoritmos baseados nos "modelos semânticos de bordas" usam características específicas dos objetos de interesse para descobrir as bordas. Por exemplo, vários objetos como estradas ou rios em uma fotografia aérea, ou costelas em um filme de radiografia médica, tem propriedades de bordas que são dependentes da natureza dos próprios objetos. Os algoritmos desta categoria são elaborados para procurar as bordas de uma classe particular de objetos. Por exemplo, para detectar estradas em

2.6. DETECÇÃO DE DESCONTINUIDADES E HOMOGENEIDADES LOCAIS (BORDA, TEXTURA, COR,

uma imagem aérea com resolução baixa, o operador de bordas de Duda (figura 2.4) especificamente requer que o fragmento da estrada tenha valores de intensidade e de largura relativamente constantes ao longo de um segmento de linha. Além disso, o requisito genérico para a direção do segmento é utilizar o gradiente de alto valor normal à direção do segmento.

O operador de bordas de Duda é proposto especificamente para detectar estradas, produz intuitivamente mais resultados óbvios que outros mais genéricos operadores de bordas.

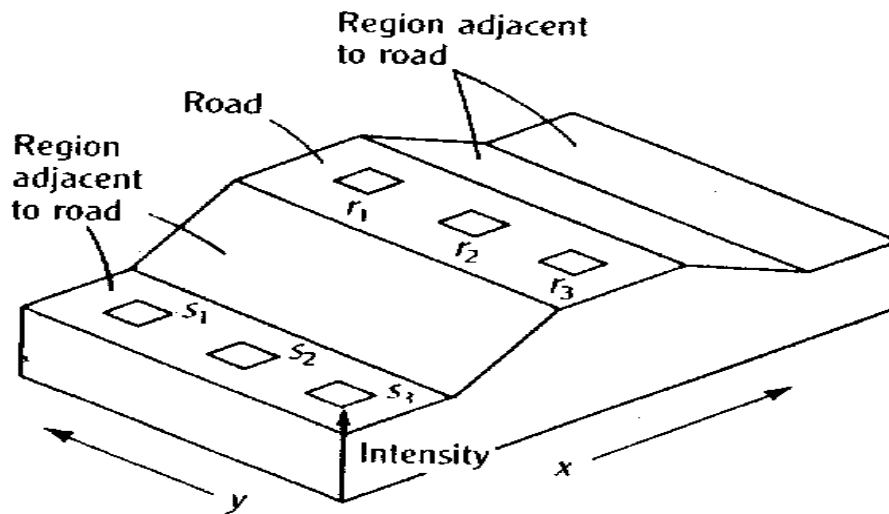


figura 2.4: O operador de Duda

2.6.2 Detecção de Borda Local Baseada em Inibição Lateral

Os sistemas visuais orgânicos universalmente empregam um mecanismo, chamado de inibição lateral que oferecem um contraste aos problemas de detecção de bordas encontrados através do uso do tipo de gradiente simples.

Computacionalmente a inibição lateral envolve a "assinatura" de borda de um pixel como sendo a diferença ponderada das médias das intensidades entre duas máscaras de tamanho diferentes centradas no pixel. Em algumas implementações simples, cada máscara é uma caixa retangular uniformemente ponderada. Em versões mais sofisticadas, e especialmente como encontrado em sistemas biológicos, as máscaras têm uma forma de curva gaussiana, apresentando uma distribuição adequada de valores em vez de uma distribuição uniforme de valores (figura 2.5). Aplicando um tal operador para uma região de uma imagem, na qual há um gradiente uniforme, não importa quanto forte, resultará em todos lugares em uma resposta zero (assumindo que a soma dos pesos nas duas máscaras são iguais). Se há uma descontinuidade aguda de intensidade sobreposta ao gradiente uniforme, então como o operador é descontinuamente movido ao longo de um caminho normal, este valor resultará em zero até a maior máscara (exterior) cruzar a descontinuidade. Então o valor retornado pelo operador aumentará até que a menor (centro) máscara intercepta a descontinuidade. Neste ponto o valor retornado pelo operador diminuirá rapidamente, propriamente zero quando a máscara menor for centrada exatamente na descontinuidade. Continuando, obtem-se agora uma resposta simétrica que aproxima a descontinuidade, mas com sinal invertido. A borda é detectada localizando a

"passagem por zero" da resposta do operador em cima de alguma porção da imagem (não só um valor zero). A força e a direção da borda devem ser determinadas analisando a resposta do operador no justicamento da passagem por zero.

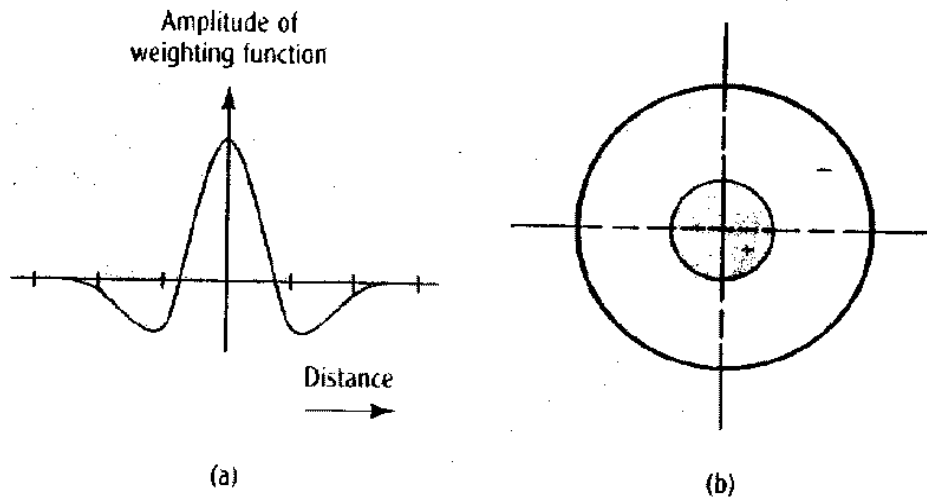


figura 2.5: Inibição Lateral usando o detetor "passagem por zero" baseado em diferenças de Gaussianas

Para determinar discontinuidades de intensidade muito agudas em uma imagem essencialmente livre de ruídos, a máscara central pode ser um pixel em diâmetro, e a máscara externa tem que ter um diâmetro ligeiramente maior: porém, se a discontinuidade de intensidade é "borrada" para vários pixels, e se a imagem é ruidosa, então a máscara exterior tem que ter um diâmetro maior que a largura da região de transição de extremidade. De fato, a melhor resposta em termos da magnitude e rampa da reversão na qual a passagem por zero acontece será obtida se a máscara exterior é feita tão grande quanto possível sem fazê-lo tão grande que cobre mais que uma borda. Aumentando o diâmetro da máscara central suaviza a resposta do operador, mas também diminui a amplitude e o inclinação da seção na qual nós estamos procurando a passagem por zero. Um assunto complexo é escolher um ótimo tamanho para a máscara central, mas provavelmente não deverá ser maior que o diâmetro da transição de borda. Para lidar com uma cena complexa, um jogo de operadores da passagem por zero com tamanhos graduados da máscara central é requerido para descobrir bordas de larguras variadas.

2.6.3 Análise da Homogeneidade local

A identificação de regiões homogêneas é feita geralmente rotulando primeiramente cada pixel da imagem com valores de atributos como por exemplo a textura e a cor.

Analisando a textura: Embora há muitos aspectos computacionais e biológicos de percepção e detecção de textura, só preocupa-se aqui com a pergunta de como a análise de textura pode ser usada para dividir uma imagem em regiões homogêneas. Considerando que o sistema visual humano pode reconhecer facilmente tipos diferentes de texturas, pode aparecer como surpreendente saber que não há nenhuma definição geralmente aceita de textura e assim nenhum consensus de como caracteriza-la. Não é fácil de caracterizar formalmente a base de nossa percepção de textura descrita em termos de adjetivos

2.6. DETECÇÃO DE DESCONTINUIDADES E HOMOGENEIDADES LOCAIS (BORDA, TEXTURA, COR,

como "fino", "grosso", "macio", "granuloso", "randômico", "mosqueado", etc. Sabe-se intuitivamente que a textura envolve uma relação estatística ou estrutural entre os elementos básicos. E para texturas figurativas (celular, macroestruturas) como uma parede de tijolo ou um chão ladrilhado, o sistema visual pode descobrir os padrões subjacentes que compõem o desenho da textura, e pode-se descrever a relação dos elementos. Para textura de microestrutura, como os campos vistos em fotografias aéreas, ou a textura de pano, os padrões subjacentes não são tão evidentes e é difícil para o humano descrever a textura (mas não para reconhecer!). A análise de micro e macro texturas estruturadas é descrito no parágrafo 2.6.4

Cor: Uma imagem colorida é representada tipicamente no computador como três matrizes distintas de números, cada matriz corresponde a uma cor primária, e o valor de um pixel representa a componente correspondente da cor primária na posição correspondente. Então cada elemento da imagem é representado por três valores. Embora todas as três imagens que são obtidas simultaneamente usam um vidicon de cor normalmente, as matrizes também podem ser obtidos percebendo a cena três vezes, cada tempo por um filtro de cor diferente. Às vezes o componente espectral (cor) de uma imagem é relativamente constante em cima de grandes áreas da superfície de um único objeto, enquanto a intensidade total pode mudar mais irregularmente devido a iluminação direta desigual e as reflexões locais. Assim pode ser usada a cor para ajudar achar regiões homogêneas em uma imagem agrupando pixels vizinhos com atributos de cor semelhantes. Erros ainda podem acontecer devido ao fato que objetos podem ser coloridos por reflexos de outros objetos coloridos, porque a suposição básica de homogeneidade sempre não é válida.

2.6.4 Análise de texturas micro e macro estruturadas

A análise de textura é uma operação de base em segmentação de imagens e consiste em para formalizar a noção intuitiva da aparência da superfície numa imagem.

- Textura micro estruturada: Existem duas abordagens para analisar a textura microestruturada:
 - análise com base na regularidade da microestrutura detectada por técnicas de força espectral ou estatística.
 - uso de técnicas que modelam a superfície original que produzem os padrões de texturas.
- Abordagem por Características estatísticas: Porque as unidades básicas da microestrutura são pequenas, técnicas que detectam regularidade em pequenas seqüências de pixels podem ser usadas para particionar cenas microtexturizadas. A estratégia básica é formar um espaço característica baseado nas medidas de suas vizinhanças sobre cada pixel na imagem. A segmentação de uma imagem pode ser adotada designando pixels para uma ou outra região com base na localização daquele pixel no espaço característica.
- Abordagem por Co-ocorrência: As relações espaciais em tons de cinza podem ser expressados como $S(i, j|d, A)$, o número de vezes que um pixel de intensidade i aparece numa distância de d pixels e um ângulo A de um pixel de intensidade j , em alguma vizinhança do pixel a ser classificado. Um pode usar funções de S como os componentes de um espaço característica. A *energia* é algo típico para estas funções, formado

pela soma do quadrado de S sobre todos i e j , para valores dados do ângulo A e de distância d .

- Abordagem por Análise de Fourier: Regularidades em padrões de tons de cinza aparecem na transformada de Fourier calculada em varias direções no pixel. O conjunto de medidas de Fourier, obtido pela convolução de um conjunto de janelas ponderadas sobre a imagem, então forma os componentes de um espaço de características. Cada pixel na imagem tem um conjunto associado de medidas de energia de Fourier, e pode ser representado no espaço de características.
- Abordagem por Modelagem: Em uma abordagem por processo de modelagem para análise de textura, uma tentativa para descrever coisas no mundo em termos de como elas aparecem, por exemplo, feitas pelo homem, coisas que crescem, ou processos de desgaste (como em um canyon). Usando este ponto de visão é possível prever como as superfícies naturais irão produzir os padrões de textura em uma imagem. Uma técnica baseada em funções de fractal pode modular texturas de imagem surgindo de processos físicos que alternam o terreno via uma sequência de trocas pequenas; a imagem correspondente vem a ter propriedades estatisticamente medidas que são invariantes sobre transformações lineares de intensidade de transformações de escala. A dimensão fractal, D de uma superfície corresponde intuitivamente a nossa noção de regularidade. É possível medir a dimensão fractal dos dados da imagem se a superfície tridimensional correspondente é áspera ou suave. Esta informação pode ser usada para particionar a imagem em regiões de características de superfícies homogêneas. A dimensão fractal é calculada para cada 8×8 blocos de pixel em uma imagem, e um histograma das dimensões fractal é calculado para a imagem inteira. Este histograma é então quebrado nos vales entre os modelos do histograma, e a imagem é segmentada em imagens pertencentes a um modelo ou outro.
- Textura macro estruturada: Técnicas para tratar texturas macroestruturadas tem encontrado um sucesso limitado. A análise de texturas macroestruturadas é bastante difícil porque uma necessidade de identificar ambas a(s) primitiva(s) e o relacionamento espacial entre elas. Outra dificuldade encontrada, são os efeitos de perspectiva que mudam o tamanho e a forma das primitivas dependendo de sua posição em uma imagem bidimensional.

2.6.5 Geometria de uma cena local a partir de uma única imagem (forma a partir do sombreamento e textura)

Sabemos que a sombra e a textura presentes em uma única imagem podem produzir uma impressão nítida de estrutura tridimensional.

Qual é a base computacional deste efeito?

Uma fonte crucial da informação sobre a estrutura tridimensional é proveniente da distribuição espacial do conteúdo da superfície em uma imagem. Já que a projeção distorce a geometria da textura de uma maneira que depende sistematicamente da forma e orientação da superfície, isolando e medindo esta distorção projetada na imagem pode-se restabelecer a estrutura tridimensional da superfície da textura. Isto não é uma tarefa direta porque as superfícies orientadas estão misturadas na imagem com as características da textura original na qual a distorção age.

2.6. DETECÇÃO DE DESCONTINUIDADES E HOMOGENEIDADES LOCAIS (BORDA, TEXTURA, COR,

Se a textura é simples e regular, tal como de um tijolo quadrado padrão, a mudança na forma dos retângulos na imagem pode ser medida para deduzir a forma da superfície. Contudo, na maioria das vezes, não existe um padrão simples de textura. Uma técnica eficaz para recuperar a orientação das superfícies das imagens em geral deve basear-se nas descrições da textura usada, que pode ser computada a partir de tais imagens. Uma hipótese natural é que as texturas não contribuem para gerar os efeitos projetados nem para cancelar estes efeitos. Logo é razoável assumir que o que parece ser uma distorção projetada realmente é uma distorção.

É muito provável que as variações observadas são atribuídas a projeções - as orientações das superfícies que melhor explicam os dados em uma imagem é a melhor posição para a atual orientação da superfície na cena.

A aparência das referências da superfície da imagem é assunto para dois tipos de distorções geométricas:

- Quando o objeto recua do observador, suas referências aparecem menores (efeito do rastro do trilho do trem); e
- quando uma superfície está inclinada em direção ao plano frontal, estas referências aparecem reduzidas em perspectiva ou comprimidas na direção da inclinação (o círculo é projetado como uma elipse).

Então, qualquer método para recuperar a superfície orientada a partir da textura deve ser expressa em termos de alguma descrição concreta da textura da imagem que seja sensível para esses dois tipos de distorção.

Quando uma textura plana é visualizada em uma orientação desconhecida, a textura original e a orientação do plano com relação aos observados não podem ser duplamente descobertas a partir da imagem. Contudo é possível produzir um conjunto de reconstruções empregando uma transformação inversa de todos os valores de inclinação do ângulo, associado com uma particular orientação da superfície plana.

O problema de redescobrir a orientação da superfície pode ser reformulada dentro de uma escolha da melhor parte do conjunto possível de reconstruções, ordenando as reconstruções por algum critério de semelhança. Por exemplo o ordenamento pode ser baseado na hipótese que todas direções da borda são igualmente como na cena. Primeiro, os pixels da borda e suas orientações na imagem são encontrados, e então, a melhor combinação da inclinação dos ângulos do plano na qual estas bordas se projetam para satisfazer o princípio "randômico da direção da borda" é encontrada. A técnica planar pode ser estendida para superfícies curvas, encontrando a estimativa do plano para uma região circular ao redor de cada ponto da imagem. Repetido sobre a imagem, este método possibilita estimar a orientação da superfície local mas esta validade depende da hipótese randômica, a qual é frequentemente violada.

O homem pode observar superfícies tridimensionais convoluídas da base das distorções projetadas nas texturas complexas e tênues, a visão computacional está distante de conseguir tal estabilidade.

2.6.6 Geometria de uma cena local a partir de múltiplas imagens (estereoscopia e fluxo ótico)

Todas as técnicas anteriormente descritas, operam utilizando uma única imagem.

Abaixo discutir-se-á duas técnicas: estereoscopia e fluxo ótico, que recuperam a cena baseando na análise de uma seqüência de imagens .

- estereoscopia: A visão estereoscópica permite um modelo tridimensional de uma cena ser derivado a partir de dois sensores que observam a cena de diferentes pontos de vista.

A diferença relativa na posição dos objetos em duas imagens é chamada de disparidade , e é causada pela pequena diferença no ângulo de um objeto qualquer dado em relação a cada sensor. Uma suposição embora desconhecida, considera que o nosso cérebro mede estas disparidades e estima as distâncias absolutas entre os objetos e o observador.

A partir de experimentos empregando imagens sintéticas compostas de pontos randômicos, sabe-se que o homem não depende da detecção das características reconhecíveis em cada imagem individual para unir um par estéreo de imagens para recuperar a profundidade da informação. Em computação visual, um par estéreo de imagens é obtido usando duas câmeras separadas ou movendo uma única câmara para dois locais.

Um passo difícil e crítico é determinar correspondências entre pontos nas duas imagens, logo a disparidade pode ser determinada. Quando isto for reduzido, a análise geométrica pode ser usada para computar a localização tridimensional de pontos na cena.

- Fluxo ótico: Supondo que tem-se um sensor de imagem movendo-se através de uma cena. Enquanto o sensor avança, os pontos da cena movem-se no plano da imagem através das curvas, isto é conhecido como curvas de fluxo ótico. Analisando estas curvas do fluxo, é possível determinar a distância dos pontos na cena para o sensor.

Estas distâncias podem então ser usadas para construir um modelo tridimensional da cena exatamente como no caso do emprego da análise estéreo.

Se o sensor está avançando com movimento de translação pura, as curvas do fluxo ótico serão linhas retas que convergem ao ponto conhecido como foco de expansão (*FOE*). Se o *FOE* pode ser localizado no plano da imagem, e a distância movida pelo sensor dquadro a quadro é conhecida, então a distância de um ponto na cena para o sensor pode ser encontrada. Se somente é conhecido que o movimento do sensor era constante, mas não quão distante o sensor move, então a profundidade relativa dos pontos na cena pode ser obtida.

Capítulo 3

RECONHECIMENTO DE PADRÕES SEGUNDO ROSENFELD

3.1 Preâmbulo

A meta geral de reconhecimento de padrões de imagens é gerar descrições de imagens, e relacionar essas descrições a modelos que caracterizam classes de imagens. A visão de Rosenfeld sobre reconhecimento de padrões de imagens permite fazer uma revisão do assunto de um ponto de vista técnico. Tópicos abordados incluem pré-processamento, registro e comparação, segmentação em partes, representação das partes e descrição em termos das propriedades de e relacionamentos entre as partes.

3.2 Introdução

Em processamento de imagens, a entrada e a saída são ambas imagens, com a saída uma versão melhorada da entrada.

Em reconhecimento de imagens, a entrada é uma imagem, mas a saída é uma descrição dos conteúdos daquela imagem. Reconhecimento de imagens tem muitas aplicações; a relação seguinte indica algumas das áreas principais de atuação:

- processamento de documentos: reconhecimento de impresso ou escrita caracter;
- automatização industrial: assembly (robótica), inspeção;
- medicina e biologia: citologia, radiologia, etc., por exemplo, contagem de células sanguíneas, detecção de tumores;
- sensoriamento remoto: agricultura, uso de terra, monitoramento ambiental, meteorologia, etc.

Em qualquer área de aplicação, as fases no processo de análise são basicamente semelhantes. A imagem de entrada é freqüentemente "pré-processada" para, por exemplo, corrigir ou unificar seus níveis de cinza, para eliminar o aspecto borrado, ou remover ruídos, tudo antes de começar a analisá-la. Se for necessário comparar duas ou mais imagens, elas podem ter que ser registradas, i.e., transformadas geometricamente para fazê-las congruentes, principalmente para compará-las. Descrições

normalmente se referem a partes de uma imagem, não somente para a imagem como um todo; assim a segmentação em partes é freqüentemente um pré-requisito para a descrição. As partes podem ser representadas de vários modos para facilitar análises adicionais - por exemplo, objetos compridos podem ser representados por seus esqueletos ou regiões pelas suas bordas. Propriedades das partes, descrevendo suas formas, texturas, etc., e relações entre as partes, podem ser então extraídas. O resultado deste processo, em geral, é uma estrutura de relacionamento que representa partes, propriedades, e relações; no caso especial de estarmos descrevendo a imagem como um todo, sem referência para partes, esta estrutura se reduz a uma lista de valores das propriedades de toda a imagem. Se tem-se um modelo que geralmente descreve uma classe de imagens, reconhece-se a determinada imagem como pertencendo àquela classe estabelecendo uma correspondência entre a descrição da imagem e o modelo.

Este paper trata de reconhecimento de imagem de um ponto de vista técnico. Enfatiza os processos envolvidos na geração de uma descrição, mas não discute o problema de controlar estes processos em um modo extensivamente dirigido as metas. A discussão neste paper estará em um nível conceitual; só serão usadas formulações matemáticas quando necessário para a definição precisa ou concisa de um conceito. As seções restantes neste paper tratam de pré-processamento, registro e comparação, segmentação em partes, representação das partes e descrição em termos das propriedades de e relacionamentos entre as partes.

3.3 Pré-processamento

Esta seção discute brevemente três classes de operações de pré-processamento: modificação de níveis de cinza, agudização e suavização. Claro que, estas operações se relacionam mais com o processamento de imagem do que com o reconhecimento; mas sua importância como passos preliminares no processo de reconhecimento justifica sua inclusão aqui. Considera-se aqui só algumas técnicas relativamente elementares; estudos baseados em modelos matemáticos, como em restauração de imagem e estimativas não serão discutidos.

3.3.1 Modificação de níveis de cinza

A classe mais simples de operações de pré-processamento envolve ajustes de níveis de cinza que são aplicados pixel-a-pixel. Um tipo de ajuste é a variante de espaço, e é usado para compensar as variações de espaço na iluminação da cena ou na resposta do sensor; tal compensação é importante se quer-se classificar os pixels baseados nos seus níveis cinzas (ou cores, etc.). Outro tipo de ajuste é o espaço invariante, e pode ser usado para normalizar brilho e contraste da imagem, modificando os níveis de cinza necessários para medir propriedades da imagem dependentes dos níveis de cinza ou para "casar" ou emparelhar duas imagens.

- Correção de níveis de cinza: O mapeamento do brilho da cena para imagens em níveis de cinza varia freqüentemente de ponto para ponto, devido a variações na resposta do sensor óptico. Se puder determinar estas variações usando imagens de cenas de brilho uniforme, poderia-se (em princípio) compensar seus vizinhos. Em particular, se o nível de cinza observado $g(x, y)$ pode ser expresso na forma $e(x, y) \times f(x, y)$, onde $f(x, y)$ é o nível cinza ideal, pode-se determinar o $e(x, y)$ de imagem de uma cena uniforme, e então obtém f a partir de g dividindo por e .

Um problema mais difícil é o de corrigir a própria informação do brilho da cena. O brilho em um ponto da cena é o resultado de vários fatores, inclusive a

iluminação e a refletividade daquele ponto, e os ângulos entre a luz incidente e a luz refletida que chegam ao sensor. Se os efeitos direcionais são desprezíveis, pode-se escrever $f(x, y) = i(x, y) \cdot r(x, y)$, onde i é iluminação e r é a refletância. Se i varia lentamente, como é freqüentemente o caso, pode-se reduzir seus efeitos aplicando um operador Laplaciano em f para remover variações preservando as bordas, e usando então o operador "Laplaciano inverso" para construir uma função constante tendo as bordas dadas. Alternativamente, pode-se usar a abordagem do filtro homomórfico: use $\log(f) = \log(i) + \log(r)$ para pôr a combinação de i e r em forma de elemento aditivo; use uma filtragem de alta-freqüência reduzir os efeitos de i ; e então use o antilog.

- Transformação em níveis de cinza: Uma transformação arbitrária (espaço-invariante) da escala de cinza tem a forma $z' = h(z)$, onde z e z' são os velhos e os novos níveis cinzas, respectivamente. Pode-se aumentar o contraste da imagem em um determinado conjunto de nível de cinza, projetando a transformação para ampliar os níveis cinzas naquele conjunto. Note que em geral, isto implica que outros conjuntos podem ser comprimidos. Suponha, contudo, que, como é comum nestes casos, os níveis de cinza em um certo conjunto R ocorram com freqüência, enquanto aqueles fora de R são raros. Se ampliar R e comprimir o resto dos níveis de cinza, estar-se-a aumentando o contraste para a maioria da imagem.

Também pode-se transformar a escala de cinza de uma imagem para gerar um histograma específico (= indicação de com que freqüência cada nível de cinza ocorre). Em particular, imagens são freqüentemente transformadas para dar histogramas planos, com cada nível de cinza ocorrendo igualmente. Note que quando faz-se isto a uma imagem na qual um certo intervalo de níveis cinzas é freqüente (e assim dá origem a um pico no histograma), está ampliando aquele conjunto e comprimindo em outro lugar, de forma que a transformação plana gera um efeito de realce de contraste. Para imagens com histogramas de múltiplos picos, pode ser mais apropriado converter cada pico de histogramas em um pico, ao invés de um plano; com efeito, isto segmenta a imagem mapeando cada subpopulação de nível de cinza em um único valor representativo.

3.3.2 Suavização

Se uma imagem foi degradada pelo ruído e tem-se um modelo para o processo de ruído, pode-se determinar estimadores ótimos para os níveis cinzas da imagem original em termos das ruidosas; este assunto também pertence ao tópico de restauração de imagem. Nesta seção considera-se processos de suavização ou smoothing que envolvam operações locais simples executadas na imagem ruidosa. Em particular, será discutido métodos de detecção e remoção de ruídos; o uso de técnicas de média para reduzir o ruído, incluindo esquemas de média seletiva e ponderada que são projetados para evitar bordas borradas; e técnicas baseadas em estatísticas de ordem local (filtro pela mediana, filtro min/max). Remoção ou redução de ruído é um processo preliminar à segmentação de imagem que deve ser feito para reduzir erros de segmentação.

- Remoção de ruído: Se o ruído pode ser distinguido do sinal, isto torna relativamente fácil removê-lo mantendo intacto o sinal. Dois exemplos simples desta situação se seguem: Se o ruído consiste em pontos isolados ("sal e pimenta"), enquanto a imagem ideal não contém tais manchas, pode-se identificar tais pontos como pixels que diferem grandemente em níveis de cinza de quase todos os seus vizinhos; tal pixel pode então ser substituído pela média de seus

vizinhos. Se o ruído tem um padrão periódico, deveria aparecer na transformada de Fourier da imagem como um conjunto de picos isolados, pode-se detectá-los, substituí-los pela média dos vizinhos, e executar a transformada inversa para obter uma imagem suavizada.

- Média: A média de n variáveis independentes e idênticas, distribuídas randomicamente, cada uma tendo média μ e desvio padrão σ , ainda gera média μ , mas seu desvio padrão $\frac{\sigma}{\sqrt{n}}$, protnato menor. Assim a ampliude de ruído de média zero pode ser reduzida calculando a média de instâncias independentes do ruído. Dois exemplos: se tem-se n cópias independentemente ruidosas da imagem (por exemplo, quadros de uma cena de TV estática), pode-se reduzir o ruídos calculando a média de seus pointwise. Se a imagem ideal é composta de regiões de nível de cinza constante, pode-se reduzir o ruído no interior das regiões calculando a média local de cada pixel com seus vizinhos. (Infelizmente, isto também atinge as extremidades das regiões; será discutido como evitar calcular a média sobre as bordas)
- Média seletiva: Para evitar atingir as bordas quando calculamos a média de uma imagem localmente, pode-se usar qualquer dos esquemas que seguem: 1). Execute o processo de detecção de bordas para cada pixel; se uma borda está presente, só calcule a média na direção ao longo da borda, ou só com aqueles vizinhos no mesmo lado da borda como o pixel dado. b) Examine um conjunto de vizinhos assimétricos nos vários lados de cada pixel; ache aquele que tenha a mais baixa variabilidade, e substitua o pixel pela média de sua vizinhança. c) Calcule a média de cada pixel só com os vizinhos cujos níveis de cinza são mais próximos ao do pixel, ou com aqueles vizinhos cujas taxas de mudança de nível de cinza são mais baixas que a do pixel. Em quaisquer destes esquemas, pode-se fazer escolhas mais "soft" que "hard", i.e., em lugar de selecionar uma vizinhança ou um conjunto de vizinhos, pode-se dar pesos aos vizinhos, escolhendo de tal forma que vizinhos que pertencem à mesma região do pixel determinado tenham pesos altos. Tudo estes esquemas podem ser incrementados se desejado, para aumentar o grau de suavização.
- Mediana e Filtros Min/Max : Outro método de suavização local que remove ruído mas não borra as extremidades é o filtro da mediana, onde substituí-se cada pixel pela mediana do nível de cinza de sua vizinhança. Note que este processo trata curvas finas como ruído, e arredonda cantos retos.

Padrões pequenos e finos de 1's em uma imagem binária podem ser removidos por um processo de encolhimento e reexpansão de 1's, i.e., (repetidamente) substituindo 1's por 0's se eles têm 0's como vizinhos, e então invertendo o processo. Analogamente, padrões pequenos ou finos de 0's podem ser removidos se expandindo e reencolhendo os 1's. Pode-se generalizar esta idéia para imagens não binárias usando operações locais de min e max : Pequenos ou finos padrões alto-estimados em um fundo baixo-estimado podem ser removidos executando min seguidos por max local, e vice-versa.

3.3.3 Agudização

O efeito de borrar debilita mais as altas frequências que as baixas; assim a aproveitamento básico do "desborrar" ocorre nas filtragens passa-alta. Porém, isto não pode ser feito indiscriminadamente, pois o ruído é normalmente mais forte que o sinal de imagem em frequências altas. Se puder modelar o processo de borrar e o ruído,

poder-se-á projetar vários tipos de filtros ótimos que darão boas estimativas da imagem ideal; isto é restauração de imagem, e não será discutido mais adiante.

Até mesmo na ausência de tais modelos, pode-se freqüentemente usar o "des-borrar" aplicando operadores simples para a imagem. O exemplo clássico envolve o operador Laplaciano $\nabla^2(f) \equiv (\delta^2 f / (\delta^2 x) + (\delta^2 f / (\delta^2 y))$; a versão digital deste operador usa diferenças em vez de derivadas e é equivalente a $f - \bar{f}$ onde $\bar{f}$ é obtido calculando a média local de f . Pode ser demonstrado que se f foi borrado calculando-se a média local, somando-se um múltiplo de $f - \bar{f}$ a f inverte-se o efeito deste borrão, para uma primeira aproximação. De fato, as baixas freqüências de espaço de f são mantidas relativamente intactas em $\bar{f}$, enquanto as altas são debilitadas ou enfraquecidas; assim em $f - \bar{f}$, as baixas são anuladas e as altas são preservadas, de forma que em $f + (f - \bar{f})$, as baixas permanecerão inalteradas e as altas serão aumentadas.

3.4 Segmentação

A segmentação de uma imagem em partes pode ser considerado como um processo de agrupamento e classificação de pixels. A cada pixel, associa-se um ou mais valores de propriedades - por exemplo - seu nível cinza ou cor; seu intervalo, sua velocidade, etc., se estas informações estão disponíveis; ou as propriedades de sua vizinhança estão mensuradas, por exemplo, a média local dos níveis de cinza, a magnitude do gradiente, o grau de correlação para um determinado modelo, etc. Isto mapeia o pixel em um ponto do espaço de propriedade, definido por seu vetor de valores de propriedade. Conjuntos neste espaço são presumidamente distintos, subpopulações homogêneas de pixels; dividindo o espaço para separar estes agrupamentos, pode-se então classificar os pixels. Isto define uma divisão da imagem em subconjuntos, identificando os conjuntos de pixels pertencentes às diferentes classes.

Métodos de detecção e classificação de conjuntos são parte do assunto geral de reconhecimento de padrões, e não serão discutidos aqui. Nesta seção serão discutidos os tipos de propriedades que comumente se usam para classificar pixels. Também considerar-se-ão algumas possíveis estratégias de classificação, por exemplo, paralela versus seqüencial, e tratar-se-á brevemente um ponto de vista alternativo em segmentação na qual é considerado como um processo de divisão da imagem em máximas regiões homogêneas. Finalmente, serão discutidos métodos de extrair informação tridimensionais sobre uma cena de uma ou mais imagens.

3.4.1 Subpopulações homogêneas

Se uma imagem é composta de alguns tipos de regiões homogêneas, a sua população de pixels consistirá em algumas subpopulações homogêneas, e deve ser possível descobrir estas subpopulações como agrupamentos em um espaço de propriedades satisfatório.

A classe mais simples de exemplos desta aproximação envolve imagens compostas de regiões que ocupam intervalos distintos de níveis de cinzas, por exemplo, objetos escuros sobre um fundo claro. Em tais casos, pode-se obter uma boa segmentação usando uma única propriedade de nível de cinza. Os objetos e fundo devem dar origem a conjuntos em um espaço de níveis de cinza, i.e., para picos no histograma de níveis de cinza, e pode-se classificar os pixels como objeto ou fundo dividindo o histograma, i.e., escolhendo um nível de cinza, para separar os picos. Observações semelhantes aplicam-se quando há vários tipos de regiões, a não ser

que os picos se tornem difíceis de separar.

Podem ser usados vários métodos padrões para selecionar um limiar ótimo. Por exemplo, baseado em ajustar densidades prováveis de formas conhecidas (Gaussiana, por exemplo) para os picos do histograma, ou em minimizar as variâncias dentro da classe enquanto maximizam a variância entre as classes. Os métodos baseados em modelos para a geração de classes de imagens são de maior interesse do ponto de vista do processamento de imagem. Suponha que os objetos e fundo consistem em regiões relativamente grandes separadas por extremidades suaves e íngremes. Então é razoável escolher um limiar (entre os cumes) que minimize o número de adjacências entre os pontos de baixo e alto limiar desde que tais adjacências só devem acontecer nas bordas que são relativamente infreqüentes, e não em objetos ou interiores de fundo. Também pode-se usar a magnitude de gradiente de níveis de cinza como uma propriedade auxiliar para facilitar a seleção de um bom limiar. De fato, é provável que os pixels nos quais esta magnitude está alta façam parte das bordas dos objetos/fundo, de forma que a média de seus níveis de cinza poderia ser um bom limiar. Reciprocamente, é provável que pixels nos quais a magnitude está baixa façam parte do interior de objetos/fundo, de forma que o histograma destes pixels deveriam ter picos que são mais fáceis de distinguir. Outras propriedades locais (por exemplo, média local de níveis cinza, magnitude de Laplaciano, etc.) podem ser usadas de modo análogo.

Se os conjuntos do objeto e do fundo se sobrepõem, haverá erros não importa onde escolha-se o limiar. Uma possibilidade é usar dois limiares; pixels tendo níveis de cinza abaixo do mais baixo ou acima do superior são definitivamente classificados como objeto ou fundo, enquanto entre os dois limiares usamos informações de vizinhança, i.e., pixels duvidosos são classificados como a maioria dos seus vizinhos já classificados. Outra alternativa é suavizar a imagem antes de limiarizá-la, isto reduz a variabilidade dos níveis de cinza nos objetos e no fundo, e assim torna mais fácil separar os cumes de histograma. (Note que também suaviza irregularidades nas bordas do objeto/fundo.) Considerações semelhantes se aplicam a classificação de pixel baseadas em outras propriedades que as de nível cinza.

Um único limiar pode não ser eficiente se aplicado a imagem inteira, devido a iluminação desigual ou sensibilidade de sensor. Em tais casos pode-se dividir a imagem em blocos; escolher limiares locais para esses blocos que têm dois picos de histograma; e interpolar estes limiares para definir um limiar variável para a imagem inteira. A mesma idéia pode ser usada para extrair pequenos objetos ou características locais que não dariam origem a um pico na imagem inteira, mas que o fazem nos histogramas de blocos. (em geral, sempre que descobre-se uma característica local, tenta-se extraí-la aplicando um limiar em sua vizinhança) Reciprocamente, pode-se procurar blocos que são homogêneos e usá-los para definir centros de agrupamento em um espaço de níveis de cinza; e de forma similar para classificação baseada em outras propriedades.

Classificação de pixels baseado em cor (ou mais geralmente, em valores de brilho em um determinado conjunto de faixas espectrais) é um método padrão de segmentar imagens de sensores remotos. Para imagens preto-e-brancas, pode-se classificar os pixels baseado nos valores das propriedades locais mensuradas em suas vizinhanças; tais valores são freqüentemente altamente variáveis, mas pode-se reduzir essa variação através de processos de suavização, como mencionado anteriormente. Em lugar de lidar com todas as propriedades ao mesmo tempo, pode-se tentar usá-las uma de cada vez para segmentar a imagem recursivamente, i.e., achar um pico de

histograma bom, extrair a subpopulação correspondente, e repetir o processo para os pixels restantes.

3.4.2 Bordas

Se imagem contém muitos tipos diferentes de regiões homogêneas, pode não ser possível segmentar a imagem em subpopulações homogêneas, mas ainda assim pode-se detectar as bordas entre as regiões. (Note, porém, que isto não separa as regiões automaticamente, visto que as bordas detectadas podem ter falhas devido a borões ou ruídos.). Para detectar bordas, mede-se a taxa de mudança dos níveis de cinza (ou cores, etc.) a cada pixel, e se esta taxa é suficientemente alta, decide-se que uma borda está presente, i.e., classifica-se o pixel como estando em uma borda. Esta seção discute vários métodos de medir a taxa de mudança de nível de cinza; considerações semelhantes se aplicam para detecção de bordas baseadas em outras propriedades de pixels diferentes da de nível de cinza.

Em duas dimensões, a taxa de mudança de uma função f depende de direção; o gradiente de f é um vetor cujo magnitude, $\sqrt{(\delta f/\delta x)^2 + (\delta f/\delta y)^2}$, de um determinado ponto é a taxa de máxima mudança de f naquele ponto, e cuja direção, $\tan^{-1}(\frac{\delta f/\delta x}{\delta f/\delta y})$, é a direção deste máximo. Para imagens digitais, pode-se definir um "gradiente digital" usando diferenças em lugar de derivadas, em um par de direções perpendiculares; dependendo em como estas diferenças são definidas, o máximo de seus valores absolutos, em lugar da raiz quadrada da soma dos quadrados, pode ser menos sensível à orientação de bordas.

Pode ser vantajoso usar diferenças de médias (ponderadas), em lugar de simples níveis de cinzas; isto permite suavizar na direção ao longo da borda enquanto diferenciamos transversalmente a ela. Pode-se usar isto em pontos espaçados de forma irregular para detectar bordas, e então usar melhores diferenças para localizar as bordas com maior precisão. Em geral, pode-se usar outras medidas de diferença entre populações de níveis de cinza em lugar de simplesmente as diferenças entre suas médias. Mais que fazer diferenças em duas direções perpendiculares e combina-las para calcular o gradiente, pode-se tomar as diferenças em um conjunto maior de direções e simplesmente escolher o valor máximo. Outra alternativa para o gradiente é usar o Laplaciano digital que também responde às bordas; porém, este é até mais sensível às linhas e aos pontos ruidosos. Uma objeção semelhante se aplica ao uso de filtros de frequência passa-alta para detecção de bordas. O Laplaciano também não responde a rampas; pode-se definir outros operadores com esta propriedade que são menos sensíveis ao ruído, por exemplo, a diferença entre a média e a mediana local.

Outras técnicas para detecção de bordas estão baseadas em ajustar uma superfície (por exemplo, planar ou quadrática) à vizinhança de cada pixel, e achar o gradiente desta superfície. Isto mostra que pode ser expresso diretamente em termos dos níveis de cinza do pixel e de seus vizinhos. Alternativamente, pode-se ajustar uma função degrau para a vizinhança, e medir o "edgeness" pela altura do degrau. Para vizinhanças pequenas, estas técnicas tem o mesmo rendimento que alguns dos operadores padrão de diferenças (de média).

Freqüentemente é difícil decidir se um determinado valor de gradiente se deve ao ruído ou a uma borda, visto que a imagem pode conter bordas de alturas diferentes. Para tomar-se esta decisão, precisa-se de um modelo para as populações de valores de gradiente de níveis de cinza dentro de regiões e através de bordas.

A resposta para um detetor de borda deveria tender a cair em curvas suavizadas; esta detecção de curva pode ser usada para limpar (e se desejar, ligar) estas respostas. Outro modo de limpeza e de ligação de bordas é aplicar uma limiarização (local) para a imagem; pode-se descartar respostas que não caem nas bordas (acima e abaixo das regiões da limiarização), e vincular as respostas ao longo destas bordas. Reciprocamente, pode-se usar respostas de bordas para ajudar a selecionar limiarizações (locais); dado um conjunto de limiarizações, escolhe-se aquela para a qual as respostas ao longo das bordas são mais altas.

3.4.3 Características locais

Nesta seção, nós discutiremos brevemente como descobrir características como manchas, linhas, ou curvas em uma imagem, i.e., como classificar pixels como pertencendo a tal característica.

A técnica clássica para detectar características locais usa filtros lineares de comparação, i.e., convolve a imagem com o modelo da característica desejada (ou, mais comumente, de sua derivada segunda). Infelizmente, esta técnica não é nenhuma característica específica; por exemplo, um modelo de linha responde mais fortemente para uma borda de alto contraste ou para um ponto ruidoso do que para uma linha de baixo contraste. Desta forma é preferível, usar filtros não lineares; por exemplo, para descobrir uma linha vertical fina e brilhante em um determinado pixel, poderia exigir que o pixel fosse mais brilhante que seus vizinhos horizontais (isto exclui as bordas), e o mesmo se aplica aos seus vizinhos verticais (isto exclui pontos ruidosos). Um conjunto de tais operadores é necessário para detectar linhas em todas as orientações; a pessoa pode descobrir então também curvas arbitrárias suaves, desde que tal uma curva se pareça como uma linha reta naquele local (exceto onde se inclina por mais que 45° graus em uma vizinhança de 3×3). Curvas espessas podem ser detectadas da mesma maneira, usando-se os operadores baseados em blocos de pixels (de tamanho igual às grossuras da curva) em lugar de simples pixels. Operadores não lineares análogos podem ser definidos para descobrir pontos isolados, picos, ou manchas, requerendo que um bloco seja mais brilhante que seus vizinhos laterais.

Características globais, tais como bordas ou curva de uma formas dadas, podem ser detectadas pelo mapeamento em características locais em um espaço da Transformada de Hough. Por exemplo, para descobrir bordas colineares ou respostas de curva, pode-se proceder como segue: dado uma resposta em (x_i, y_i) com orientação θ_i , seja l_i uma linha de inclinação θ_i através de (x_i, y_i) , e seja p_i a menor distância menor de l_i até a origem; então nós mapeamos a resposta no ponto (p_i, θ_i) no espaço (p, θ) . Evidentemente, um conjunto de respostas colineares será mapeado no mesmo espaço (p, θ) , dando origem a um pico no espaço (p, θ) . Em geral, se tem-se duas famílias de parâmetros de curvas, uma resposta em (x_i, y_i) com inclinação q_i determina um par de valores de parâmetro; se há muitas respostas que recaem na mesma curva, ter-se-á um pico neste par de valores. Além disso, uma resposta de inclinação desconhecida em (x_i, y_i) determina uma subfamília de curvas de um parâmetro; se há muitas respostas que recaem na mesma curva, estas subfamílias deverão ter interseção no par de valores de parâmetro da curva.

3.4.4 Segmentação Sequencial e Iterativa

As técnicas de segmentação descritas até agora tem sido "paralelas", no sentido que são tomadas decisões para cada pixel independentemente. Serão discutidas agora esquemas de segmentação sequenciais nos quais cada decisão pode depender

das decisões anteriores. Tais esquemas são potencialmente mais baratos visto que só precisa executar processos computacionais baratos em cada pixel, e se reserva executar os processos mais caros para lugares onde a segmentação já tenha sido iniciada. Eles também são potencialmente mais aplicáveis à natureza dos objetos que estão sendo extraídos, visto que podem se modificar dependendo do que foi previamente encontrado. Os métodos "paralelos", por outro lado, grandemente podem ser acelerados se um hardware paralelo é disponibilizado, e os seus resultados não dependem da ordem na qual são examinados os pixels.

O exemplo clássico de segmentação seqüencial é a perseguição de borda ou curva. Aplica-se um critério simples e conservador de detecção (por exemplo, limiar da magnitude de gradiente) a cada pixel para descobrir pixels de uma possível borda (ou curva); tenta-se estender então ou seguir estes fragmentos de borda examinando dos vizinhos nas direções apropriadas. O critério para aceitar novos pixels de borda pode depender do nível de cinza, do contraste, da curvatura, etc., dependendo de quanto é conhecido sobre as bordas desejadas. Para evitar problemas, devem ser permitido backups, pelo menos de forma limitada. Programação matemática ou algoritmos heurísticos de procura podem ser usados para extrair "ótimas" bordas da imagem.

Geralmente, pode-se extrair seqüencialmente regiões bi-dimensionais de uma imagem, começando com (pequenos) pedaços homogêneos e somando a eles, ou combinando a eles, enquanto os resultados permaneçam homogêneos. O critério de aceitação pode depender do nível de cinza, da textura, do tamanho, da forma, etc., dependendo do que é conhecido sobre as regiões desejadas.

A meta da segmentação geralmente é dividir a imagem dada em um conjunto de máximas regiões homogêneas onde "homogêneo" poderia significar, por exemplo, ter uma variação de nível de cinza abaixo de um determinado limiar. O crescimento de regiões ou técnicas de fusão podem ser usadas para este propósito, começando com uma partição trivial (por exemplo, em componentes conectados de nível de cinza constante). A fusão pode ser feita de um modo paralelo computando-se as taxas de fusão para todos os pares de regiões, e fundindo-se todos os pares que têm um mérito mutuamente mais alto. Pode-se usar a divisão em vez da fusão, onde uma imagem é recursivamente subdividida em blocos (por exemplo, em quadrantes, subquadrantes,...) e subdivide-se um bloco se este não é homogêneo, ou uma combinação de fusão e divisão pode ser usada. Aqui novamente, o processo deveria ser guiado por um modelo para os tipos de partições que são esperadas.

Recentemente tem havido interesse em técnicas iterativas para segmentação (ou para classificação de objetos inter-relacionados em um determinado conjunto de classes) nos quais procede-se como se segue: inicialmente, calcula-se a probabilidade de cada objeto (por exemplo, um pixel) pertencer a cada classe, e então iterativamente ajusta-se estas probabilidades de classificação para cada objeto, e dependendo das probabilidades dos objetos vizinhos e as compatibilidades entre os membros da classe para pares vizinhos de objetos. Por exemplo, para detectar bordas suavizadas, poderia inicialmente calcular uma probabilidade e orientação de borda para cada pixel, baseado em respostas de operadores locais; deixa-se estas probabilidades então interagir, com bordas reforçando bordas que dão continuidade de forma suave. Esta técnica permite maior uso de contexto que técnicas paralelas one-shot, mas ainda é potencialmente rápida (cada iteração é paralela), como também é independente da ordem.

Capítulo 4

SEGMENTAÇÃO POR BINARIZAÇÃO SEGUNDO O’GORMAN

4.1 Introdução

A captura de dados de documentos através da digitalização ótica ou vídeo digital gera um arquivo de elementos de imagem ou pixels, que é a essência da entrada para a análise de documentos. Estes pixels são exemplos de valores de intensidade tomados em uma grade sobre a página do documento, onde estes valores podem ser desligados (0) ou ligados (1) para imagens binárias, de 0 a 255 para imagens com níveis de cinza e três canais de 0 a 255 valores de cores para imagens coloridas. O primeiro passo na análise do documento é executar um processamento desta imagem para prepará-la para posterior análise. Tal processamento inclui a limiarização para reduzir uma imagem em nível de cinza ou colorida para uma imagem binária; a redução do ruído para reduzir dados estranhos e o afinamento e detecção da região para capacitar a subsequente detecção de características pertinentes e objetos de interesse. O processamento a nível de pixel (também chamado de pré-processamento ou processamento baixo nível em outra literatura) é o tema deste capítulo.

4.2 Limiarização

Neste tratamento do processamento do documento, lida-se com imagens contendo texto e gráficos de informações binárias. Isto é, estas imagens contêm um simples nível de primeiro plano que são o texto e os gráficos de interesse e um simples nível em segundo plano com o qual o primeiro plano contrasta. Chamar-se-á o primeiro plano de objetos, regiões de interesse ou componentes. (É claro que os documentos também podem conter informações em níveis de cinza verdadeiros ou coloridos, tais como fotografias; entretanto, além do reconhecimento da presença de uma figura em nível de cinza em um documento, deixamos a análise destas para campos mais genéricos da análise da imagem e visão de máquina). Apesar da informação ser binária, o dado (na forma de pixels com valores de intensidade) provavelmente não tem somente dois níveis, mas sim uma faixa de intensidades, como resultado de uma impressão ou refletância não uniforme da página ou como resultado de transições de intensidade nas bordas das regiões localizadas entre as regiões de fundo e de frente (primeiro e segundo plano). O objetivo na binarização é marcar os pixels que pertencem às verdadeiras regiões de primeiro plano com uma simples intensidade

(ON) e as regiões de fundo (segundo plano) com uma intensidade diferente (OFF). A figura 4.1 ilustra o resultado da binarização da imagem de um documento com diferentes valores de limiar. Os valores ON são mostrados em preto e os valores OFF são mostrados em branco.

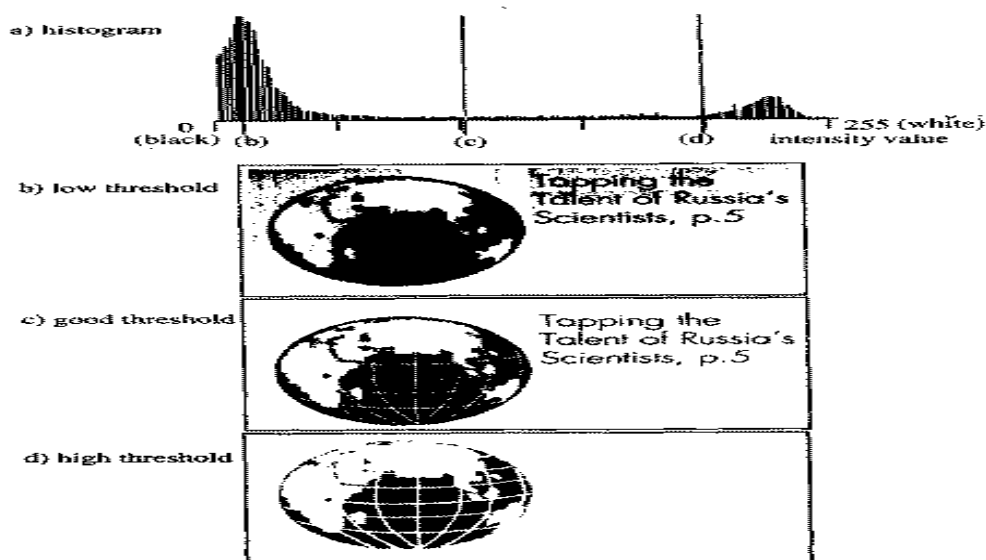


figura 4.1: Binarização da imagem de um documento com diferentes valores de limiar.

Para documentos com um bom contraste de componentes contra um fundo uniforme, os scanners binários estão disponíveis. Eles combinam a digitalização com a limiarização para gerar uma imagem binária. Entretanto, para vários documentos que possuem uma vasta faixa de fundo e intensidade de objetos, este nível de limiar fixo freqüentemente não gera imagens com uma clara separação entre os componentes de frente e os de fundo. Por exemplo, quando um documento é impresso em papel colorido diferentemente, ou quando os componentes de primeiro plano ficam desbotados por causa da fotocópia, ou quando vários scanners possuem diferentes níveis de luz, o melhor valor de limiar também será diferente. Para estes casos, existem duas alternativas. Uma é determinar empiricamente a melhor configuração da binarização no scanner (a maioria dos scanners binários possuem este ajuste), e fazer isto cada vez que a imagem for fraca "binariamente". A outra alternativa é começar com imagens em níveis de cinza (possuindo uma variedade de intensidades, geralmente de 0 a 255) a partir do estágio da digitalização e então usar métodos para a obtenção automática do limiar que melhor execute a binarização. Enquanto a última alternativa requer mais dados de entrada e mais processamento, sua vantagem é que um bom nível de limiar pode ser encontrado automaticamente, assegurando boas e consistentes imagens e evitando a necessidade de consumo de tempo através do ajuste manual e digitalização repetida. A discussão seguinte presume uma digitalização inicial para imagens em níveis de cinza.

Se os valores dos pixels dos componentes e aqueles do fundo são razoavelmente consistentes em seus respectivos valores sobre a imagem inteira, então um simples valor de limiar pode ser encontrado para esta imagem. Esta utilização de um simples limiar para todos os pixels da imagem é chamada de limiarização global. Métodos de processamento são descritos abaixo que automaticamente determinam o melhor valor para o limiar global para imagens diferentes. Contudo, para vários documentos

um simples valor de limiar global não pode ser usado igualmente para uma simples imagem por causa das deformidades dentro das regiões de primeiro e segundo planos. Por exemplo, para um documento contendo áreas de fundo brancas bem como áreas de destaque de um fundo colorido diferente, os melhores limiares mudarão por área. Para este tipo de imagem, valores de limiar diferentes são requeridos para áreas de locais diferentes. Este tipo de uso é chamado de limiarização adaptativa e está descrito abaixo.

4.2.1 Limiarização global

O modo mais direto para selecionar automaticamente um valor global de limiar é através do uso de um histograma das intensidades dos pixels na imagem. O histograma de intensidade traça o número de pixels com valores de cada nível de intensidade. Para uma imagem com intensidades de primeiro e segundo planos bem diferenciados, o histograma terá dois picos distintos. O vale entre estes picos pode ser encontrado como o mínimo entre dois máximos, e o valor da intensidade aí é escolhido como o limiar que melhor separa os dois picos. Já que os métodos de detecção de vales e picos estão propensos a localizar o local máximo e o mínimo diferentes dos desejados, um método [Rosenfeld e De La Torre [RT83]] ajusta um envelope convexo ao histograma bimodal e então detecta o vale como a máxima distância entre a cobertura e a altura da caixa.

Existe um número de desvantagens para a seleção do limiar global baseada na forma da distribuição da intensidade. A primeira é que as imagens nem sempre contém intensidades de primeiro e segundo planos bem diferenciados como um resultado de um pobre contraste e ruído. A segunda é que, especialmente para uma imagem de componentes de primeiro plano esparsos, tais como a maioria das imagens gráficas, o pico representando estas intensidades será muito menor do que o pico das intensidades de segundo plano (fundo). Esta diferença freqüentemente torna difícil encontrar o vale entre estes dois picos. Além disso, métodos de detecção de vales e picos confiáveis são problemas separados entre si.

Uma maneira de melhorar este enfoque é compilar um histograma das intensidades dos pixels que são ponderados pelo inverso dos seus valores de comprimento das bordas [Mason [Mas75]]. As intensidades de regiões de pixels com valores de bordas baixos serão mais ponderados do que as intensidades das bordas e os pixels de ruído com valores das bordas mais altos, cortando, deste modo os picos do histograma causados por estas regiões e facilitando a detecção do limiar entre eles.

Uma técnica análoga é ponderar altamente estas intensidades de pixels com valores altos de bordas e então escolher o limiar no pico do histograma, correspondendo a transição entre as regiões [Katz [Kat65]] [Watanabe [Wat74]] [Weszka, Nagel e Rosenfeld [WNR74]]. Este método requer a detecção do pico de um simples máximo, o qual é freqüentemente mais fácil do que a detecção de vales entre dois picos. Este enfoque também reduz o problema de uma grande discrepância em termos de tamanho entre os picos das regiões de primeiro e segundo planos, porque os pixels de borda estão acumulados no histograma ao invés dos pixels de região; a diferença entre uma área de tamanho grande e pequena é uma quantidade linear de bordas versus uma quantidade muito maior para as regiões.

Um terceiro método utiliza a ponderação Laplaciana. O Laplaciano é um operador de segunda derivada, o qual ponderará altamente as transições das regiões em bordas (a primeira derivada pondera altamente as bordas). Este método irá ponderar altamente os pixels de borda de ambas as regiões de primeiro plano e seus

fundos (segundo plano) circunvizinhos; assim, o histograma terá dois picos de áreas similares. Embora estas técnicas de histograma ofereçam a vantagem de que as detecções pico e vale são intuitivas, a detecção de picos é suscetível ao erro devido ao ruído e a regiões pobremente separadas. Além disso, quando a região de frente (primeiro plano) e a de fundo (segundo plano) consistem em várias regiões estreitas - tais como - textos e bordas, a medida do Laplaciano pode ser pobre por causa de transições muito abruptas (bordas estreitas) entre o primeiro e o segundo plano.

Um número de técnicas determinam as classes através do reconhecimento de padrões formal que otimiza algumas medidas de separação. Um enfoque é o limiar do erro mínimo [Kittler e Illingworth [SSW88]] [Ye e Danielson [HD88]] figura 4.2).

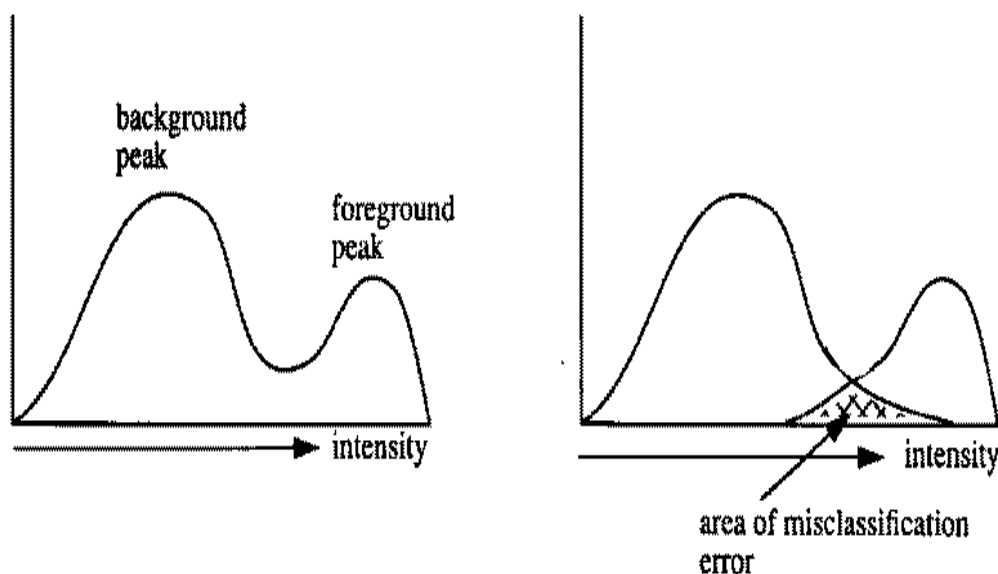


figura 4.2: Binarização a partir de funções de densidade de probabilidade Gaussianas.

Aqui, as distribuições de intensidade do primeiro e do segundo plano são modeladas como funções de densidade de probabilidade normal (Gaussiana ou forma de sino). Para cada valor de intensidade (de 0 a 255, ou uma faixa muito pequena se o limiar é conhecido e é limitada por ele), as médias e as variâncias são calculadas para as classes do primeiro e do segundo plano, e o limiar é escolhido tal que o erro da não-classificação entre as duas classes seja minimizado. O último método é classificado como uma técnica paramétrica pela suposição que a distribuição da escala de nível de cinza pode ser modelada como uma função densidade de probabilidade. Este é um método popular para várias aplicações em visão computacional, mas alguns experimentos indicam que alguns documentos não seguem bem este modelo. Deste modo, os resultados com este método são mais pobres do que com abordagens não paramétricas [Abutaleb [Abu89]].

Uma abordagem não paramétrica é o método de [Otsu [SSW88]] [Reddi, Rudin e Keshavan [RRK84]]. Primeiro os cálculos são feitos da relação entre as classes de variância (interna e externa) para cada valor de limiar em potencial. As classes aqui são os pixels de primeiro e segundo planos, e o objetivo é encontrar o limiar que maximiza a variância das intensidades entre duas classes e as minimiza dentro de cada classe. Esta relação (razão) é calculada para todos os níveis de limiar em

potencial. Nível este no qual a razão é máxima se é o limiar escolhido.

Um outro enfoque similar ao de Otsu emprega uma medida teórica de informação, chamada entropia, a qual é a medida da informação de uma imagem expressada como a média do número de bits requeridos para representar a informação [Abutaleb [Abu89]] [Johannsen e Bille [JB82]] [Sahoo e Wong [SSW88]] [Pun [Pun81]]. Aqui a entropia para duas classes é calculada para cada limiar em potencial e o limiar cuja soma de duas entropias é o maior, é escolhido como o melhor. Outra abordagem sobre limiarização é a preservação de momento [Tsai [Tsa85]]. Este é menos popular do que os métodos acima; entretanto, achamos que tem sido o mais efetivo na binarização de imagens de documentos contendo textos. Para este método, um limiar escolhido que melhor preserva a estatística de momento na imagem binária resultante é comparada com a imagem inicial em nível de cinza. Estes momentos são calculados do histograma de intensidade - os primeiros quatro momentos são requeridos para binarização.

4.2.2 Limiarização local adaptativa

Uma forma comum de executar a limiarização adaptativa é analisando as intensidades de níveis de cinza dentro de uma janela local sobre a imagem para determinar limiares locais. [Bernsen [Ber86]] [Casey e Wong [CW94]] [Kamel e Zhao [KZ93]] [Wong [Won78]]. [White e Rohrer [WR83]] descreve um algoritmo para a tal limiarização separando caracteres do fundo (segundo plano). O limiar é alterado continuamente através da imagem estimando o nível de fundo como uma média móvel bidimensional dos valores locais dos pixels tomados para todos estes pixels nessa imagem.

3).

[Mitchell e Gillies [MG89]] descrevem um método similar de limiarização onde uma normalização de nível branco no fundo é feita pela primeira estimativa de nível de branco e subtraindo este nível da imagem original. Então, a segmentação de caracteres é realizada aplicando uma faixa de limiares e selecionando a imagem resultante com um conteúdo mínimo de ruído. Este conteúdo é medido como a soma das áreas ocupadas pelos componentes mais pequenos e finos do que empiricamente através de parâmetros determinados. Olhando para os resultados da binarização da fig. 1 anteriormente mostrados, pode-se observar que a escolha do melhor limiar produz um ruído mínimo visível. O problema principal com qualquer técnica de binarização adaptativa é a escolha do tamanho da janela. O tamanho escolhido deve ser grande o suficiente para garantir que o número de pixels de fundo (segundo plano) incluídos sejam também grandes o suficiente para obter uma boa estimativa do valor da média, mas não tão grande como a média sobre as intensidades de segundo plano (fundo) não uniformes. Porém, as características na imagem frequentemente variam em tamanho tal que existem problemas com o tamanho da janela fixo. Para remediar este problema, a informação dependente de domínio pode ser usada para verificar que os resultados da binarização geram as características esperadas (por exemplo, uma bolha grande de uma região com valor ON não é esperada na página dos símbolos pequenos). Se o resultado é inesperado, então o tamanho da janela pode ser modificado e a binarização é aplicada outra vez.

4.2.3 A escolha de um método de limiarização

Se os métodos de limiarização global e adaptativo são usados para binarização, não se pode nunca esperar resultados perfeitos. Dependendo da qualidade do original, pode haver buracos nas linhas, bordas rompidas na região limítrofes e regiões

de pixels estranhas com valores ON e OFF. A suposição de que os resultados do processamento não serão perfeitos é geralmente verdadeiro com outros métodos de processamento de documento e certamente com o processamento da imagem em geral.

O procedimento recomendado é processar tanto quanto possível cada passo, mas adiar decisões que não necessitam ser executadas até passos recentes, para evitar erros irreparáveis. Nessas últimas etapas, existe mais informação do resultado do processamento do que aquele ponto e esta informação provê o maior contexto e descrições de alto nível para auxiliar na tomada de decisões corretas e por conseguinte, o reconhecimento. O adiamento, quando possível, é um princípio apropriado para todos os passos da análise de documentos (exceto, é claro, o último).

Um número de diferentes métodos de limiarização foram apresentados nesta seção. Nenhum método é melhor para todos os tipos de imagens e aplicações. Para problemas mais simples, onde as características da imagem não variam muito dentro dela ou entre diferentes imagens, os métodos mais simples serão suficientes. Para problemas mais difíceis de ruído ou características da imagem variantes, métodos mais complexos geralmente serão requeridos. Os produtos comerciais variam em suas capacidades de limiarização. Os scanners atuais executam a binarização com um valor de limiar fixo. Sistemas de documentos mais sofisticados provêm técnicas manuais ou automáticas baseadas em histogramas para a limiarização global. O uso mais comum de limiarização adaptativa está em sistemas de objetivos especiais usados em bancos para representar cheques. A melhor maneira de escolher um método neste momento é primeiro estreitando as escolhas através de descrições de métodos e então apenas ir experimentando-os com os outros e examinando seus resultados.

Como não existe o melhor método de limiarização, há bastante espaço para futuras pesquisas neste campo. Um problema que requer mais trabalho é identificar métodos de limiarização ou enfoques que melhor trabalham com documentos com características particulares. Muito dos métodos descritos acima não foram formulados para documentos em particular, e seus desempenhos não são bem conhecidos. Os documentos possuem características, tais como linhas finas, que favorecerão um método sobre outro. Relacionado a questão de qual método é o melhor para usar é como melhor quantificar os resultados da limiarização. Para texto, uma forma é executar o reconhecimento ótico de caracteres (OCR) nos resultados binarizados e medir a taxa de reconhecimento para limiares diferentes.

Outro problema que requer trabalho adicional é o de multilimiarização ou limiarização multimodal. Às vezes, os documentos não possuem dois, mas três ou mais níveis de intensidade. Por exemplo, vários jornais contendo caixas destacadas dentro do texto, onde o texto é colocado contra o fundo (segundo plano) com um diferente nível de cinza ou cor. Enquanto as capacidades de multilimiarização tem sido reivindicadas para alguns dos métodos discutidos acima [Otsu [SSW88]], poucos trabalhos tem sido focalizados neste problema.

Capítulo 5

GLOSSÁRIO DE HARALICK

1. A IMAGEM

1 - IMAGEM (**IMAGE**): Uma imagem é a representação espacial de um objeto em cenas de duas ou três dimensões ou outra imagem. Pode ser real ou virtual como na área de óptica. Numa visão computacional, "imagem" usualmente significa imagem gravada como uma imagem de vídeo, imagem digital ou um quadro. Ela pode ser abstrata como uma função contínua I de duas variáveis definidas no mesmo limite e região retangular de um plano. O valor da imagem localizado em coordenadas espaciais (r, c) é indicado por $I(r, c)$. Para sensores ópticos ou fotográficos, $I(r, c)$ é tipicamente proporcional à energia radiante recebida de uma banda eletromagnética para o qual o sensor ou detetor é sensitivo numa pequena área ao redor de (r, c) . Para sensores de busca, $I(r, c)$ é uma função de linha da distância de visão de $I(r, c)$ para um objeto do mundo tridimensional. Para um sensor tátil, $I(r, c)$ é proporcional à quantidade da superfície que deforma o sensor em e ao redor de (r, c) . Quando a imagem é um mapa, $I(r, c)$ é um índice ou símbolo associado com alguma categoria como uma cor, um tipo de solo ou um tipo de pedra. Uma imagem gravada pode ser em formato fotográfico, formato de sinal de vídeo ou formato digital.

2 - IMAGEM DE VÍDEO (**VIDEO IMAGE**): Uma imagem de vídeo é uma imagem em formato de sinal eletrônico capaz de ser mostrado num tubo de raios catódicos ou monitor. O sinal de vídeo pode ser gerado a partir de uma saída como uma câmera CCD, um scanner spot, um sensor tátil, um sensor de alcance ou um buffer de frames para um conversor digital-analógico. Imagens de vídeo possuem dois formatos comuns. No formato de frame, o próprio sinal de vídeo é uma sequência de sinais, cujo i -ésimo sinal representa a i -ésima linha de imagem. O i -ésimo sinal é separado do $(i-ésimo + 1)$ por um sync horizontal ou pulso. Cada frame de vídeo é separado do próximo frame de vídeo por um pulso sync vertical. No formato entrelaçado, o sinal de vídeo é dividido em dois campos. O primeiro campo contém todas as linhas ímpares; no segundo campo contém todas as linhas pares. Como no formato frame, a i -ésima linha do campo é o i -ésimo sinal e ela é separada da próxima linha do campo por um pulso de sync horizontal. Sucessivos campos são separados por pulsos sync verticais.

3 - NÍVEL DE CINZA (**GRAY LEVEL, GRAY SHADE, GRAY TONE, INTENSITY, IMAGE INTENSITY, IMAGE DENSITY, BRIGHTNESS, IMAGE VALUE**): Um nível de cinza, régua de cinza, tom de cinza ou intensidade

de tom cinza, intensidade da imagem, densidade da imagem, ou valor da imagem é um número ou valor assumido para uma posição sobre a imagem. Para sensores ópticos ou fotográficos, a intensidade da imagem em (r, c) é proporcional a saída integrada, refletância ou transmitância de uma pequena área, usualmente chamando uma de célula de resolução ou píxel, centralizado na posição (r, c) . Este valor pode ser relatado como transmitância, refletância ou coordenada de tristimulus, ICI , YIQ ou sistema de coordenada de cores RGB , brilho, radiância, luminescência, densidade, voltagem ou corrente.

4 - RESOLUÇÃO (**RESOLUTION**): Resolução é um termo genérico que descreve muito bem um sistema, processo, componente, material ou imagem que pode reproduzir um objeto isolado construído de objetos espaçados fechados separados ou linhas. A limitação de resolução, limite de resolução ou resolução espacial é descrita em termos como dimensões pequenas de um objeto ou objeto que pode ser discriminado ou observado. Resolução pode ser uma função do contraste de um objeto ou posição espacial conhecido como elemento de régua (ponto simples, n° de pontos num grupo, continuum ou linha, etc.)

5 - CÉLULA DE RESOLUÇÃO (**RESOLUTION CELL**): Uma célula de resolução é uma pequena parte elementar aérea constituída e associada com intensidade da imagem numa imagem digital. Uma célula de resolução é referenciada por uma coordenada espacial a qual é o centro da coordenada da área. A resolução de célula ou formação espacial da resolução da célula constitui a unidade básica para um baixo nível de processamento de dados de imagens digitais. Resolução de células usualmente possuem áreas quadradas, retangulares ou hexagonais.

6 - ACUTANCE (**ACUTANCE**): A acutance é uma medida de grandeza de bordas numa fotografia ou imagem. É definida como qualquer margem pela taxa quadrada média dividida pelo total de intensidade de imagens diferentes de um lado da margem pelo outro lado da margem.

7 - CONTRASTE (**CONTRAST**): O contraste de um objeto oposto é o background e pode ser medido por: (1) a taxa de contraste a qual é a taxa entre a maior transmitância e a menor transmitância. (2) a diferença de contraste, o qual é a diferença entre a maior densidade do objeto ou do fundo; (3) a modulação de contraste, a qual é a diferença entre um objeto escuro ou intensidade da imagem de fundo das duas dividido pelo somatório da intensidade da imagem do objeto e a intensidade da imagem de fundo.

8 - PIXEL (**PIXEL, PICTURE ELEMENT, PEL**): Um píxel, elemento de imagem ou pel é um par no qual o primeiro membro é uma resolução de célula ou (linha ou coluna) posição espacial e o segundo membro é o valor de intensidade de imagem ou vetor de valor de imagem associado com a posição espacial.

9 - VOXEL (**VOXEL**): Um voxel, contração das palavras elemento de volume, é um par ordenado cujo primeiro componente é uma localização de um volume retangular paralelepípedo (linha, coluna, fatia) e cujo segundo componente é um vetor de propriedades num volume paralelepípedo retangular.

10 - EDGEL (**EDGEL**): Um edgel, atalho para borda elemento é uma tripla formada por três elementos cujo primeiro componente (fila, coluna) é a localização de um píxel, cujo segundo componente é a posição orientação da execução de uma borda através de um píxel e cujo terceiro componente é o comprimento da borda.

11 - **RASTER SCAN (RASTER SCAN ORDER)**: Raster scan order refere-se à seqüência de localização de pixels obtido pela scannerização de domínio espacial de uma imagem da esquerda para direita de cada linha de imagem com linhas obtidas de cima para baixo ordenadamente. As imagens de vídeo em formato de frame são imagens scannerizadas numa ordem raster scan.

12 - **RANGE IMAGE (RANGE IMAGE)**: Uma imagem range é uma imagem na qual cada valor de pixel é uma função da distância entre o pixel e a faixa da superfície do objeto no pixel. Dependendo do sensor e do pre-processamento usado para criar uma imagem intervalo, a distância pode ser a distância entre a imagem plana e o alcance da faixa da superfície, a linha da distância de visão entre o pixel e a correspondente faixa da superfície ou ainda algumas funções dessas distâncias e da posição dos pixel.

13 - **IMAGEM DIGITAL (DIGITAL IMAGE, DIGITIZED IMAGE, DIGITAL PICTURE FUNCTION)**: Uma imagem digital, imagem digitalizada é uma imagem em formato digital e é obtida pelo particionamento da área da imagem numa matriz bidimensional finita de regiões exclusivas mutuamente e uniformemente fatiadas chamadas de células de resolução e assumidas como valor de imagem representativa para cada região espacial. Uma imagem digital pode ser de forma abstrata definida como uma função cujo domínio é um conjunto finito bidimensional de células de resolução e cujo intervalo é o conjunto finito das possíveis intensidades da imagem.

14 - **RANGEL (RANGEL)**: Rangel é a fatia de elementos de dados produzidos por um sensor range. É um par cujo primeiro membro é a posição espacial linha-coluna e cujo segundo membro é o valor range ou vetor cujo primeiro componente é o valor range e cujo segundo componente é o valor da intensidade da imagem.

15 - **MAPA DE PROFUNDIDADE (DEPTH MAP, RANGE MAP)**: Um mapa de profundidade é uma imagem digital range cujo valor range em cada posição do pixel é a distância entre o plano da imagem e a fatia da superfície range correspondente ao pixel.

16 - **MAPA DE ORIENTAÇÃO (ORIENTATION MAP)**: Um mapa de orientação é uma imagem digital na qual cada pixel contem o vetor de orientação $3D$ para superfícies $3D$ correspondente à posição do pixel.

17 - **CONJUNTO MULTI IMAGENS (MULTI IMAGE SET)**: Um conjunto multi imagens ou imagens multi banda é um conjunto de imagens registradas onde cada está relacionada com o mesmo assunto, mas obtido diferentes vezes, em diferentes posições, com diferentes focos, com diferentes sensores, com diferentes frequências eletromagnéticas, com diferentes polarizações, ou partir de diferentes sessões. Embora exista alto grau de informação redundante, entre imagens e multi imagens, cada imagem usualmente possui alguma informação não disponível em qualquer uma das combinações de outras imagens. Se o conjunto de multi imagens possui N imagens, então cada linha da resolução é associada com N tuplas de valores de imagens.

18 - **IMAGEM MULTI SPECTRAL (MULTI SPECTRAL IMAGE)**: Uma imagem multi-spectral é uma imagem multi banda, sobre a qual cada banda é uma imagem obtida no mesmo instante, porém sensitivamente numa parte diferente do espectro eletromagnético.

19 - IMAGEM MULTI TEMPORAL (TIME VARYING IMAGE, MULTI-TEMPORAL IMAGE, DYNAMIC IMAGERY, IMAGE TIME SEQUENCE):

Uma imagem time varying, imagem multi temporal, imagem dinâmica, ou sequência de tempo de imagem, num conjunto, é um conjunto de multi imagens a qual cada imagem sucessiva num conjunto é adquirida na mesma cena num tempo sucessivo. Entre sucessivas fotos instantâneas, o objeto da cena pode mover-se.

20 - IMAGEM BINÁRIA (BINARY IMAGE): Uma imagem binária é uma imagem onde cada pixel toma o valor 0 ou o valor 1.

21 - IMAGEM EN NÍVEIS DE CINZA (GRAY SCALE IMAGE, GRAY LEVEL IMAGE): Uma imagem em tons de cinza, ou em níveis de cinza é uma imagem onde cada pixel possui valor num intervalo entre 0 e 1. Uma escala de cinza possui valores entre 0 e 63, ou 0 e 255, ou 0 e 1023, correspondente a 6, 8 ou 10 bits.

22 - IMAGEM SIMBÓLICA (SYMBOLIC IMAGE): Uma imagem simbólica é uma imagem cujo valor de cada pixel é um índice ou um símbolo.

23 - HISTOGRAMA (HISTOGRAM, IMAGE HISTOGRAM): Um histograma ou histograma de imagem é uma função h definida sobre um conjunto de imagens de valor intensitivo para inteiros não negativos. O valor $h(k)$ é dado pelo número de pixels da imagem que possuem a intensidade k . Para imagens que possuem uma larga escala de tons de cinza, a imagem freqüentemente será quantificada antes de ser "histogramada" ou durante os processos de histogramização.

2. FOTOMETRIA E ILUMINAÇÃO

24 - FLUXO LUMINOSO (LUMINOUS FLUX): O fluxo luminoso é a força radiante avaliada de acordo com a capacidade de produzir sensação visual. A intensidade luminosa numa direção dada é medida em termos o fluxo de luminosidade por esteradian. A unidade de intensidade é a candela. A luminância do corpo preto de um radiador na temperatura de solidificação do platino é 60 candelas por centímetro quadrado. A unidade do fluxo de luminosidade é o lumen. O fluxo luminoso emitido por uma fonte pontual uniforme de intensidade luminosa de 1 candela num ângulo sólido de 1 esteradian é 1 lumen.

25 - ILUMINAÇÃO (ILLUMINATION): A iluminação de um ponto sobre uma superfície é o fluxo de luminosidade incidente num elemento infinitesimal de uma superfície centralizada no ponto dividido pela área do elemento da superfície. A unidade de iluminação é o lux ou metro-candela. Um lux é igual a um lumen por metro quadrado. Outra unidade de iluminação é o pé-candela e é igual a um lumen por pé quadrado. A iluminação de um ponto da superfície a partir de uma fonte pontual de luz é proporcional à intensidade luminosa da fonte na direção do ponto da superfície e o coseno do ângulo entre esta direção e a direção normal da superfície. Isto é inversamente proporcional ao quadrado da distância entre o ponto da superfície e a fonte.

26 - ILUMINÂNCIA (ILLUMINANCE): A iluminância numa direção dada de um ponto na superfície é a intensidade luminosa naquela direção de um elemento de fonte infinitesimal contendo esse ponto, dividido pela área da projeção octogonal de um elemento no plano perpendicular à direção dada.

27 - RADIÂNCIA (RADIANCE): A radiância de um objeto é a medida de potência por unidade de área da superfície pela unidade de ângulo sólido ou refletido

pelo objeto através de uma direção específica. A radiância pode ser uma função do ângulo de visão e dos comprimento e largura da onda espectral.

28 - INTENSIDADE RADIANTE (**RADIANT INTENSITY**): A intensidade radiante de um ponto de um objeto é a medição da potência radiante por steradian ou refletida pelo objeto. A intensidade radiante pode ser uma função do ângulo de visão e dos comprimento e largura da onda espectral.

29 - IRRADIÂNCIA (**IRRADIANCE**): A irradiância é a potência por área unitária de energia radiante incidente na superfície.

30 - REFLETÂNCIA (**REFLECTANCE, REFLEXION COEFFICIENT, BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTION**): A refletância, o coeficiente de reflexão, ou a função de distribuição reflectante bidirecional de uma superfície é a potência radiante por área unitária refletida dividida pela superfície de potência radiante por área unitária incidente na superfície. A reflectância pode ser uma função do ângulo incidente da radiância, do ângulo de visão do sensor e dos comprimento e largura da onda espectral.

31 - MAPA DE REFLETÂNCIA (**REFLECTANCE IMAGE, REFLECTANCE MAP**): Uma refletância de imagem ou mapa de refletância é uma imagem digital onde o valor de cada pixel é proporcional à reflectância da fatia da superfície da imagem na posição do pixel para dadas iluminação e direção de visão.

32 - SUPERFÍCIE LAMBERCIANA (**LAMBERTIAN SURFACE**): Uma superfície Luminosa é uma superfície uniformemente difusa. Apresenta-se como uma imagem fosca e tem uma função refletância que é uma constante. A função de refletância da superfície Lamberciana não depende do ângulo de visão; portanto, uma superfície plana tem uma refletância Lamberciana tornar-se igualmente brilhante em todos os ângulos de visão. Para uma superfície Lamberciana, a intensidade luminosa por área unitária em uma dada direção varia com o coseno do ângulo entre a direção e a direção normal à superfície.

33 - BACKLIGHTING : O backlighting refere-se a um sistema de iluminação onde a fonte de luz está no lado oposto do objeto a partir da câmera. Backlighting tende a produzir imagens os quais são silhuetas pretas ou brancas.

34 - FRONTLIGHTING : O frontlighting refere-se a um sistema de iluminação onde a fonte de luz está no mesmo lado do objeto que a câmera.

35 - ILUMINAÇÃO AMBIENTE (**AMBIENT LIGHT**): A iluminação ambiente refere-se à luz que está presente no local ao redor de um sistema de visão artificial e que é gerada de fontes fora do sistema. Do ponto de vista do sistema de visão artificial, a luz ambiente é uma luz não planar que pode atingir desfavoravelmente o processamento da imagem. Usualmente, todo o cuidado é tomado para minimizar este resultado.

36 - COORDENADAS *rgb* (*r*, *g*, *b* **CHROMATICITY**): As coordenadas de cromaticidade de um pixel multi-espectral *r*, *g*, *b*, onde *R* é o brilho vermelho, *G* é o brilho verde e *B* é o brilho azul, são dadas por:

$$r = \frac{R}{R + G + B}$$

$$\begin{aligned} g &= \frac{G}{R + G + B} \\ b &= \frac{B}{R + G + B} \end{aligned} \quad (5.1)$$

37 - COORDENADAS H, S (H, S **COORDINATES**): As coordenadas de matiz e de saturação H, S de um pixel multi-espectral cujas coordenadas de cromaticidade são r, g, b , são dadas por:

$$\begin{aligned} H &= \begin{cases} \theta & \text{if } b < g \\ 360 - \theta & \text{if } b > g \end{cases} \\ e \\ S &= 1 - 3\min\{r, g, b\} \end{aligned} \quad (5.2)$$

onde

$$\theta = \frac{2r - g - b}{\sqrt{6}[(r - 1/3)^2 + (g - 1/3)^2 + (b - 1/3)^2]^{1/2}} \quad (5.3)$$

38 - COORDENADAS YIQ (YIQ **COORDINATES**): As coordenadas YIQ usadas nas transmissões de TV colorida $NTSC$ estão relacionadas com as coordenadas RGB pela seguinte transformação linear:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.144 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (5.4)$$

3. FOTOGRAMETRIA

39 - FOTOGRAMETRIA ANALÍTICA (**ANALYTIC PHOTOGRAMMETRY**):

A fotogrametria analítica refere-se às técnicas matemáticas analíticas que permitem a inferência das relações geométricas entre pontos e linhas no espaço de imagem das projeções perspectivas bidimensionais e no espaço dos objetos tridimensionais.

40 - FOTOGRAMETRIA DIGITAL (**DIGITAL PHOTOGRAMMETRY**):

A fotogrametria digital refere-se a um processamento computacional de imagens digitais de projeções perspectivas com as técnicas de fotogrametria analítica bem como outras técnicas computacionais para a interpretação automática de cenas ou de conteúdos de imagens.

41 - ORIENTAÇÃO RELATIVA (**RELATIVE ORIENTATION**): A orientação relativa em fotogrametria analítica é a posição e a orientação relativa de um frame comum de referência com respeito a outro. Quando dois frame comuns de referência estão na orientação conhecida relativa, os raios emanados do mesmo ponto do objeto localizado em cada imagem de câmera cruzarão exatamente em um ponto do espaço $3D$.

42 - ORIENTAÇÃO EXTERNA (**EXTERIOR ORIENTATION** or **OUTER ORIENTATION**): A orientação externa refere-se à posição e à orientação de um frame de referência da câmera com respeito ao frame de referência do mundo real.

43 - ORIENTAÇÃO ABSOLUTA (**ABSOLUTE ORIENTATION**): A orientação absoluta em fotogrametria analítica é a transformação de rotação e de translação pela qual um ou muitos frames de referência da câmera podem corresponder ao frame de referência do mundo real.

44 - EIXO ÓTICO (**OPTICAL AXIS** or **PRINCIPAL AXIS**): O eixo ótico ou eixo principal é a reta que passa pelos centros de curvaturas das superfícies das lentes.

45 - PONTO PRINCIPAL (**PRINCIPAL POINT**): O ponto principal é aquele ponto da imagem que está na interseção da imagem plana com o eixo ótico.

46 - PONTO DE FUGA (**CENTER OF PERSPECTIVITY**): O ponto de fuga de uma projeção perspectiva é o ponto comum onde todos os raios se encontram.

47 - DISTÂNCIA PRINCIPAL (**PRINCIPAL DISTANCE**): A distância principal é a distância entre o ponto de fuga e o plano de projeção da imagem. Para uma imagem fixa, esta é, muitas vezes, chamada de constante de câmera ou comprimento focal Gaussiano.

48 - ORIENTAÇÃO INTERNA (**INNER ORIENTATION** or **INTERNAL ORIENTATION**): A orientação interna é dada pela tríplice (u_0, v_0, f) onde (u_0, v_0) é a posição do ponto principal na medição do sistema de coordenada planar da imagem e f é a distância principal. A orientação interna pode também incluir valores dos parâmetros livres que descrevem a distorção da lente.

49 - PROJEÇÃO PERSPECTIVA (**PERSPECTIVE PROJECTION**): A projeção perspectiva é definida em termos do plano de projeção e do centro de fuga. Os sensores de imagens cujo modelo ideal é a câmera "pinhole" gera imagens de projeções perspectivas. Existem três estruturas de imagens comumente usadas para definir a relação entre um ponto $3D$ e a sua projeção perspectiva $2D$. Todas consideram a projeção plana como sendo perpendicular ao eixo z . Na primeira estrutura de imagens, o centro de fuga é a origem e o plano de projeção é a distância f da origem até eixo z positivo. Neste caso, a projeção (P_x, P_y) do ponto (x, y, z) é dada por:

$$\begin{aligned} P_x &= \frac{f_x}{z} \\ P_y &= \frac{f_y}{z} \end{aligned} \quad (5.5)$$

Na segunda estrutura de imagens, o centro de fuga é $(0, 0, -f)$ e o plano de projeção passa pela origem. Neste caso, a projeção (P_x, P_y) do ponto (x, y, z) , $z > 0$ é dado por:

$$\begin{aligned} P_x &= \frac{f_x}{f + z} \\ P_y &= \frac{f_y}{f + z} \end{aligned} \quad (5.6)$$

Na terceira estrutura de imagens, o centro de fuga é $(0, 0, f)$ eo plano de projeção passa pela origem. Neste caso, a projeção (P_x, P_y) do ponto (x, y, z) , $z < 0$ é dado por:

$$\begin{aligned} P_x &= \frac{f_x}{f - z} \\ P_y &= \frac{f_y}{f - z} \end{aligned} \quad (5.7)$$

50 - PROJEÇÃO PARALELA ou ORTOGRÁFICA (**PARALLEL PROJECTION** or **ORTHOGRAPHIC PROJECTION**): A projeção paralela ou ortográfica num plano perpendicular ao eixo z do ponto (x, y, z) produz as projeções (P_x, P_y) definidas por:

$$\begin{aligned} P_x &= s x \\ P_y &= s y \end{aligned} \quad (5.8)$$

onde s é o fator escalar da projeção.

51 - PARALAXE (**PARALLAX**): O paralaxe é a diferença em termos de posição observada a partir do ponto 3D projetado num par de imagens perspectivas 2D. A diferença na posição é causada pelo deslocamento da posição dos centros de fuga e nas orientações dos eixos óticos. A parte do paralaxe na direção do eixo x é chamada de $x - paralaxe$. A parte do paralaxe na direção do eixo y é chamada de $y - paralaxe$. Para o par de imagens stereo onde a linha juntando os centros de fuga é paralela ao eixo x , o paralaxe estará inteiramente na direção do eixo x quando duas imagens planas estão na mesma orientação. Caso tiver um plano de imagem inclinado com relação ao outro, o $y - paralaxe$ não será zero.

52 - PONTO MORTO (**VANISHING POINT**): Um ponto morto é o ponto no plano de projeção perspectiva 2D da imagem onde um sistema de linhas paralelas 3D convergem. Os pontos mortos de todos os sistemas de linhas paralelas 3D paralelos a um certo plano encontrar-se-ão numa linha no plano de projeção perspectiva 2D da imagem chamada de linha morta de um dado plano.

4. OPERADORES DE IMAGENS

53 - OPERADOR DE IMAGEM ou TRANSFORMAÇÃO DE IMAGEM ou OPERADOR DE TRANSFORMAÇÃO DE IMAGEM (**IMAGE OPERATOR** or **IMAGE TRANSFORM** or **IMAGE TRANSFORM OPERATOR**): Um Operador de Imagem, um Operador de Transformação ou Operador de Transformação de Imagem é uma função que pega como entrada uma imagem e produz como saída uma imagem. O domínio do operador de transformação é muitas vezes chamado de domínio espacial. A área do operador de transformação é muitas vezes chamado de domínio da transformação. Alguns operadores de transformação de imagem têm domínios espaciais e de transformação diretamente diferentes da geometria ou caracter; a imagem no domínio espacial pode aparecer diretamente diferente e poder ter uma interpretação diferente da imagem no domínio da transformação. Exemplos específicos destes tipos de transformações de imagens incluem transformações de Fourier, Seno, Coseno, Slant, Harr, Hadamard, Mellin, Karhunen-Loeve e Hough. Os operadores de imagens que têm domínios de transformação e espaciais da geometria ou caracter similar incluem operadores pontuais, operadores de vizinhança e filtros espaciais.

5. OPERADORES PONTUAIS

54 - **OPERADOR PONTUAL (POINT OPERATOR)**: Um operador pontual é um operador de imagem onde o valor de cada pixel da imagem de saída numa posição específica depende somente do valor do pixel corespondente na imagem de entrada na mesma posição.

55 - **BINARIZAÇÃO ou LIMARIZAÇÃO (THRESHOLDING)**: A binarização ou limiarização é um operador pontual de imagem que produz uma imagem binária a partir de uma imagem em níveis de cinza. Um valor binário 1 é produzido na imagem de saída sempre que o valor do pixel na imagem de entrada está acima do limiar mínimo especificado. Um valor binário 0 é produzido caso contrário. Alternativamente, a binarização pode produzir valor binário 1 na imagem de saída sempre sempre que o valor do pixel na imagem de entrada está abaixo do limiar máximo específico. Um valor binário 0 é produzido caso contrário.

56 - **FATIAR EM NÍVEL ou EM DENSIDADE (LEVEL SLICING or DENSITY SLICING)**: Fatiar em nível ou em Densidade é a ação de um operador pontual que emprega dois limiares e produz uma imagem binária. Um valor binário 1 é produzido na imagem de saída sempre que o valor do pixel na imagem de entrada está entre os limiares mínimo e máximo especificados. Um valor binário 0 é produzido caso contrário.

57 - **BINARIZAÇÃO ou LIMARIZAÇÃO MULTI-NÍVEIS (MULTI-LEVEL THRESHOLDING)**: A binarização ou limiarização multi níveis é um operador pontual que usa duas ou mais limiares. Os valores dos pixels que estão no intervalo entre dois limiares sucessivos recebem um índice associado a esse intervalo.

58 - **ALONGAMENTO DE CONTRASTE (CONTRAST STRETCHING)**: O alongamento de Contraste refere-se a algum operador pontual de forma monótona crescente cujo efeito é aumentar ou realçar a visibilidade de detalhes de uma imagem.

59 - **QUANTIZAÇÃO (QUANTIZING)**: A quantização é operador pontual de forma monótona crescente que associa a cada valor de intensidade de imagem numa imagem digital um novo valor dentro de um conjunto finito de valores quantizados. A imagem quantizada tem menos níveis de cinza distintos, mas pode fazer o melhor uso da área dinâmica. Deste modo, a quantização freqüentemente relace a aparência da imagem. Existem quatro métodos mais usados de quantização: quantização de intervalo igual, quantização de probabilidade igual, quantização de variância mínima e hiperbolização do histograma. Em cada método, o intervalo de valores de imagens do valor máximo para o valor mínimo é dividido em intervalos próximos e a cada valor da imagem é associado seja o valor médio da classe quantizada a qual pertence ou seja o índice da classe quantizada a qual pertence:

(a) 60 - **Quantização de intervalo igual** ou *quantização linear*: Na quantização de intervalo igual ou quantização linear, a área dos valores de imagem do máximo valor para o mínimo valor é dividido em cada intervalos próximos de igual comprimento, e cada imagem valor é nomeado na classe quantizada a qual corresponde ao intervalo interno que se encontram;

(b) 61 - **Quantização de probabilidade igual** Na quantização de probabilidade igual, o intervalo de valores de imagens é dividido em intervalos próximos tal como depois os valores de imagens recebam as suas classes quantizadas que é uma freqüência de ocorrência igual para cada valor quantizado na imagem digital

quantizada; a quantização de probabilidade igual é as vezes referenciada com um histograma de quantização;

(c) 62 - **Quantização de variância mínima**: Na quantização de variância mínima, o intervalo de valores de imagens é dividido em intervalos próximos tal como a soma ponderada das variâncias dos intervalos quantizados seja minimizada. Os pesos são geralmente escolhidos como sendo probabilidades das classes quantizadas que são computadas como as áreas proporcionais na imagem que tem valores nos intervalos de quantização;

(d) 63 - **Quantização da hiperbolização de histograma**: Na quantização da hiperbolização de histograma, o intervalo de valores de imagens é dividido em intervalos próximos e a cada valor de imagem é associada a média de sua classe quantizada. A divisão é feita no modo que a imagem quantizada tenha um brilho uniforme. A hiperbolização de histograma leva em consideração a não linearidade da combinação cérebro olho humano.

64 - **MASCARAGEM (MASKING)**: A mascaragem é um operador pontual aplicado para uma imagem de duas bandas. A primeira banda de imagem é uma imagem binária B e é chamada de banda da imagem máscara. A segunda banda de imagem I é chamada de imagem a ser mascarada. A mascaragem produz como resultado a imagem J cujos pixels tomam o valor zero onde a imagem mascarada tem o valor zero e cujos pixels tomam o valor da imagem I onde a imagem mascarada tem como valor um. Isto é:

$$J(r, c) = \begin{cases} 0 & \text{se } B(r, c) = 0 \\ I(r, c) & \text{se } B(r, c) = 1 \end{cases} \quad (5.9)$$

65 - **DETECÇÃO DE MUDANÇA (CHANGE DETECTION)**: É o processo pelo qual duas imagens registradas podem ser comparadas, pixel por pixel, e quando pixels nas imagens de entradas tem significativamente diferentes níveis de cinza, um valor binário 1 é dado para o correspondente pixel de saída, caso contrário é gerado um valor binário 0 na imagem de saída. Um operador de detecção de mudança é um operador de ponto.

6. OPERADORES ESPACIAIS

66 - **PROCESSAMENTO DE SINAIS BIDIMENSIONAIS (TWO DIMENSIONAL SIGNAL PROCESSING)**: Processando sinal bi-dimensional refere-se à área de processamento de imagem em que cada sinal de uma dimensão processa técnicas de filtragem de ruídos, restauração, compressão de dados e detecção tem sido generalizada para bi-dimensional e por meio disso sendo aplicada para dados de imagem.

67 - **OPERADOR DE VIZINHANÇA (NEIGHBORHOOD OPERATOR)**: Um operador de vizinhança é um operador de imagem, onde o valor em cada posição de pixel da imagem de saída depende somente dos valores da imagem de entrada em uma vizinhança, contendo ou envolvendo a correspondente posição do pixel de entrada.

68 - **FILTRO ESPACIAL (SPATIAL FILTER)**: é um operador de imagem, em que o domínio espacial e de transformação têm geometrias similares, onde o valor em cada pixel de saída da imagem, depende mais do que um valor de pixel na imagem de entrada. Usualmente, mas nem sempre, o valor da imagem de saída tem maior dependência nos valores de entrada da imagem em alguma vizinhança

centralizada do correspondente pixel na imagem de entrada.

69 - FILTRO LINEAR ESPACIAL (LINEAR SPATIAL FILTER): Um filtro espacial linear é um filtro espacial onde, a intensidade da imagem nas coordenadas (r, c) na imagem de saída é uma média ponderada ou uma combinação linear das intensidades da imagem localizada em uma modelo espacial particular ao redor das coordenadas (r, c) da imagem de entrada. Por exemplo, um filtro linear espacial que enfatiza altas frequências espaciais terá tendência a escurecer a imagem e reduzir o ruído sal-e-pimenta. Quando o propósito do filtro é aumentar a vizinhança tendo certas formas, o operador é chamado às vezes de "mask matching".

70 - KERNEL (KERNEL): O kernel de um filtro linear espacial é uma função definida no domínio do modelo espacial do filtro, cujo valor de cada pixel do domínio é o peso ou coeficiente da combinação linear que define o filtro linear espacial.

71 - CAIXA DE FILTRO (BOX FILTER): Uma caixa de filtro é um filtro linear espacial de suavização, em que cada pixel na imagem filtrada é igual a média ponderada dos pixels em uma janela retangular centrada em sua posição espacial na imagem de entrada.

72 - FILTRO GAUSSIANO (GAUSSIAN FILTER): O filtro Gaussiano é um filtro linear espacial de suavização, cujo o kernel é dado pelo Gaussiano bi-dimensional:

$$k(r, c) = \frac{1}{2\pi} e^{-\frac{1}{2}(\frac{r^2}{\sigma_r^2} + \frac{c^2}{\sigma_c^2})} \quad (5.10)$$

Filtrando uma imagem com um filtro Gaussiano, a imagem será suavizada.

73 - CONVOLUÇÃO (CONVOLUTION): Convoluir uma imagem I com um kernel k , tendo suporte ou domínio K produz uma imagem convoluída denotada por $I * k$, que é definida por:

$$(I * k)(r, c) = \sum_{(i,j) \in K} I(r - i, c - j) k(i, j) \quad (5.11)$$

A convolução é uma operação linear.

74 - FILTRO POR CONVOLUÇÃO (CONVOLUTION FILTER): Uma operação de filtragem bi-dimensional é chamada separável, se a convolução puder ser decomposta em duas sucessivas convoluções de uma dimensão. Uma operação de convolução na imagem linha a linha, e a segunda operação de convolução na imagem coluna a coluna.

75 - CORRELAÇÃO (CORRELATION): Correlacionando uma imagem I com um kernel k tendo como suporte ou domínio K produz uma imagem correlatada J , definida por:

$$J(r, c) = \sum_{(i,j) \in K} I(r + i, c + j) k(i, j) \quad (5.12)$$

A correlação é uma operação linear.

76 - TRANSFORMADA DE FOURIER (DISCRETE FOURIER TRANSFORM): A transformação discreta de Fourier $\hat{I}$ de uma imagem digital I representa a imagem em termos de uma combinação linear de exponenciais complexos, a transformação linear Fourier I é definida por:

$$\hat{I}(Wr, Wc) = \frac{1}{RC} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} I(r, c) e^{2j\pi(\frac{rWr}{R} + \frac{cWc}{C})} \quad (5.13)$$

1 $\hat{I}(Wr, Wc)$ é um coeficiente da exponencial complexa $e^{2j\pi(\frac{rw}{R} + \frac{cw}{C})}$ na combinação linear representando o I como pode imediatamente ser visto pela correspondente relação:

$$I(r, c) = \sum_{w_r=0}^R \sum_{w_c=0}^C \hat{I}(Wr, Wc) e^{2j\pi(\frac{rw}{R} + \frac{cw}{C})} \quad (5.14)$$

o qual é chamado a inversa da transformação discreta de Fourier. As variáveis Wr e Wc tem a interpretação sendo linha e coluna frequências espacial.

77 - FILTRO PASSA ALTO (HIGH PASS FILTER): Um filtro passa-alto é um filtro espacial linear que atenua as baixas frequências espaciais de uma imagem, e acentua as altas frequências espaciais de uma imagem. Ele é geralmente usado para aumentar pequenos detalhes, extremidades e linhas.

78 - FILTRO PASSA BAIXO (LOW PASS FILTER): Um filtro passa-baixo é um filtro espacial linear que atenua as altas frequências espaciais de uma imagem, e acentua as baixas frequências espaciais de uma imagem. Ele é geralmente usado para suprimir pequenos detalhes indesejados, eliminar ruído, realçar as características grosseiras da imagem ou suavizar a imagem.

79 - FILTRO PASSA BANDA (BAND PASS FILTER): Um filtro de passagem de banda é um filtro espacial linear que atenua aquelas frequências espaciais exteriores a banda, e acentua as frequências espaciais dentro da banda. Ele é geralmente usado para realçar detalhes da imagem, cujo as características de tamanho espacial são relatadas pelas frequências espaciais dentro da banda.

80 - IMAGEM PIRAMIDAL (PYRAMID, PYRAMIDAL IMAGE): Uma pirâmide ou imagem piramidal é uma sequência de cópias de uma imagem, em que ambas as amostras: densidade e resolução são reduzidas em passos regulares. O mais baixo nível da pirâmide é a imagem original, cada nível sucessivo é obtido do nível anterior, por uma operação de filtragem seguida de uma operação de amostragem. Em uma imagem piramidal Gaussiana, a resolução é reduzida por sucessivas convoluções da imagem do nível anterior da pirâmide com um Gaussian como kernel. Depois da passagem do filtro passa-baixa de convolução Gaussiana, a densidade da amostra é tipicamente reduzida amostrando cada outro pixel.

Na morfologia da imagem piramidal, cada nível sucessivo é obtido por uma operação de abertura e fechamento nos níveis anteriores seguido de amostragem.

Na imagem piramidal Laplaciana, cada camada sucessiva é obtida pegando o Laplaciano do nível correspondente na pirâmide de Gaussian. O kernel do filtro de convolução Laplaciano aqui é tipicamente definido como kernel obtido, pegando o Laplaciano de um Gaussian tendo uma mudança apropriada de valor para seu desvio padrão.

81 - OPERADOR DE BORDAS (EDGE OPERATOR, STEP EDGE OPERATOR): Um operador de bordas é um operador de vizinhança que determina a extensão a qual, cada vizinhança de pixel pode ser dividida por um simples arco passando através do pixel, onde pixels na vizinhança em um lado do arco tem um valor predominante, e pixels na vizinhança do outro lado do arco tem valores predominantes diferentes. Alguns operadores de pixels podem também produzir uma direção que é a direção predominante da tangente do arco, enquanto esse passa através do pixel.

Existem 4 classes de operadores de pixel: operadores gradiente, operadores Laplaciano, operadores zero-crossing e operadores morfológicos de extremidade. Os operadores gradiente dão alguma quantidade relativa a magnitude do declive da in-

tensidade do tom de cinza da imagem básica, superfície a qual os valores de pixel da imagem observada são amostras discretizadas de ruído.

O operador Laplaciano dá alguma quantidade relacionada a superfície Laplaciana de intensidade de tom de cinza da imagem básica. O operador zero-crossing determina de qualquer forma o Laplaciano digital ou a segunda direção derivada estimada que tem um zero-crossing no pixel. O operador morfológico de extremidade dá uma quantidade relacionada aos resíduos de uma região e/ou operação de dilatação.

82 - IMAGEM DE BORDAS (EDGE IMAGE): Uma imagem de bordas é uma imagem em que cada pixel é rotulado como "borda" ou "não-borda". Além do que, neste rotulamento básico, pixels em uma imagem de borda podem carregar informação adicional, tal como a direção, contraste ou força da borda.

83 - CONEXÃO DE BORDAS (EDGE LINKING): Conectar extremidades refere-se ao processo pelo qual, os pixels rotulados como vizinhos da borda podem ser agregados para constituir uma corrente ou seqüência de pixels de borda.

84 - DETECÇÃO DE BORDAS (BOUNDARY DETECTION): Detecção de bordas ou delineação de bordas refere-se a cada processo que determina uma corrente de pixels separando uma região de imagem de uma região de imagem vizinha.

85 - BORDA OCLUSA (OCCLUDING EDGE): Uma borda oclusa é uma imagem de bordas que origina-se de uma extensão ou espessura descontínua. Isto, acontece tipicamente onde um objeto de superfície projeta para um pixel em um lado da extremidade, e outro objeto de superfície que está a alguma distância além do primeiro objeto de superfície projeta para um pixel no outro lado da superfície. Degraus de extremidades em mapas de espessura são sempre extremidades oclusas.

86 - RETAS ANTI-PARALELAS : Um par de bordas retas são ditas anti-paralelas, se não houver bordas entre eles e os picos tiverem contrastes opostos.

87 - FILTRO HOMOMÓRFICO (HOMOMORPHIC FILTERING): Uma filtragem homomórfica é um processo de filtragem em que o filtro é aplicado ao logaritmo da imagem, e a imagem de saída é obtida exponenciando a imagem logarítmica filtrada.

88 - FILTRO DA MEDIANA (MEDIAN FILTER): Um filtro da Mediana é um filtro espacial não-linear de suavização de imagem baseado em vizinhança, em que o valor de uma saída de pixel é o valor médio de todas as entradas de pixels na vizinhança suporte do filtro sobre as posições dadas dos pixels. Filtros da mediana são usados para suavizar e remover ruídos das imagens.

89 - SUAVIZAÇÃO (IMAGE SMOOTHING): Suavizar uma imagem refere-se alguma filtragem espacial produzindo uma saída de imagens que espacialmente simplifica e aproxima a imagem de entrada. Suavizar uma imagem suprime pequenos detalhes da imagem e aumenta o tamanho ou a espessura das estruturas de imagem.

90 - IMAGEM DE ESPAÇO ESCALAR (SCALE SPACE IMAGE): Uma imagem espacial escalar é uma imagem em que cada valor de pixel é uma função, indicando para cada desvio padrão o valor na posição do pixel da convolução da imagem com um kernel Gaussiano tendo desvio padrão σ .

91 - **ESTRUTURA DE ESPAÇO ESCALAR (SCALE SPACE IMAGE STRUCTURE)**: Uma estrutura espacial escalar refere-se à análise de uma imagem de escala espacial, em que cada valor de pixel é uma função, especificando para cada possível desvio padrão σ , qual pixel contém um zero-crossing de alguma combinação de uma ordem fixa das derivadas parciais espaciais na posição do pixel.

7. OPERADORES MORFOLÓGICOS

92 - **MORFOLOGIA MATEMÁTICA (MATHEMATICAL MORPHOLOGY)**: A morfologia matemática refere-se a uma área do processamento de imagem voltada à análise de formas. As operações morfológicas básicas consistem em: dilatação, erosão, abrir e fechar uma imagem com elemento estruturado.

93 - **ELEMENTO ESTRUTURANTE (STRUCTURING ELEMENT)**: O elemento estruturante de uma operação morfológica é uma função definida no domínio do modelo espacial do operador morfológico e, cujo valor de cada pixel do domínio é o peso ou coeficiente empregado pelo operador morfológico para posição daquele pixel. O elemento estruturante do operador morfológico tem uma função na morfologia exatamente análoga a função do kernel e uma operação convolução.

94 - **DILATAÇÃO (DILATION)**: Dilatando uma imagem I por um elemento estruturante s tendo suporte ou domínio S , produz uma imagem dilatada denotada por $I \oplus s$, que definido por:

$$(I \oplus s)(r, c) = \text{Max}_{[(i, j) \in S]} \{I(r - i, c - j) + s(i, j)\} \quad (5.15)$$

A dilatação é uma operação comutativa, associativa, invariante por translação e crescente. A Dilatação é a operação contrária da operação de erosão.

95 - **EROSÃO (EROSION)**: Erodindo uma imagem I , por um elemento estruturante s , tendo suporte ou domínio S , produz uma imagem erodida denotada por $I \ominus s$, que definido por:

$$(I \ominus s)(r, c) = \text{Min}_{[(i, j) \in S]} \{I(r + i, c + j) - s(i, j)\} \quad (5.16)$$

A erosão é uma operação invariante por translação e crescente. É a operação contrária a dilatação.

96 - **ABERTURA (OPENING)**: Abrindo uma imagem I , por um elemento estruturante s , tendo suporte ou domínio S , produz uma imagem aberta denotada por $I \circ s$, que definido por:

$$I \circ s = (I \ominus s) \oplus s \quad (5.17)$$

Abrir é uma operação crescente, anti-extensiva e idempotente. É a operação contrário de fechar. Abrindo uma imagem com um disco formatado de elementos estruturantes de contornos suavizados, quebra estreitos isthmuses, e elimina ilhas e cabos menores em tamanho ou largura que o disco do elemento estruturante.

97 - **FECHAMENTO (CLOSING)**: Fechando uma imagem I , por um elemento estruturante s , tendo suporte ou domínio S , produz uma imagem fechada denotada por $I \bullet s$, que definido por:

$$I \bullet s = (I \oplus s) \ominus s \quad (5.18)$$

Fechar é uma operação crescente, extensiva e idempotente. É a operação contrário de abrir. Fechar uma imagem com um disco formatado de elementos estruturados

de contornos suavizados, denota estreitos quebrados e longos e finos golgos, elimina vazios menores que o tamanho do elemento estruturado do disco e abastece fendas no contorno.

98 - AFINAMENTO (**THINNING**): Uma operação de afinamento é uma imagem simbólica do operador da vizinhança, que apaga de forma simétrica todo o interior das bordas dos pixels de uma região que não desconecta a região. Aplicações sucessivas da operação de afinamento reduz a região a um grupo de arcos que constitui um esqueleto da região.

99 - ESPESSAMENTO (**THICKING**): Uma operação de espessamento é uma imagem simbólica do operador da vizinhança, que em algum caminho simétrico agrega todo pixel do fundo perto suficiente da região dentro da região.

8. TRANSFORMADA DE HOUGH

100 - TRANSFORMAÇÃO DE RADON (**RADON TRANSFORM**): A transformação discreta de Radon $R : Q \rightarrow [0, \infty)$ de uma função $I : X \rightarrow [0, \infty)$ relativa à forma funcional $F : X \times Q \rightarrow [0, \infty)$ é definida por:

$$R(q) = \sum_{\{x \in X | F(x, q) = 0\}} I(x) \quad (5.19)$$

101 - TRANSFORMAÇÃO DE HOUGH (**HOUGH TRANSFORM**): A transformação de Hough $H : \mathfrak{S} \rightarrow [0, \infty)$ de uma função $I : X \rightarrow [0, \infty)$ relativa à forma funcional $F : X \times Q \rightarrow [0, \infty)$, onde $\mathfrak{S}$ é uma partição do espaço dos parâmetros Q , é definida por:

$$H(B) = \sum_{\{x \in X | \text{para } q \in B, F(x, q) = 0\}} I(x), \quad B \in \mathfrak{S} \quad (5.20)$$

Quando I tem a forma $I(x) = 1$, para todo $x \in X_0$, onde $X_0 \subseteq X$, e $I(x) = 0$ em outro lugar qualquer, a transformação de Hough de X_0 , relativa a forma funcional $F : X \times Q \rightarrow [0, \infty)$, tem a forma simples definida por:

$$H(B) = \#\{x \in X_0 | \text{para algum } q \in B, F(x, q) = 0\} \quad (5.21)$$

A transformação de Hough é uma transformação que pode auxiliar na detecção de arcos de figuras ou objetos em formato 3D. Cada formato ou figura tem alguns parâmetros livres que quando precisamente especificados definem o arco, formato ou figura. O formato que tem o parâmetro livre q corresponde ao conjunto $\{x \in X | F(x, q) = 0\}$. Os parâmetros livres constituem o domínio da transformação ou o parâmetro espacial da transformação de Hough. Dependendo da informação disponível à transformação de Hough, cada vizinho da imagem ou superfície do objeto sendo transformado mapeará a um ponto ou um conjunto de pontos no parâmetro espacial de Hough. A transformação de Hough modela o parâmetro espacial de Hough em depósitos e conta, para cada depósito, quantos vizinhos na imagem ou superfície do objeto tem um dos seus pontos transformados situados no volume determinado para o depósito.

102 - ESFERA GAUSSIANA (**GAUSSIAN SPHERE**): A esfera Gaussiana refere-se a uma unidade esférica e seu sistema de coordenadas esféricas associado. As quantidades usualmente representadas na esfera Gaussiana são vetores orientados.

103 - ESPAÇO GRADIENTE (**GRADIENTE SPACE**): O espaço do Gradiente é um espaço bi-dimensional cujos eixos representam as derivadas parciais de primeira ordem de uma superfície da forma $z = f(x, y)$. Cada ponto do espaço Gradiente corresponde a orientação de uma possível superfície normal.

104 - IMAGEM GAUSSIANA (**EXTENDED GAUSSIAN IMAGE, ORIENTATION HISTOGRAM**): A imagem estendida Gaussiana ou orientação do histograma de um objeto 3D é um histograma bi-dimensional ou transformação de Hough das orientações da superfície normal do objeto. É calculada encaixando a superfície de uma esfera dentro das células e determinando para cada célula um valor que estima a área total da superfície do objeto tendo a orientação da superfície normal que pertence a célula.

9. GEOMETRIA DIGITAL

105 - DISTÂNCIA EUCLIDIANA (**EUCLIDIAN DISTANCE**): A distância Euclidiana $d(p, q)$ entre dois pontos $p = (p_1, \dots, p_n)$ e $q = (q_1, \dots, q_n)$ é definida por:

$$d(p, q) = \sqrt{\sum_{n=1}^N (p_n - q_n)^2} \quad (5.22)$$

106 - DISTÂNCIA CITY-BLOCK (**CITY-BLOCK DISTANCE**): A distância block ou city $d(p, q)$ entre dois pontos $p = (p_1, \dots, p_n)$ e $q = (q_1, \dots, q_n)$ é definida por:

$$d(p, q) = \sum_{n=1}^N |p_n - q_n| \quad (5.23)$$

107 - DISTÂNCIA QUADRADA (**SQUARE DISTANCE**): A distância quadrada ou distância máxima $d(p, q)$ entre dois pontos $p = (p_1, \dots, p_n)$ e $q = (q_1, \dots, q_n)$ é definida por:

$$d(p, q) = \max_n |p_n - q_n|, \quad n = 1, \dots, N \quad (5.24)$$

108 - TRANSFORMAÇÃO DE DISTÂNCIA (**DISTANCE TRANSFORM**): A transformação da distância de uma imagem binária é uma imagem tendo na posição de cada pixel a sua distância do pixel binário zero mais próximo da imagem de entrada. A distância pode ser distância block, distância Euclidiana, ou distâncias quadradas.

109 - SUBIMAGEM (**SUBIMAGE**): Uma figura F ou uma subimagem F em uma imagem contínua ou digital I é qualquer função F cujo domínio é algum subconjunto A do conjunto de coordenadas espaciais ou células resolutivas, cujo intervalo é o conjunto G de intensidades da imagem e que é definido por $F(r, c) = I(r, c)$ para qualquer (r, c) pertencente a A .

110 - REGIÃO (**REGION**): Uma região R de uma imagem é qualquer subconjunto de células resolutivas no domínio espacial da imagem.

111 - 4-8 VIZINHANÇA (**4-8 NEIGHBORHOOD**): Um par de pixels vizinhos são ditos para ser 4-conectados se eles compartilham um lado comum. Um par de pixels vizinhos são ditos para ser 8-conectados se eles compartilham um lado comum ou um canto comum.

112 - REGIÃO CONECTADA : Uma região R é conectada se houver um caminho entre duas células resolutivas quaisquer contidas em R . Mais precisamente,

R é 4-conectado (8-conectado) se para cada par de células resolutivas (r, c) e (u, v) pertencentes a R , existir alguma seqüência $[(a1, b1), (a2, b2), \dots, (am, bm)]$ de células resolutivas pertencentes a R tal como $(r, c) = (a1, b1)$, $(u, v) = (am, bm)$, e (ai, bi) , é 4-conectado (8-conectado) a $(ai + 1, bi + 1)$, $i = 1, 2, \dots, m - 1$.

113 - COMPONENTE CONECTADO (**CONNECTED COMPONENT**): Um componente conectado ou componente bolha é uma região conectada de tamanho máximo.

114 - SEGMENTO DE RETA DIGITAL (**DIGITAL STRAIGHT LINE SEGMENT**): Um segmento de linha reta digital entre células resolutivas $(r1, c1)$ e $(r2, c2)$ é aquele subconjunto de todos os pixels tais que alguma parte do segmento de reta unindo $(r1, c1)$ e $(r2, c2)$ tem uma interseção não-vazia com a área do pixel.

115 - REGIÃO CONVEXA (**CONVEX REGION**): Uma região R é convexa se para todo par de células resolutivas em R , R contém o segmento de reta digital que une o par de células resolutivas.

116 - PIXEL DE BORDA INTERNA (**INTERIOR BORDER PIXEL**): Um pixel é um pixel de uma borda interna de uma região R se o pixel pertence a R e tem vizinhos fora de R .

117 - PIXEL DE BORDA EXTERNA (**EXTERIOR BORDER PIXEL**): Um pixel é um pixel de uma borda externa de uma região R se o pixel não pertence a R e tem vizinhos pertencentes a R .

118 - PIXEL INTERNO (**INTERIOR PIXEL**): Um pixel é um pixel interno de uma região R se todos os pixels vizinhos pertencem a R .

119 - BORDA SIMPLES (**SIMPLE BOUNDARY**): Uma borda simples é uma curva orientada fechada que não toca ou atravessa a si mesma. Pixels que estão dentro de uma borda simples constituem uma região conectada que não têm buracos.

120 - CONTORNO (**BOUNDING CONTOUR**): O contorno da borda de uma região R consiste da borda simples que envolve os pixels de R .

121 - BURACO (**HOLE**): Um conjunto de pixels H constitui um buraco de uma região R se H é um conjunto máximo conectado de pixels que não pertencem a R mas que estão ao redor de R .

122 - BORDA (**BORDER, BOUNDARY**): A borda ou contorno de uma região conectada R consiste de seu contorno e do (possivelmente vazio) conjunto de bordas simples cada qual ao redor dos pixels pertencentes a algum buraco de R .

123 - PERSEGUIÇÃO DE CONTORNO (**BOUNDARY FOLLOWING**): A perseguição de contorno refere-se ao procedimento seqüencial pela qual a lista de pixels da borda de uma região podem ser determinadas.

124 - CONCURVA (**CONCURVE**): Uma concurva é uma curva contínua, usualmente representando um contorno conexo, consistindo de uma corrente conectada de arcos simples descritos.

125 - POLÍGONO PERIMETRAL MÍNIMO (**SIMPLE BOUNDARY**): O polígono perimetral mínimo de uma curva digital C é o polígono de menor tamanho cuja digitalização é C . O formato do polígono perimetral mínimo é freqüentemente similar ao formato geral de uma curva digital.

126 - BUSCA DE CONTOURNO (**CONTOUR TRACING**): A busca de contorno é um processo de procurar ou percorrer onde o contorno da borda de uma bolha pode ser identificado.

10. DESCRIÇÃO 2D

127 - PERÍMETRO (**PERIMETER**): O perímetro de uma região conectada R é o comprimento do contorno da borda de R .

128 - ÁREA (**AREA**): A área A de uma região R é definida por:

$$A = [\#R].s \quad (5.25)$$

onde s é o fator de escala que especifica a área de um pixel.

129 - CENTRÓIDE (**CENTROID**): O centróide de uma região R é o centro de massa da região $(\bar{r}, \bar{c})$. Esta é a posição média (linha, coluna) para todos os pixels na região e é dado por:

$$\bar{r} = \frac{1}{\#R} \sum_{(r,c) \in R} r, \quad e \quad \bar{c} = \frac{1}{\#R} \sum_{(r,c) \in R} c \quad (5.26)$$

130 - MOMENTO (**MOMENT**): O (j, k) -ésimo momento M_{jk} de um formato digital S é dado por:

$$M_{jk} = \sum_{(r,c) \in S} r^j c^k \quad (5.27)$$

O centro de gravidade $(\bar{r}, \bar{c})$ de S pode ser expresso em termos dos momentos de S :

$$\bar{r} = \frac{M_{10}}{M_{00}} \quad e \quad \bar{c} = \frac{M_{01}}{M_{00}} \quad (5.28)$$

131 - MOMENTO CENTRAL (**CENTRAL MOMENT**): O (j, k) -ésimo momento central μ_{jk} de um formato digital S é dado por:

$$\mu_{jk} = \sum_{(r,c) \in S} (r - \bar{r})^j (c - \bar{c})^k \quad (5.29)$$

132 - MOMENTO CENTRAL NORMALIZADO (**NORMALIZED CENTRAL MOMENT**): O (j, k) -ésimo momento central normalizado η_{jk} de S é dado por:

$$\eta_{jk} = \frac{\mu_{jk}}{\mu_{00}^\gamma}, \quad \text{onde } \gamma = \frac{j+k}{2} + 1 \quad (5.30)$$

133 - MOMENTOS DO MOVIMENTO DE ROTAÇÃO (**ROTATION MOVEMENT MOMENTS**): Os momentos do movimento de rotação ϕ_k de S são dados por:

$$\begin{aligned} \phi(1) &= \eta_{20} + \eta_{02} \\ \phi(2) &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi(3) &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \end{aligned}$$

$$\begin{aligned}
\phi(4) &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi(5) &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\phi(6) &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi(7) &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] \\
&\quad - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{5.31}$$

134 - NÚMERO DE EULER (**EULER NUMBER**): O número de Euler de uma região é o número de seus componentes conectados menos o número de seus buracos.

135 - COMPACIDADE (**COMPACTNESS**): A compacidade de uma objeto pode ser medida pelo comprimento de seu perímetro quadrado dividido por sua área ou alternativamente medida pelo desvio padrão dos raios que vão do centroid a borda dividido pelo raio médio. A medida clássica do perímetro quadrado dividido pela área, que tem a desvantagem de que no domínio digital leva seu menor valor não por um círculo digital mas por um octógono ou diamante dependendo se a 8-conectividade ou 4-conectividade, é usada no cálculo do perímetro.

136 - RETÂNGULO DELIMITADOR (**BOUNDING RECTANGLE**): O retângulo do contorno de uma região R é um retângulo que circunscreve R . Tem os seus lados alinhados com a direção das linhas e das colunas, seu lado mais à esquerda alinhado com a coluna de número mais baixo de R , seu lado mais à direita alinhado com a coluna de número mais alto de R , seu lado mais alto alinhado com o número mais baixo da linha de R , e seu lado mais baixo alinhado com o maior número da linha de R .

137 - PIXEL EXTREMO (**EXTREMAL PIXEL**): Um pixel extremo de R é um pixel de R tendo dentre de todos os pixels em R ambos:

- (a) uma coordenada de linha extrema com valor r e uma coordenada de coluna extrema com valor entre todas as posições da coluna c tais que $(r, c) \in R$,
- (b) uma coordenada de coluna extrema com valor c e uma coordenada de linha extrema com valor entre todas as posições da linha r tais que $(r, c) \in R$.

Uma região pode ter até oito pontos extremos distintos, cada qual deve estar situado no retângulo da borda da região. Pixels extremos podem ser usados para representar a área estendida de uma região e para inferir o comprimento e a orientação do eixo dominante da região.

138 - MATRIZ DO SEGUNDO MOMENTO (**SECOND MOMENT MATRIX**): A matriz do segundo momento de uma região R é definida por:

$$\sigma = \begin{pmatrix} \mu_{rr} & \mu_{rc} \\ \mu_{cr} & \mu_{cc} \end{pmatrix} \tag{5.32}$$

de uma região R é definida por:

$$\begin{aligned}
\mu_{rr} &= \frac{s_r^2}{A} \sum_{(r,c) \in R} (r - \bar{r})^2 \\
\mu_{rc} &= \frac{s_r s_c}{A} \sum_{(r,c) \in R} (r - \bar{r})(c - \bar{c})
\end{aligned}$$

$$\mu_{cc} = \frac{s_c^2}{A} \sum_{(r,c) \in R} (c - \bar{c})^2 \quad (5.33)$$

onde A é a área da região, S_r é o fator de escala da linha, e S_c é o fator de escala da coluna, $\bar{r}$ é a linha do centroid, e $\bar{c}$ é a coluna do centroid.

139 - **ELONGAÇÃO (ELONGATION)**: O alongamento de uma região conectada pode ser medida de várias maneiras. Uma técnica é usar a razão do comprimento do máximo comprimento da corda na bolha com o comprimento de seu comprimento perpendicular máximo da corda. A segunda técnica é usar a raiz quadrada da razão do maior com o menor autovalor da matriz do segundo momento central da região. Uma terceira técnica é usar a razão da maior distância entre um par oposto de pontos extremos e a distância entre aquele e o par de pontos extremos tendo próximo a maior distância.

140 - **EIXO SIMÉTRICO (MEDIAL AXIS)**: O eixo simétrico ou eixo médio de uma região é um subconjunto de pixels de bolha que são o centro das linhas máximas, quadrados ou discos que estão contidos na região. A cada pixel que é parte de um eixo simétrico pode também estar associado informação adicional tal como o tamanho da linha máxima ou quadrado ou o raio máximo do disco em que está centrado.

141 - **COMPONENTE CONECTADO (CONNECTED COMPONENT)**: O operador do componente conectado tem como entrada uma imagem binária e produz como saída uma imagem em que cada pixel binário um é dado um único rótulo de máximo componente conectado de pixels tendo um valor binário um para o qual ele pertence.

142 - **ANÁLISE DE COMPONENTES CONECTADOS (CONNECTED COMPONENT ANALYSIS)**: Na análise dos componentes conectados ou análise de conjuntos, a posição e as propriedades de forma de cada componente conectado são medidas. As propriedades típicas de forma incluem área, perímetro, número de buracos, retângulo de fronteira (bounding rectangle), pontos extremos, centróide, momentos de segunda ordem, e orientação derivada dos momentos de segunda ordem ou pontos extremos. Os componentes conectados são então identificados ou classificados por uma regra de decisão com base em suas características medidas.

143 - **ANÁLISE DE ASSINATURA (SIGNATURE ANALYSIS)**: A assinatura de uma imagem binária analisa a imagem binária em termos de suas projeções. Projeções podem ser vertical, horizontal, diagonal, circular, radial, espiral, ou projeções gerais. A análise consiste calcular as projeções, segmentando cada projeção, pegando as propriedades medidas de cada segmento de projeção. A análise da assinatura também pode usar a segmentação de projeção para induzir a segmentação de uma imagem.

144 - **PROJEÇÃO DE UMA IMAGEM BINÁRIA (BINARY IMAGE PROJECTION)**: A projeção de uma imagem binária é o histograma de uma imagem em tons de cinza produzido pela máscara da imagem índice de projeção com a imagem binária dada. Cada pixel da imagem índice de projeção contém um número que é o índice do valor da projeção bin para o qual o pixel pertence. Um histograma da imagem de projeção mascarada então contém na projeção bin i o número de pixels os quais na imagem binária tem o valor binário 1 e os quais na imagem índice de

projeção contém o valor de índice i .

11. ESTRUTURA DE DADOS DE CURVAS E IMAGENS

145 - DESCRITORES DE FOURIER (**FOURIER DESCRIPTORS**): Os descritores de Fourier de uma curva planar fechada são os coeficientes das séries de Fourier de posições espaciais de uma curva considerada como uma função de comprimento de arco. Tipicamente os menores coeficientes de frequência são os de maior interesse.

146 - AJUSTE ITERATIVO DE CURVAS POR PONTOS FINAIS (**ITERATIVE END POINT CURVE FITTING**): O ajuste iterativo de curvas a partir de pontos finais refere-se a um processo iterativo de segmentação de uma curva em um set de segmentos lineares de *piecewise* o qual aproxima a curva. O processo começa construindo uma linha direta entre os pontos finais da curva. Se a distância promovida entre a curva e a linha direta é menor que uma tolerância especificada então a aproximação é considerada para ser adequada e o segmento de curva é dividido sem promover. Se a distância promovida entre a curva e a linha direta for maior que a tolerância especificada, então a aproximação é considerada ser não adequada. A curva então é dividida em dois segmentos deste ponto de distância promovida e o processo de ajuste da linha direta continua independentemente em cada segmento.

147 - CÓDIGO DE CADEIA (**CHAIN CODE**): A representação por código de cadeia de um arco digital ou fronteira de conjunto é uma seqüência onde cada elemento é um símbolo representando um vetor unindo dois pixels vizinhos de um arco digital ou fronteira de conjunto. O código de cadeia mais comum usa os símbolos 0 a 7 para representar os vetores $(0, 1)$, $(-1, 1)$, $(-1, 0)$, $(-1, -1)$, $(0, -1)$, $(1, -1)$, $(1, 0)$, e $(1, 1)$ das coordenadas de linhas de colunas, os quais podem unir dois pixels vizinhos. Códigos de cadeias mais complexos tem mais símbolos e podem representar o vetor unindo 2 pixels mais distantes os quais define o início e o final do segmento da linha direta digital o qual é a parte de um arco ou fronteira de conjunto.

148 - QUADTREE (**QUADTREE**): Um quadtree é uma estrutura de dados em árvore que representa uma imagem. Cada nó de um quadtree representa um subconjunto quadrado do domínio espacial de uma imagem. O nó raiz de uma quadtree representa o domínio espacial de uma imagem inteira. Se todos os pixels do subconjunto do domínio espacial representado pelo nó tiverem os mesmos valores, então o valor do nó é o valor dos pixels do subconjunto. Tal nó é chamado nó puro. Se o nó é um nó misturado ou impuro, então o quadrado representado pelo nó é particionado em 4 quadrantes e o nó terá 4 nós filhos, um nó filho em cada quadrante. Se a imagem inicial representar uma imagem binária, então o quadtree correspondente é chamado de *quadtree binário*. Se a imagem inicial representar uma imagem em tons de cinza, então o quadtree correspondente é chamado *quadtree em tons de cinza*.

149 - CURVA PIRAMIDAL (**PYRAMIDAL CURVE**): Uma curva "piramidal" consiste na seqüência de imagens simbólicas representando curvas de múltiplas resoluções. A operação principal em uma construção bottom-up consiste em conectar pequenos segmentos locais de uma curva em segmentos maiores. Se os segmentos são descritos por "relações de curva" binárias de lados rotulados de uma célula quadrada, a concatenação dos pequenos segmentos podem ser guardados formalmente pegando o fechamento transitivo das relações da curva. A sobreposição das pirâmides são necessárias se o resultado da pirâmide tem a "propriedade de redução

de comprimento”: curvas longas com muitos segmentos sobrevivem a altos níveis, considerando que curvas pequenas desaparecem depois de poucos passos de redução.

150 - OCTREE (**OCTREE**): Um octree é um dado em uma estrutura de dados em árvore que representa a função definida como um volume de espaço 3-D. Cada nó de um octree representa o volume de um subconjunto de um cubo. O nó raiz do quadtree representa o volume inteiro. Se todos os voxels representados por um nó tem o mesmo valor de funções, então o valor do nó é os valores daquelas funções. De tal modo um nó é chamado de nó puro. Se o nó é um nó misturado ou impuro, então o cubo representado pelo nó é particionado em 8 volumes octantes e o nó terá oito filhos, um filho para cada octante. Se a função é binária, então o octree correspondente é chamado *octree binário*. O octree binário é útil para representar volumes tridimensionais. Se a função inicial representada é não-binária, tal como uma função valendo real ou integral, então a octree correspondente é chamada *octree em tons de cinza*.

151 - RUN LENGTH ENCODING (**RUN LENGTH ENCODING**): O Run length encoding é uma maneira de representar de forma compactada imagens binárias. Há uma variedade de formatos para o run length encoding. Cada formato tem uma maneira de representar o início da posição da coluna de uma cadeia horizontal (de máxima extensão) de pixels binários valendo 1(um) assim como o número de pixels em uma varredura. Muitos sistemas de visão que reconhecem objetos de imagens binárias usam o run length encoding para reduzir o volume de dados a ser processados.

152 - CONE GENERALIZADO ou CILINDRO GENERALIZADO (**GENERALIZED CONE or CYLINDER**): Um cone generalizado ou cilindro generalizado é uma estrutura de dados para uma representação volumétrica de objetos 3D. O volume é gerado varrendo uma seção arbitrária em forma de cruz ao longo de uma curva 3D chamada de "eixo generalizado do cone" ou "espinha generalizada do cilindro". O eixo passa pelo centróide da seção em cruz e é posicionado num ângulo fixo (usualmente ortogonal) para eles. A seção em forma de cruz é permitido para ter alguns parâmetros livres tal como tamanho ou alongamento. Estes valores são especificados para cada medida de eixo pela função de seção em cruz ou regra de varredura.

153 - GEOMETRIA CONSTRUTIVA DE SÓLIDOS (**CONSTRUCTIVE SOLID GEOMETRY**): Um geometria construtiva de sólidos é um mecanismo para representar volumes tridimensionais por um processo construtivo que começa com volumes de formas simples e que são combinadas e subtraídas uns dos outros por um conjunto de operações de união, interseção e conjunto diferença entre conjuntos.

154 - DESCRIÇÃO DE SUPERFÍCIE DE BORDAS ou REPRESENTAÇÃO DE BORDAS (**BOUNDARY SURFACE DESCRIPTION or BOUNDARY REPRESENTATION**): Uma descrição de superfície de borda ou representação de borda de um objeto tridimensional ou volume é uma representação que contém cada uma das bordas superficiais do volume. Cada borda de superfície é representada em termos de pedaços simplesmente descritas, cada um dos quais tem sua própria borda de arco. Cada um dos arcos é representado em termos de pedaços simplesmente descrito os quais começam e terminam em pontos finais ou vértices.

155 - SUPERQUADRÁTICA (**SUPERQUADRATIC**): Uma superquadrática é uma superfície fechada média por um vetor cujos componentes x , y e z são especificados a partir dos ângulos η e ω , via produto esférico de duas curvas parametrizadas bidimensionais:

$$\begin{aligned}
h &= \begin{pmatrix} h_1(\eta) \\ h_2(\eta) \\ e \end{pmatrix} \\
m &= \begin{pmatrix} m_1(\omega) \\ m_2(\omega) \end{pmatrix}
\end{aligned} \tag{5.34}$$

o qual provêm de um das três formas trigonométricas básicas. O produto esférico de h com m escalado pelo vetor $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ é definido por:

$$h \otimes m = \begin{pmatrix} a_1 h_1(\eta) m_1(\omega) \\ a_2 h_2(\eta) m_2(\omega) \\ a_3 h_2(\omega) \end{pmatrix} \tag{5.35}$$

156 - STRIP TREE (STRIP TREE): Um strip tree é uma estrutura de dados de árvore binária que representa de forma hierárquica um segmento de arco planar. O nó raiz de uma árvore representa o retângulo de tamanho mínimo que cerca o arco. Cada nó não-terminal de uma árvore divide o arco que seu retângulo de fronteira (bounding rectangle) aproxima em dois pedaços contínuos. Cada um dos dois nós filhos contém o retângulo de tamanho mínimo que cerca o pedaço de segmento do arco pertencente a ele. Os nós terminais de uma árvore tem retângulos de fronteira que são suficientemente pertos do segmento de arco que eles contém.

157 - ESBOÇO PRIMITIVO (PRIMAL SKETCH): Um esboço primitivo é uma estrutura de dados para representar as mudanças de intensidade de tons de cinza, distribuição geométrica, e organização em cada vizinhança da imagem. As primitivas de um esboço primitivo incluem passagem por zero, conjunto, terminações e descontinuidades, segmentos de bordas, linhas virtuais, grupos, organizações curvilíneas, e fronteiras.

158 - MODELO DE FACETAS (FACET MODEL): Em um processamento de imagem de modelo de facetas, os valores dos pixels da imagem digital são considerados como discretizadas ruidosas de sua subjacente e desconhecida superfície de intensidade em tons de cinza. Qualquer operação a ser realizado em uma imagem é definida em termos desta superfície de intensidade em tons de cinza subjacente. Assim, em ordem para fazer qualquer processamento, a superfície de intensidade subjacente deve ser estimada. Isto requer um modelo que descreva qual seria a forma geral da superfície em qualquer vizinhança da imagem se não tivesse ruído. Estimar a superfície a partir da vizinhança envolta do pixel exige estimar os parâmetros livres para a forma geral. Operações de processamento de imagens proveitosas que então podem ser feitos usando a aproximação do modelo de faceta incluem a detecção de bordas por gradiente, detecção de bordas por passagem por zero, segmentação de imagem, detecção de linhas, detecção de cantos, estimação de formas 3D por sombreamento, e determinação de fluxo óptico.

12. TEXTURAS

159 - CARACTERÍSTICA TONAL DISCRETA (DISCRETE TONAL FEATURE): Uma característica tonal discreta em uma imagem digital é um conjunto conectado de todas as células de resolução que tem a mesma ou quase a mesma intensidade da imagem.

160 - TEXTURA (**TEXTURE**): Uma Textura está relacionada com a distribuição espacial da intensidade da imagem e as características tonais discretas. Quando uma área pequena de uma imagem tem uma pequena variação de características de tons discretos, a propriedade dominante naquela área é o tom cinza. Quando uma pequena área tem grande variação de características tonais discretos, a propriedade dominante naquela área é a textura. Há 3 fatos cruciais para esta distinção: (1) o tamanho das áreas pequenas, (2) os tamanhos relativos das características tonais discretos, e (3) o número das características de tons discretos distinguíveis. Textura pode se descrita na forma de dimensões de uniformidade, densidade, espessura, aspereza regular, intensidade e direção.

161 - TEXTEL (**TEXTEL**): Um texel, abreviatura para elemento de textura, é uma tripla cujo primeiro componente é a localização (linha, coluna) de uma vizinhança pequena, cujo segundo componente é o tamanho da vizinhança, e cujo terceiro componente é o vetor das propriedades da textura.

162 - MATRIZ DE DEPENDÊNCIA ou DE CO-OCORRÊNCIA EM NÍVEIS DE CINZA (**GRAY LEVEL DEPENDENCE MATRIX or GRAY LEVEL CO-OCURRENCE MATRIX**): A matriz de dependência em níveis de cinza ou matriz de co-ocorrência em níveis de cinza caracteriza a micro-textura de uma região de uma imagem pela medida da dependência entre os pares de níveis de cinza dos pixels em uma relação espacial especificada. Para um par de nível de cinza (i, j) a matriz de dependência em nível de cinza P para a região R da imagem I tem o valor $P(i, j)$ quando $P(i, j)$ é o número de pares de pixels em uma região que tenha a relação espacial desejada onde o primeiro pixel tem o nível de cinza i e o segundo pixel tenha o nível de cinza j . Se S designar o conjunto de todos os pares de pixels em uma relação espacial desejada, então $S \subseteq R \times R$ e:

$$P(i, j) = \# \{ (r_1, c_1), (r_2, c_2) \in S \mid I(r_1, c_1) = i \text{ e } I(r_2, c_2) = j \} \quad (5.36)$$

A matriz co-ocorrência em níveis de cinza P pode ser normalizada. Uma forma normalizada que produz uma probabilidade de união é dada por:

$$P_1(i, j) = \frac{P(i, j)}{\#} \quad (5.37)$$

A segunda forma normalizada que produz a probabilidade condicional é dada por:

$$P_2(i, j) = \frac{P(i, j)}{\sum_j P(i, j)} \quad (5.38)$$

163 - MEDIDAS ESTATÍSTICAS DE TEXTURA (**STATISTICAL TEXTURE MEASURES**): As medidas estatísticas de textura incluem os momentos dos níveis de cinza de uma região dada, tipicamente a variância, a slewness e a curtose.

164 - HISTOGRAMA DE DIFERENÇAS DE NÍVEIS DE CINZA (**GRAY LEVEL DIFFERENCE HISTOGRAM**): Um histograma de diferença dos níveis de cinza a uma distância d de uma imagem I é o histograma com os valores

$$\|I(r_1, c_1) - I(r_2, c_2)\|; \mid (r_1 - r_2)^2 + (c_1 - c_2)^2 = D^2 \quad (5.39)$$

165 - DESCRIÇÃO ESTRUTURAL DE TEXTURA (**STRUCTURAL TEXTURE DESCRIPTION**): Uma descrição de textura estrutural é dada por um conjunto de primitivas e regras de posicionamento que governam as relações espaciais estocásticas entre eles.

166 - DESCRIÇÕES DE TEXTURA RELACIONADAS A FOURIER (FOURIER RELATED TEXTURE DESCRIPTIONS): As descrições de textura relacionadas a Fourier incluem o espectro de potência e função de auto-correlação.

13. SEGMENTAÇÃO

167 - SEGMENTAÇÃO DE IMAGEM (IMAGE SEGMENTATION): Segmentação de imagem é um processo que tipicamente particiona o domínio espacial de uma imagem em subconjuntos mutuamente exclusivos, chamados regiões, onde cada uma destas é uniforme e homogêneo com respeito a algumas propriedades como tom ou textura e cujos valores destas propriedades diferem em alguns aspectos e significados das propriedades de cada região vizinha. Regiões produzidas pelo processo de segmentação de imagens usam a intensidade da imagem como uma propriedade da região gerada que são chamadas de características discretas dos tons.

168 - CRESCIMENTO DE REGIÃO (REGION GROWING) : O crescimento de região refere-se ao processo de segmentação seqüencial da imagem onde pixels são sucessivamente adicionados a regiões incompletas ou a novas regiões geradas quando este processo não é apropriado para criar partes de uma região incompleta existente. Existem três classes básicas de region growing: region tracking, region aggregation e region merging.

(a) **169 - BUSCA DE REGIÃO (REGION TRACKING):** Em busca de regiões, a imagem é explorada em uma ordem de procura por rastreamento. A similaridade de cada pixel é comparada com as regiões já processadas com 4 ou 8 pixels vizinhos conectados cada um em seu próprio lugar. Se um dos pixels vizinhos já processado estiver em um lugar na região de similaridade suficiente com o pixel corrente, então o pixel é adicionado à região. Se a região para a qual todos os pixels já foram processados não estiver similar com o pixel corrente, então o pixel corrente inicia uma nova região.

(b) **170 - AGREGAÇÃO DE REGIÃO (REGION AGGREGATION):** Em agregação de regiões, são encontrados primeiro os pixels que servem como protótipo para as regiões que se deseja segmentar. Então de forma seqüencial, pixels que tem vizinhos em alguma região incompleta unem-se dentro da região se eles forem similares o suficiente. O processo de agregação continua até que todos os pixels façam parte de alguma região.

(c) **171 - FUSÃO DE REGIÃO (REGION MERGING):** Em fusão de regiões, as regiões vizinhas de uma segmentação inicial são sucessivamente fundidos na mesma região se elas possuem propriedades similares o suficiente. Após cada interação merge, as propriedades das novas regiões são recalculadas. As interações merge continuam até que as propriedades de cada par de regiões vizinhas sejam suficientemente diferentes umas das outras.

172 - PREENCHIMENTO DE CONTORNO (CONTOUR FILLING): O preenchimento de contorno é o processo pelo qual todos os pixels internos a um contorno são marcados com um único rótulo.

14. MATCHING

173 - EMPARELHAMENTO DE PADRÕES (TEMPLATE MATCHING): O emparelhamento de padrões é uma operação que pode ser usada para encontrar como saída uma sub-imagem temporária igual a uma janela de imagem dada. Os graus de semelhança é muitas vezes determinada através da translação de todas as sub-imagens temporárias sobre a imagem dada e para cada posição avaliar a

correlação cruzada ou a soma da raiz quadrada ou ainda as diferenças de intensidade da imagem absoluta dos pixels correspondentes. Template matching também pode ser usado para melhor adequar um modelo observado com o modelo protótipo.

174 - **FILTRO DE EMPARELHAMENTO (MATCHED FILTERING)**: O filtro de emparelhamento é uma operação de emparelhamento executada usando a magnitude da função de correlação cruzada para medir o grau do emparelhamento.

175 - **EMPARELHAMENTO DE IMAGENS (IMAGE MATCHING)**: O emparelhamento de imagens refere-se ao processo de determinar o pixel-por-pixel, o arco-por-arco, ou região-por-região, correspondente entre duas imagens obtidas de algum cenário porém com diferentes sensores, diferentes iluminações ou um diferente ângulo de visão. Image matching pode ser usado em um modelo de classificação spectral/temporal de sentido remoto, ou em determinados pontos correspondentes no espaço, posição, mudança de análise e movimentação da análise. Em um grupo de aproximação, sub-imagens de uma imagem são transladadas provocando uma segunda imagem excedente. Para cada translação as diferenças entre transformações apropriadas de intensidades de tons de cinza e/ou são limitadas pelas bordas. Neste nível aproximado, a única função existente com estes pixels, é determinar as diferenças entre os valores dos pixels que delimitam duas imagens.

176 - **EMPARELHAMENTO SIMBÓLICO (SYMBOLIC REGISTRATION)**: Em emparelhamento simbólico, trabalha-se por métodos de alto nível. Por exemplo a cena pode ser segmentada e a region matching é portanto efetuada usando características como área, posição, $(\text{perímetro})^2/\text{área}$, orientação, proporção entre comprimento e largura, $\text{área}/(\text{área do retângulo mínimo})$, $\text{área}/\text{área do bounding ellipse}$, intensidades de tons de cinza ou cor do segmento e número de vizinhos correspondentes. Esse processo é chamado de igualdade simbólica ou registro simbólico porque esse matching emprega uma unidade de nível mais elevado que o pixel.

177 - **EMPARELHAMENTO DE PONTOS CARACTERÍSTICOS (FEATURE POINT MATCHING)**: Em um emparelhamento de pontos característicos, pontos selecionados de cada imagem são determinados primeiro, na base dos valores distintos da imagem da vizinhança ou na base de uma interseção entre duas linhas características. O local de cada ponto pode ser com precisão de sub-pixel. Após a localização dos pontos distintos ter sido determinada, um processo correspondente associa muitos dos possíveis pontos selecionados de uma imagem com pontos selecionados em uma segunda imagem. A correspondência é baseada na similaridade das características dos pontos.

178 - **DESCRIÇÃO ESTRUTURAL (STRUCTURAL DESCRIPTION)**: Uma descrição estrutural é uma representação relacional de entidades 2D ou 3D. Isto consiste em uma série de primitivas onde cada uma contém a descrição de seus atributos e um conjunto de nomes de relações que consiste de tuplas cujos componentes são primitivas que encontram-se na relação especificado pelo nome da mesma.

Uma função $h : A \rightarrow B$ é um homomorfismo relacional da N-array relação $R \subseteq A^N$ para a N-array relação $S \subseteq B^N$ se

$$R \circ h \subseteq S \quad (5.40)$$

onde $R \circ h = \{(b_1, \dots, b_N) \in B^N \mid \text{para algum } (a_1, \dots, a_N) \in R, b_n = h(a_n), n = 1, \dots, N\}$

179 - **RELAÇÃO DE EMPARELHAMENTO (MATCHING RELATION)**: A relação R é chamada relação de matching ou de emparelhamento S se existe nela uma relação de homomorfismo satisfazendo:

$$R \circ h = S \text{ e } S \circ h^{-1} = R \quad (5.41)$$

180 - **EMPARELHAMENTO RELACIONAL (RELATIONAL MATCHING)**: O emparelhamento relacional refere-se ao processo pelo qual é determinado se duas relações emparelham ou não emparelham. Um casamento estrutural é um matching que estabelece uma correspondência ou homomorfismo das primitivas de uma primeira estrutura de descrição para as primitivas de uma segunda estrutura de descrição. Em um casamento ideal, uma tupla de primitivas que encontram-se em uma dada relação na primeira descrição tem esta correspondência com a tupla de primitivas situada em uma dada relação na segunda estrutura de descrição.

181 - **MÉTODO DE FOCALIZAÇÃO DE CARACTERÍSTICAS LOCAIS (LOCAL FEATURE FOCUS METHOD)**: O local feature focus method é um modelo baseado no reconhecimento do objeto e na técnica de localização em que uma característica, referenciada como sendo a característica de foco, na imagem é estabelecido e é usado junto com o modelo do objeto para um predicado onde outra característica aproxime-se dele. Após localizar um conjunto de características, um matching relacional é realizado para inferir uma correspondência consistente entre todas as características localizadas na imagem e as características do objeto. Uma vez que uma correspondência consistente tenha sido estabelecida, a posição e orientação do objeto pode ser suposta. A suposta orientação e posição do objeto é então verificada pelo matching temporário.

182 - **RELAXAÇÃO (RELAXATION)**: A relaxação refere-se a um mecanismo computacional que emprega uma série de processos paralelos interagindo localmente, associado a uma única imagem, que, de uma forma iterativa, atualiza o processo de rotulação de cada unidade corrente de maneira a efetuar uma consistência global na interpretação dos dados da imagem. Em relaxação discreta, a taxação do estado de cada unidade corrente consiste no subconjunto de rótulos ainda não controlado. Em relaxação probabilística, a taxação do estado de cada unidade corrente é a função probabilística associando a cada possível rótulo uma probabilidade de seu estado correto.

15. LOCALIZAÇÃO

183 - **OPERADOR DE INTERESSE (INTEREST OPERATOR)**: O operador de interesse é um operador de vizinhança, que é projetado para localizar, com alta exatidão espacial, posições de pixels ou de subpixels, cujas vizinhanças centrais têm padrões de níveis de cinza distintos. Tais vizinhanças são tipicamente aquelas cuja função de autocorrelação diminua rapidamente. Interest operators são freqüentemente usados para marcar pixels num par de imagens obtidas da mesma cena, apenas deslocando-se a posição de cada câmera ou do objeto de uma primeira imagem para outra. Os pixels marcados, são pixels candidatos de entrada para um algoritmo que estabelece correspondência entre os pixels selecionados.

184 - **ÁREA DO OPERADOR DE INTERESSE (AREA OF INTEREST OPERATOR)**: A área do operador de interesse é um operador que delimita regiões de uma imagem que têm potenciais padrões interessantes. Estas regiões de interesse são aquelas que serão futuramente processadas.

185 - **SCREENING (SCREENING)**: O Screening é uma operação de seleção de fotografias ou imagens contendo áreas de interesse potencial num conjunto de fotografias, sendo que algumas ou até mesmo várias das mesmas contém nenhuma área interessante.

186 - **ANÁLISE DE ÁREA (AREA ANALYSIS)**: Numa análise de área, a área da imagem contendo objetos ou entidades a serem processadas é localizada por um algoritmo simples e um algoritmo mais complexo é somente aplicada na área localizada. Esta estratégia de processamento pode muitas vezes aumentar a velocidade de execução. O algoritmo de localização da área a ser processada é chamado por (**focus of attention mechanism**).

16. PROCESSAMENTO GERAL DE IMAGENS

187 - **PREPROCESSAMENTO (PREPROCESSING)**: O preprocesamento é uma operação aplicada antes que a identificação do modelo seja realizada. O preprocesamento produz, para as categorias de interesse, características que tendem a ser invariáveis frente a mudanças como translação, rotação, mudança de escala, nível de iluminação e o ruído. Na essência, o preprocessing converte os padrões de medição para uma forma que possibilite uma simplificação da regra de decisão. Preprocessing pode trazer um registro, trazer uma coerência, remover ruídos, realçar imagens, segmentar padrões alvo, detectar, centralizar, e normalizar objetos de interesse.

188 - **REGISTRATION (REGISTRATION)**: O Registering ou registration é um processo de alinhamento de translação ou translação e rotação pelo qual duas imagens de mesma geometria e de mesmo conjunto de objetos são posicionados coincidentemente com uma outra tal como pontos correspondentes da imagem da cena apareçam na mesma posição nas imagens registradas. Desta maneira, os valores da imagem correspondente podem ser produzidos para representar a saída do sensor para o mesmo ponto do objeto na sequência completa de imagens.

189 - **CONGRUÊNCIA (CONGRUENCING)**: A Congruência é um processo de deformação geométrica pelo qual duas imagens de diferentes geometrias, embora de mesmo conjunto de objetos, são espacialmente transformadas tal como, nas duas imagens, o tamanho, a forma, a posição e a orientação de qualquer objeto sejam idênticos.

190 - **COMPRESSÃO DE IMAGEM (IMAGE COMPRESSION)**: A compressão de imagens é uma operação que preserva toda ou pelo menos a maioria das informações na imagem e que reduz a quantidade de memória necessária para armazenar uma imagem ou o tempo necessário para transmitir uma imagem.

191 - **RESTAURAÇÃO DE IMAGEM (IMAGE RESTORATION)**: A restauração de imagens é o processo pelo qual uma imagem degradada é restaurada, da melhor maneira possível, até sua condição ideal. A restauração perfeita de uma imagem é possível somente na extensão que a degradação é matematicamente invertível. Formas comuns de restauração incluem (**inverse filtering**, **Wiener filtering** e (**constrained least squares filtering**).

192 - **RECONSTRUÇÃO DE IMAGEM (IMAGE RECONSTRUCTION)**: A reconstrução de imagens refere-se ao processo de reconstrução de uma imagem a partir de um conjunto de suas projeções. As projeções podem ser capturadas ao longo de uma série de raios paralelos, no caso deles serem chamados *parallel projections*, ou eles podem ser capturados ao longo de um conjunto de raios proveniente

de um ponto, nesse caso eles serão chamados de fan beam projections. As técnicas de reconstrução mais comumente empregadas são a filtered back projection e a algebraic reconstruction. Técnicas de reconstrução de imagem são importantes na tomografia computadorizada, na medicina nuclear e nas imagens de ultra-som.

193 - REALCE DE IMAGEM (IMAGE ENHANCEMENT): O realce de imagens é qualquer processo do grupo de operações que melhora a detecção dos objetos. Estas operações incluem, porém não são limitadas à elas, o alongamento de contraste, o realce de bordas, filtragem espacial, a supressão de ruídos, a suavização da imagem e a agudização da imagem.

194 - PROCESSAMENTO DE IMAGEM (IMAGE or PICTURE PROCESSING): O processamento de imagens inclui todas as mais variadas operações que podem ser aplicadas numa imagem. Estas incluem, porém não são limitadas à elas, a compressão de imagem, a restauração de imagem, o realce da imagem, o pré-processamento, a quantização, a filtragem espacial, o matching e as técnicas de reconhecimento.

195 - PROCESSAMENTO INTERATIVO DE IMAGEM (INTERACTION IMAGE PROCESSING): O processamento interativo de imagens é realizado por um operador ou analista em um console com meios de um acessar, de pré-processar, de extrair de características, de classificar, de identificar e de mostrar a imagem original ou a imagem processada para uma avaliação subjetiva e para interações futuras.

17. VISÃO

196 - LUZ ESTRUTURADA (STRUCTURED LIGHT): Luz estruturada refere-se à técnica de projeção de padrões de luz cuidadosamente elaborados numa cena e à técnica de visualização da cena a partir de uma direção diferente. Usualmente o padrão consiste em sucessivos planos de luz em diferentes posições e orientações. Estes pixels que imagem uma superfície que é varrida por um padrão conhecido de luz tem informações suficientes para determinar as coordenadas 3D da superfície. Isto porque o padrão de luz é elaborado de tal maneira que a linha de visão passando pelo pixel e pela lente intercepte o padrão conhecido de luz em um unico ponto. Para propósitos de matching estereo, o padrão estruturado de luz pode ser "desestruturado" no sentido de ser um padrão de textura ou consistindo de pontilhados randômicos.

197 - LIGHT STRIPING (LIGHT STRIPING): O Light striping refere-se à simples forma de iluminação estruturada na qual o modelo de luz consiste em sucessivos planos de luz que são todos paralelos.

198 - STEREOPSIS (STEREOPSIS): Stereopsis refere-se à capacidade de determinar a profundidade de um ponto 3D pela observação do ponto em duas imagens projetadas em perspectiva capturadas de diferentes posições.

199 - UM PAR DE IMAGENS ESTÉREAS : Um par de imagens estéreas refere-se a duas imagens de projeção de perspectiva obtidas da mesma cena a partir de posições ligeiramente diferentes. A área comum que aparece em ambas as imagens do par estéreo é usualmente de 40% a 80% do total da área da imagem.

200 - UM PAR DE PONTOS CORRESPONDENTES : Um ponto p numa primeira imagem e um ponto q em uma segunda imagem são ditos formar um

par de pontos correspondentes (p, q) , se p e q são cada um a projeção de sensor diferente do mesmo ponto $3D$. O problema de correspondência visual consiste na correspondência de todos os pares de pontos correspondentes das duas imagens da mesma cena.

201 - DISPARIDADE OU DISPARIDADE ESTÉREA (**DISPARITY** or **STEREO DISPARITY**): A disparidade ou disparidade estéreo refere-se á diferença em posições das imagens do mesmo ponto $3D$ em duas imagens de projeção de perspectiva tiradas de posições diferentes.

202 - EMPARELHAMENTO ESTÉREO (**STEREO MATCHING**): O emparelhamento estéreo refere-se ao processo de correspondência ou identificação dos pontos em um par de imagens estéreas.

203 - TRIANGULAÇÃO (**TRIANGULATION**): A triangulação refere-se ao processo de determinação das coordenadas (x, y, z) de um ponto $3D$ a partir da posição observada por duas projeções perspectivas do ponto. Os centros de perspectiva e os planos de projeção perspectiva são assumidos como conhecidos.

204 - EIXO EPIPOLAR (**EPIPOLAR AXIS**): O eixo epipolar de um par de imagens estéreas é a linha passando através do centro de perspectivalidades das imagens.

205 - EPIPÓLOS (**EPIPOLES**): Os dois epipólos de um par de imagem estéreas consistem em um ponto em cada um dos planos de projeção perspectiva determinados como a interseção dos planos da imagem com o eixo epipolar. Para pares de imagem estéreas que possuam planos de imagem de projeção perspectiva, os epipólos estão infinitamente posicionados à direita e à esquerda.

206 - RAIO EPIPOLAR (**EPIPOLAR RAY**): Um raio epipolar é o segmento de linha entre um epipólo e um ponto no plano da imagem de projeção perspectiva.

207 - LINHA EPIPOLAR (**EPIPOLAR LINE**): Uma linha epipolar em uma imagem estérea correspondendo a um ponto dado em outra imagem estérea é a projeção perspectiva na primeira imagem estérea do raio $3D$ que é a projeção perspectiva inversa de um ponto dado de uma outra imagem estérea.

208 - PLANO EPIPOLAR (**EPIPOLAR PLANE**): Um plano epipolar relativo a um par de imagens estéreas é qualquer plano determinado por um ponto $3D$ observado, a posição em espaço $3D$ da sua projeção perspectiva na imagem estérea esquerda e a posição em espaço $3D$ de sua projeção perspectiva na imagem estérea direita. Cada plano epipolar contém o eixo epipolar e cada plano que contém o eixo epipolar é um plano epipolar.

209 - BORDA ESCONDIDA (**OCCCLUDING BOUNDARY**): Uma fronteira oclusa é uma fronteira aparecendo em uma imagem devido a descontinuidade nos limites de variação ou na profundidade de um objeto na cena observada.

210 - MAPA DE REFLETÂNCIA (**REFLECTANCE MAP**): Para cada posição de visualização fixada e para cada fonte de iluminação pontual, o mapa de refletância é uma função definida pelo espaço gradiente que especifica uma superfície de reflexibilidade. Desta maneira, a cada possível orientação da superfície normal (definida pelas derivadas parciais de primeira ordem da superfície) o mapa de reflexão da superfície especifica o coeficiente de reflexão da superfície.

211 - FORMA A PARTIR DO SOMBREAMENTO (SHAPE FROM SHADING): Forma de sombreamento refere-se à capacidade de determinação das características de forma 3D de um objeto a partir do sombreamento em tons de cinza manifestado pela superfície do objeto em uma projeção perspectiva ou em uma imagem de projeção ortográfica, ou seja, a partir do seu mapa de refletância.

212 - ANÁLISE DO SOMBREAMENTO LOCAL (LOCAL SHADING ANALYSIS): A análise de sombreamento local refere-se à capacidade de inferência da forma e da inclinação predominante de uma seção de uma superfície de um objeto pelas intensidades da imagem em uma vizinhança local de uma vista de projeção perspectiva ou paralela.

213 - ESTÉREO FOTOMÉTRICO (PHOTOMETRIC ESTEREO): O estéreo fotométrico refere-se à capacidade de determinação da orientação da superfície por meio de variações de sombreamento presentes em duas ou mais imagens tiradas da mesma cena à partir de uma mesma posição e orientação mas, com a fonte de luz em posições diferentes.

214 - CAMPO DE MOVIMENTO ou FLUXO DE IMAGEM (MOTION FIELD or IMAGE FLOW): O campo de movimento ou fluxo de imagem é uma imagem na qual o valor de cada pixel é a velocidade translacional projetada surgindo em um ponto da superfície de um objeto em movimento relativo em relação à câmera. Cada vetor de velocidade translacional é chamado de vetor de fluxo ótico.

215 - FLUXO ÓTICO (OPTICAL FLOW): Um fluxo ótico ou imagem de fluxo ótico é uma imagem na qual o valor de cada pixel é a projeção estimada da velocidade translacional surgindo de um ponto da superfície de um objeto em movimento relativo à câmera. Porque a velocidade projetada pode não ser estimável a cada pixel, pode haver alguns pixels em uma imagem de fluxo ótico que não possuam informações de fluxo ótico.

216 - FOCO DE EXPANSÃO (FOCUS OF EXPANSION): O foco de expansão de um campo de movimento de imagem surgindo de uma câmera em movimento e uma cena estacionária é o ponto na imagem no qual o fluxo ótico é zero e tal que o fluxo ótico nos pontos de vizinhança estão direcionados afastando-se deste ponto. Em casos de movimento relativo em direção à câmera, haverá exatamente um ponto de foco de expansão em tal imagem de campo de movimento. Em uma imagem de campo de movimento surgindo de um objeto em movimento relativo em relação à câmera, o foco de expansão é aquele ponto na imagem que possui todos os vetores de fluxo ótico surgindo do objeto em movimento e afastando-se dele.

217 - FOCO DE CONTRAÇÃO (FOCUS OF CONTRACTION): O foco de contração de uma imagem de campo de movimento surgindo de uma câmera móvel e uma cena estacionária é aquele ponto na imagem para o qual o fluxo ótico é zero e tal que o fluxo ótico dos pontos da vizinhança são direcionados em sua direção. Em casos de movimento relativo afastando-se da câmera, haverá exatamente um ponto de foco de contração em tal imagem de campo de movimento. Em uma imagem de campo de movimento surgindo de um objeto em movimento relativo em relação à câmera, o foco de contração é aquele ponto na imagem que possui todos os vetores de fluxo ótico surgindo do objeto móvel em sua direção.

218 - ESTRUTURA A PARTIR DO MOVIMENTO (STRUCTURE FROM MOTION): A estrutura a partir do movimento refere-se à capacidade de determina-

ção das características de forma de um objeto móvel bem como sua posição e velocidade à partir de uma seqüência de duas ou mais imagens tiradas dos objetos móveis. Equivalentemente, se uma câmera está em movimento, a estrutura do movimento refere-se à capacidade de determinação das características de forma de um objeto e sua posição e a velocidade da câmera a partir de uma seqüência de duas ou mais imagens tiradas do objeto pela câmera móvel. Um tipo fundamental de estrutura do problema de movimento é determinar a posição fixa de M pontos $3D$ a partir de uma seqüência de tempo de N vistas e contendo a perspectiva de projeção $2D$ dos M $3D$ pontos.

219 - NAVEGAÇÃO PASSIVA (**PASSIVE NAVIGATION**): A navegação passiva refere-se à determinação do movimento da câmera a partir de uma seqüência de imagens variantes no tempo.

220 - ANÁLISE DINÂMICA DE CENAS (**DYNAMIC SCENE ANALYSIS**): A análise dinâmica de cena refere-se à análise de imagens variantes no tempo. O propósito da análise pode ser obter o caminho de objetos móveis, determinar o movimento dos objetos, reconhece-los, determinar sua posição espacial em cada momento em que a imagem foi obtida, ou determinar a descrição da forma ou caracterização de um ou mais objetos.

221 - FORMA A PARTIR DA TEXTURA (**SHAPE FROM TEXTURE**): A forma a partir da textura refere-se à capacidade de determinar as características de forma $3D$ de uma superfície de textura homogênea a partir das variações de densidade da textura manifestadas pela superfície em uma imagem da projeção perspectiva ou ortográfica.

222 - RECONSTRUÇÃO DE SUPERFÍCIE (**SURFACE RECONSTRUCTION**): A reconstrução de superfície refere-se ao processo pelo qual uma superfície $3D$ é analiticamente descrita pela representação de suas fronteiras baseado no processamento de um par de imagens estéreas, um mapa de limites, ou uma seqüência de imagens variantes no tempo da superfície observada.

223 - FORMA A PARTIR DO CONTORNO (**SHAPE FROM CONTOUR**): A forma a partir do contorno refere-se à capacidade de inferimento da forma $3D$ de um objeto a partir da vista de projeção perspectiva $2D$ de um conjunto de contornos regularmente marcados na superfície do objeto.

224 - $2^{1/2}D$ SKETCH : O esboço $2^{1/2}D$ é uma imagem multibanda na qual cada pixel provém informações sobre a profundidade e A orientação da superfície projetada nesta imagem, bem como provendo uma indicação da existência de uma descontinuidade profunda próxima e, superficial próxima.

225 - CARACTERÍSTICAS INTRÍNSECAS DE UMA CENA (**INTRINSIC SCENE CHARACTERISTICS**): As características intrínsecas de uma cena para cada ponto da superfície do objeto são: sua profundidade à partir do plano focal da imagem, sua orientação superficial, sua refletância e, sua iluminação incidental.

226 - IMAGEM INTRÍNSECA (**INTRINSIC IMAGE**): Uma imagem intrínseca é uma imagem multibanda na qual cada pixel contém as características de cena intrinsecamente predominantes da projeção do padrão da superfície para sua posição. Assim, cada pixel da imagem intrínseca especifica profundidade, orientação da superfície, refletância e, iluminação incidental para a projeção do padrão de superfície

para sua posição.

227 - ÓTICA INVERSA (INVERSE OPTICS): A ótica inversa refere-se à capacidade de inferência da posição *3D* e/ou da superfície normal de cada ponto na superfície de um objeto e/ou da forma da superfície de uma ou mais imagens de projeção perspectiva do objeto. Incluídas nas técnicas de ótica inversa estão estéreo, estéreo fotométrico, forma do sombreamento, forma da textura e estrutura do movimento. Técnicas de ótica inversa podem ser imaginadas como técnicas que invertem os processos de projeção perspectiva e, que, entretanto pertencem à escola reconstrucionista de visão computacional de física aplicada.

228 - MUNDO DE BLOCOS (BLOCKS WORLD): O mundo de blocos é referente a um mundo no qual todos os objetos possuem superfícies simples. Os tipos mais comuns de objetos no mundo de blocos são os objetos poliédricos.

229 - ASPECTO DE VISÃO (VIEW ASPECT): Um aspecto de visão é uma região conectada máxima de um espaço de ponto de vista tendo a propriedade de, quando observando de um dado centro de um objeto de qualquer ponto de um aspecto de vista, as vistas resultantes são topologicamente idênticas.

230 - GRAFO DE ASPECTO (ASPECT GRAPH): Um grafo de aspecto é um grafo no qual os nós são os aspectos de vista e os arcos conectam aspectos de vista adjacentes. As vistas representadas pelos nós no grafo são todas as vistas estáveis, características e principais.

231 - VISÃO 3D (3D VISION): Visão *3D* refere-se a capacidade de um sistema de visão por máquina ser capaz de inferir algumas características *3D* ou atributos de um objeto tais como sua posição, dimensão, orientação ou movimento ou algumas características *3D* de um ponto ou uma superfície do objeto tal como sua posição *3D* ou orientação normal de superfície.

232 - INSPEÇÃO VISUAL AUTOMÁTICA (AUTOMATIC VISUAL INSPECTION): Inspeção visual automática ou inspeção de visão automática refere-se a um processo de inspeção que usa um sensor produzindo dados de imagem e técnicas de processamento de imagens, reconhecimento de padrões, ou visão computacional para medir e/ou interpretar as imagens dos objetos com o intuito de determinar se eles foram fabricados dentro das tolerâncias permitidas. Sistemas de inspeção visual automática usualmente integram as tecnologias de manuseio de materiais, iluminação, aquisição de imagem e hardware computacional de propósito especial juntamente com algoritmos apropriados de análise de imagens em um sistema desenvolvido para ser de uso prático em indústrias. Entre os benefícios da utilização de inspeção visual automatizada pode-se incluir maior precisão, confiabilidade, repetibilidade e, uma completa garantia de qualidade com um custo menor e a uma velocidade maior de inspeção do que poderia ser possível pelo trabalho manual.

233 - SISTEMA DE VISÃO POR MÁQUINA (MACHINE VISION SYSTEM): Um sistema de visão por máquina é um sistema capaz de adquirir uma ou mais imagens de um objeto, processando, analisando e medindo várias características das imagens adquiridas, e interpretando os resultados das medições de tal modo que algumas decisões úteis possam ser tomadas sobre o objeto. Funções de sistemas de visão por máquina incluem localização, inspeção, calibragem, identificação, reconhecimento, contagem e estimação de movimento.

234 - **FIXAÇÃO VISUAL** ou **DETERMINAÇÃO VISUAL DA POSIÇÃO (VISUAL FIXTURING** or **VISUAL POSE DETERMINATION)**: A fixação visual ou determinação de pose visual refere-se a capacidade de inferir a posição e a orientação de um objeto conhecido usando um modelo de objeto adequado e uma ou mais câmeras, sensores de raio de ação, ou triangulação baseados em sensores de visão.

235 - **CALIBRAGEM VISUAL (VISUAL GAUGING)**: A calibragem ótica ou calibragem visual refere-se a habilidade de medir posições específicas ou dimensões de um objeto manufaturado pelo uso de sensores sensíveis de luz sem contato para comparar estas medidas com limites de tolerância pré-ajustados para inspeção de qualidade e decisões de ordenação. Calibragem tem largas aplicações em fabricação desde que pode determinar os diâmetros de orifícios, aberturas, ou chanfros, a largura de eixos, componentes, aberturas, cabos, ou bastões, e a localização relativa de orifícios, dobras, características, componentes, aberturas ou quebras.

236 - **SISTEMA DE INSPEÇÃO FIXA** ou **FLEXÍVEL (FIXED INSPECTION SYSTEM)**: Um sistema de calibragem pode ser tanto uma inspeção fixa ou um sistema de inspeção flexível. O sistema de inspeção fixa engloba a parte em um teste de fixação de precisão e tem um ou mais sensores para tomar as medidas requeridas. Inspeção flexível utiliza sensores que são movidos sobre a parte que está sendo inspecionada, sendo que o movimento é feito sobre uma trajetória de caminho programado. Calibragem de inspeção flexível é algumas vezes chamada de calibragem robótica.

237 - **PROCEDIMENTO ROBUSTO DE VISÃO** : Um procedimento de visão é dito ser robusto se pequenas mudanças no modelo assumido no qual o procedimento ou técnica foi desenvolvido para produzir somente pequenas mudanças no resultado. Frações pequenas de dados que não se encaixam no modelo assumido e que de fato estão muito longe de se encaixar no modelo assumido, constituem uma pequena mudança no modelo assumido. Dados não encaixam em um modelo assumido por erros de arredondamento ou quantificação, erros grosseiros, ou porque o modelo por si mesmo é somente uma aproximação idealizada da realidade.

238 - **VISÃO COMPUTACIONAL, COMPREENSÃO DE IMAGEM** ou **ANÁLISE DE CENA (COMPUTER VISION, IMAGE UNDERSTANDING, SCENE ANALYSIS)**: A visão computacional, compreensão de imagem, ou análise de cena é aquela combinação de processamento de imagem, reconhecimento de padrões e tecnologias de inteligência artificial que se focam na análise computacional de uma ou mais imagens, tiradas com um sensor simples/multibanda ou tiradas em uma sequência de tempo. A análise reconhece, localiza a posição e orientação e provê uma descrição simbólica suficientemente detalhada ou o reconhecimento daquelas imagens de objetos julgados serem de interesse no ambiente tridimensional. Os processos de visão computacional freqüentemente usam modelamento geométrico e complexas representações do conhecimento em uma expectativa ou modelo baseado em semelhanças ou metodologia de busca. A busca pode incluir bottom up, top down, blackboard, hierárquica e estratégias de controle heterarquical.

239 - **ESTRATÉGIA DE CONTROLE BOTTOM UP (BOTTOM UP CONTROL STRATEGY)**: A estratégia de controle Bottom up é uma abordagem para resolver problemas que é dirigido pelos dados. Ela não emprega modelos de objetos em seus estágios iniciais e somente usa conhecimentos gerais sobre o mundo que está sendo analisado. Em um sistema de visão computacional usando um controle de estratégia bottom up, os dados da imagem observada são interpretadas e agregadas.

As interpretações e uniões são então sucessivamente manipuladas e agregadas até que o nível de descrição suficientemente alto para a cena seja gerada.

240 - ESTRATÉGIA DE CONTROLE TOP DOWN (TOP DOWN CONTROL STRATEGY): A estratégia de controle top down é uma abordagem para resolver problemas que é a meta-dirigida ou expectativa dirigida. A forma de solução é gerada ou na forma de hipóteses. Assumindo que a hipótese é verdadeira e usando a informação de uma base de dados de conhecimento, o mecanismo de inferência então infere, se possível, alguns conjuntos consistentes de valores para as variáveis ou parâmetros desconhecidos. Se um conjunto consistente não for inferido, então uma nova forma de solução é gerada ou na forma de hipóteses. Em um sistema de visão computacional usando uma estrutura de controle top down, o número ou tipos de objetos que começam a ser analisados em uma imagem usualmente é altamente restrigido e o conhecimento sobre os objetos, relações entre objetos, e partes dos objetos são totalmente conhecidos. O sistema gera hipóteses que a imagem mostra um conjunto particular de objetos, infere valores para os parâmetros, e então testa para verificar que a hipótese é consistente com os dados observados.

241 - ESTRATÉGIA DE CONTROLE HIERÁRQUICA (HIERARQUICAL CONTROL STRATEGY): Uma estratégia de controle hierárquico é uma abordagem para resolver problemas em que dado o problema é resolvido dividindo ele em um conjunto de sub-problemas, cada qual encapsula um aspecto importante ou maior do problema original. Então cada sub-problema é sucessivamente dividido em maiores sub-problemas detalhados. O refinamento continua até os maiores sub-problemas refinados possam ser resolvidos diretamente.

242 - ESTRATÉGIA DE CONTROLE DE QUADRO NEGRO (BLACKBOARD CONTROL STRATEGY): Uma estratégia de controle de quadro-negro é uma abordagem para resolver problemas em que vários componentes de um mecanismo de inferência comunique com um outro por um trabalho comum em uma área comum de armazenamento de dados chamada de quadro-negro. Quando um quadro-negro tem dados suficientes para permitir um componente de um mecanismo de conclusão fazer uma dedução, o mecanismo de conclusão vai trabalhar e escreve seus resultados no quadro-negro onde ele se torna disponível para outros componentes do mecanismo de conclusão. Desta maneira as restrições de inferência são sucessivamente propagadas e procura requerida é feita de forma mais limitada.

243 - VISÃO POR COMPUTADOR BASEADA EM MODELOS (MODEL BASED COMPUTER VISION): A visão computacional baseada em modelos é um processo de visão computacional o qual emprega um modelo explícito do objeto para ser reconhecido. Reconhecimento procede em uma maneira top down pelo casamento da estrutura de dados do objeto inferida com a imagem observada para o modelo de estrutura de dados.

244 - VISÃO BASEADA EM CONHECIMENTO (KNOWLEDGE BASED VISION): A visão computacional baseada em conhecimento refere-se a um processo de visão computacional que tem um componente de uma imagem processada, um argumento ou o componente de inferência, e uma base de dados de conhecimento do componente. A base de dados de conhecimento armazena informações sobre o ambiente inicial da imagem. O componente da imagem processada extrai as primitivas ponto, linha, curva, e informação das regiões da imagem observada. O argumento ou o componente de inferência é tipicamente uma regra baseada e integra a informação produzida pelo componente da imagem processada com a informação na base de dados de conhecimento e razões sobre qual hipótese vai ser gerada, qual

hipótese vai ser validada, qual nova informação vai ser inferida e a partir do que já foi estabelecido, e qual novas primitivas do componente da imagem processada deveriam ser extraídas na próxima vez.

17. RECONHECIMENTO DE PADRÕES

245 - RECONHECIMENTO DE PADRÕES (PATTERN RECOGNITION): Técnicas de reconhecimento de padrões podem ser usadas para construir regras de decisão que habilita unidades a ser identificadas com base em seus padrões de medidas. Técnicas de reconhecimento de padrões podem também ser empregadas para agrupar unidades que tenham padrões de medidas bastante similares. No *reconhecimento de padrões estatístico*, os padrões de medidas tem a forma de n-tuplas ou vetores. No *reconhecimento de padrões sintático*, os padrões de medida tem a forma de sentenças de uma linguagem em frase de uma gramática de estrutura. No *reconhecimento de padrões estrutural*, as medidas não tem a forma de uma n-tupla ou vetor. Mas apropriadamente, a unidade a ser medida é codificada em termos de suas partes e suas relações assim como das propriedades das partes.

246 - RECONHECIMENTO DE PADRÕES PICTORAIS (PICTORIAL PATTERN RECOGNITION): O Reconhecimento de padrões pictoriais refere-se as técnicas o qual tratam a imagem como um padrão ou categorizam a imagem ou produzem a descrição de uma imagem.

247 - UNIDADE (UNIT): A unidade é uma entidade que é observada e cujas propriedades medidas constituem o padrão de medida. A mais simples e prática unidade observada e medida nos dados de uma imagem no reconhecimento de padrões é freqüentemente o pixel (a intensidade do tom de cinza ou a intensidade n-tupla do tom de cinza em uma particular célula de resolução). Isto é que faz o reconhecimento de padrões pictorial tão difícil, porque os objetos que requerem análise ou identificação não são pixels únicos mas são freqüentemente formações espaciais complexas de pixels.

248 - PADRÃO DE MEDIÇÃO (MEASUREMENT PATTERN): Um padrão de medição ou padrão é a estrutura de dados de medidas resultantes da observação da unidade.

249 - VETOR N-TUPLA DE MEDIÇÃO (MEASUREMENT N-TUPLE VECTOR): Um vetor n-tupla de medição é a n-tupla ordenada de medidas obtidas de uma unidade sobre observação. Cada componente da n-tupla é uma medição de uma qualidade particular, propriedade, característica, ou característica da unidade. No reconhecimento de padrões de imagem, as unidades são usualmente pixels ou simples formações de pixels e as n-tuplas de medições são as intensidades de tons de cinza correspondentes, intensidade n-tupla dos tons de cinza, ou propriedades das formações das intensidades dos tons de cinza.

250 - PRODUTO CARTESIANO (CARTESIAN PRODUCT): O produto cartesiano de dois conjuntos A e B , denotado por $A \times B$, é o conjunto de todos os pares ordenados onde o primeiro componente do par é alguns elementos do primeiro conjunto e o segundo componente do par é alguns elementos do segundo conjunto. O produto cartesiano de N conjuntos pode ser definidos indutivamente.

251 - ESPAÇO DE MEDIÇÃO (MEASUREMENT SPACE): Um espaço de medição é um conjunto suficiente para incluir nele o conjunto de todos padrões

de medidas possíveis que pode ser obtidas pela observação de alguns conjuntos de unidades.

252 - CONJUNTO DE INTERVALOS (RANGE SET): O conjunto de intervalos R_i de um i -ésimo sensor, que produz a i -ésima imagem em um conjunto de multi-imagens, é o conjunto de todas as medidas que podem ser produzidas pelo i -ésimo sensor. Simplificando, ele é o conjunto de todas as intensidades de tons de cinza que podem possivelmente existir na i -ésima imagem. Quando as unidades são os pixels, o espaço de medição de M é o produto cartesiano do conjunto de distâncias dos sensores: $M = R_1 \times R_2 \times \dots \times R_n$.

253 - CLASSE DE PADRÕES (PATTERN CLASSES): Cada unidade é assumido ser de um e só um dado tipo. O conjunto de tipos é chamado de conjunto de classes de padrões ou categorias C , cada tipo começa uma categoria particular.

254 - CARACTERÍSTICA ou PADRÃO DE CARACTERÍSTICA (FEATURE or FEATURE PATTERN): Uma característica, ou um padrão de característica, ou uma n -tupla de característica, ou vetor de característica ou característica padrão é uma n -tupla ou vetor cujos componentes são funções das variáveis iniciais do padrão de medida ou alguns subconjuntos das variáveis iniciais do padrão de medida. Característica n -tuplas ou vetores são concebidas para conter uma alta quantia de informação relativa da discriminação entre unidades de tipos de categorias em um dado conjunto de categoria. Às vezes as características são pré-determinadas e outras vezes elas são determinadas no tempo em que o problema de discriminação de padrões está sendo resolvido. No reconhecimento de padrões de imagens, as características freqüentemente contém informações relativas à intensidade de tons de cinza, textura, ou forma da região.

255 - ESPAÇO DE CARACTERÍSTICAS (FEATURE SPACE): O espaço de características é o conjunto de todas as n -tuplas de características possíveis.

256 - SELEÇÃO DE CARACTERÍSTICAS (FEATURE SELECTION): A seleção de características é o processo pelo qual as características para ser usadas em um problema de reconhecimento de padrões são determinadas. Às vezes a seleção de características é chamada seleção de propriedades.

257 - EXTRAÇÃO DE CARACTERÍSTICAS (FEATURE EXTRACTION): A extração de características é o processo pelo qual um padrão de medida inicial ou um subconjunto de padrões de medida é transformado para uma nova característica padrão. Às vezes a extração de características é chamada extração de propriedades.

A palavra "padrão" pode ser usada em três sentidos distintos:

- como padrão de medida;
- como padrão de característica; e
- como um padrão de dependência ou padrões de relações entre os componentes de qualquer medida n -tupla ou característica n -tupla derivada das unidades de uma categoria particular e que são únicas para aquelas n -tuplas, isto é, elas são dependências que não ocorrem em qualquer outra categoria.

258 - CLASSIFICADOR (CLASSIFIER): Um classificador é um dispositivo ou processo que classifica padrões em categorias ou classes.

259 - HIPÓTESE DE COMPACIDADE (COMPACTNESS HYPOTHESIS): Um hipótese de "compacidade" determina que as medidas de padrões de uma dada

classe são mais próximas de outras medidas de padrões na classe que elas são de medidas de padrões de outras classes.

260 - REGIÃO DE CLASSE (**CLASS REGION**): A região de um espaço ocupada por medidas de padrões vindo da mesma classe ou categoria é chamada região de classe.

261 - SEPARABILIDADE (**SEPARABILITY**): Duas classes ou categorias são ditas serem separáveis se suas regiões de classes não se sobrepõem. Se para toda região de classe lá existe um hiperplano que separa a região de classe de todas as outras regiões de classes, as classes são ditas como serem linearmente separáveis.

262 - PADRÃO DE REFERÊNCIA (**PROTOTYPE or REFERENCE PATTERN**): Um padrão de protótipo ou um padrão de referência representa as medidas observáveis ou características ou padrões de características derivadas das unidades de uma categoria particular. Uma categoria é dita ter um padrão de protótipo se e somente se o padrão de característica é altamente representativo das n -tuplas obtidas das unidades daquela categoria.

263 - SEQUÊNCIA DE DADOS (**DATA SEQUENCE**): Uma sequência de dados $S_d = (d_1, d_2, \dots, d_j)$ é uma sequência de padrões derivada dos padrões ou características de medida de algumas sequências de unidades observadas. d_1 é um padrão associado com a primeira unidade; d_2 é o padrão associado com a segunda unidade; e d_j é o padrão associado com a j -ésima unidade.

264 - REGRA DE DECISÃO (**DECISION RULE**): Uma regra de decisão f usualmente associa uma e só uma categoria a cada unidade com base na sequência de padrões de medida em uma sequência de dados S_d ou com base na sequência correspondente dos padrões de característica.

265 - REGRA DE DECISÃO SIMPLES (**SIMPLE DECISION RULE**): Uma regra de decisão simples é uma regra de decisão que associa uma categoria a uma unidade somente com base nas medidas ou características associadas com a unidade. Consequentemente, as unidades são tratadas independentemente e a regra de decisão f pode ser conceituada como uma função que nomeia uma e somente uma categoria para cada padrão no espaço de medida ou para cada característica no espaço de característica.

266 - REGRA DE DECISÃO HIERÁRQUICA (**HIERARCHICAL DECISION RULE**): Uma regra de decisão hierárquica é uma regra de decisão em forma de árvore. Em árvores binárias, cada nó não-terminal de uma árvore contém uma regra de decisão simples que classifica padrões como pertencendo aos seus filhos esquerdos ou aos seus filhos direitos. Cada nó terminal de uma árvore contém classes nomeadas ou categorias da unidade observada.

267 - REGRA DE DECISÃO COMPOSTA (**COMPOUND DECISION RULE**): Uma regra de decisão composta é uma regra de decisão que associa uma unidade a uma categoria com base em algumas sub-sequências não triviais de padrões de medidas em uma sequência de dados ou numa sequência correspondente ou padrões de características.

268 - JULGAMENTO DE RESERVA (**RESERVE JUDGEMENT or DEFER ASSIGNMENT**): Uma previsão pode ser feita para uma regra de decisão para o julgamento de reserva ou para a tarefa de adiamento se o padrão é muito

fechado para a fronteira da categoria em media ou espaço de característica. Com esta provisão, a tarefa de adiamento é uma tarefa para a categoria para o "julgamento de reserva".

269 - SEQUÊNCIA DE IDENTIFICAÇÃO DE CATEGORIAS ou GROUND TRUTH (CATEGORY IDENTIFICATION SEQUENCE or GROUND TRUTH): Uma sequência de identificação de categoria ou GROUND TRUTH $S_c = (c_1, c_2, \dots, c_j)$ é uma sequência de identificações de categorias obtidas a partir de algumas sequências de unidades observadas. c_1 é uma identificação de categoria da primeira unidade; c_2 é uma identificação de categoria da segunda unidade; c_j é uma identificação de categoria da j-ésima unidade.

270 - SEQUÊNCIA DE TREINAMENTO (TRAINING SEQUENCE): Uma sequência de treinamento é um conjunto de duas sequências: (1) a sequência de dados e (2) a sequência de identificação da categoria correspondente. Um dado de treinamento é o par consistindo de um padrão na sequência de dados e a identificação da categoria correspondente na sequência de identificação de categoria. A sequência de treinamento é usada para estimar a categoria condicional de distribuições de probabilidade dos quais a regra de decisão é construída ou ela possa ser usada para estimar a própria regra de decisão.

271 - PROCEDIMENTO DE TREINAMENTO (TRAINING PROCEDURE): Um procedimento de treinamento é um procedimento que usa a sequência de treinamento para construir uma regra de decisão. Ele pode operar passando por toda a sequência de treinamento uma vez e construir a regra de decisão de uma maneira que é independente da ordem em que os dados de treinamento ocorrem na sequência de treinamento. Ele pode operar iterativamente no caso em que ele passa pela sequência de treinamento inteira muitas vezes e depois de manipular cada dado de treinamento, ele modifica ou atualiza a regra de decisão. Cada procedimento iterativo de treinamento pode ser afetado pela ordem em que os dados de treinamento ocorrem na sequência de treinamento.

18. RECONHECIMENTO DE PADRÕES

272 - PROCEDIMENTO DE TREINAMENTO POR JANELA (WINDOW TRAINING PROCEDURE): Um procedimento de treinamento por janela é um procedimento de treinamento iterativo no qual cada ajuste da regra de decisão é feito apenas quando o dado de treinamento se encaixa em uma determinada janela, um subconjunto do espaço de dados. Normalmente este subconjunto, ou janela, contém a fronteira de decisão.

273 - PROCEDIMENTO DE TREINAMENTO CORRIGINDO ERROS (ERROR CORRECTING TRAINING PROCEDURE): Um procedimento de treinamento por correção de erros é um procedimento de treinamento iterativo e sequencial no qual em cada iteração a regra de decisão é ajustada em resposta a uma má classificação de um dado de treinamento.

274 - Um classificador aprende se seu processo iterativo de treinamento aumenta a precisão de classificação do classificador após algumas iterações.

275 - PROBABILIDADE CONDICIONAL (CONDITIONAL PROBABILITY): A probabilidade condicional de uma medida ou característica, representada por uma n-tupla, d dada a categoria c , é normalmente denotada por $P_c(d)$, ou por $P(d/c)$, e é definida como a frequência relativa ou proporção de vezes que a n-tupla d é derivada de uma unidade cuja verdadeira categoria é c .

276 - DECISÃO NÃO PARAMÉTRICA (NON-PARAMETRIC DECISION):

Uma regra de decisão não paramétrica, ou livre de distribuição, é aquela que não faz suposições a respeito da forma da distribuição de probabilidades dos padrões dadas as categorias.

277 - REGRA DE DECISÃO DE SEMELHANÇA MÁXIMA (MAXIMUM LIKELIHOOD DECISION RULE):

Uma regra de decisão por semelhança é aquela que trata as unidades independentemente e atribui uma unidade u , que possui as medidas ou características d à categoria c cujas unidades são as mais prováveis de terem dado origem ao padrão ou vetor de características d , ou seja, a qual a probabilidade condicional de d dado c é a maior.

278 - REGRA DE DECISÃO BAYESIANA (BAYES DECISION RULE):

Uma regra de decisão Bayesiana simples é aquela que trata as unidades independentemente e atribui uma unidade u que possui as medidas de padrão ou características d à categoria c cuja probabilidade condicional, dadas as medidas d , é a maior.

279 - REGRA DE DECISÃO DO VIZINHO MAIS PRÓXIMO (NEAREST NEIGHBOR DECISION RULE):

Seja $\langle u_1, u_2, \dots, u_j \rangle$ uma seqüência de unidades com a correspondente seqüência de dados $\langle d_1, d_2, \dots, d_j \rangle$ e uma seqüência de classificação conhecida $\langle c_1, c_2, \dots, c_j \rangle$. Uma regra de decisão do vizinho mais próximo é tal que trata as unidades independentemente e atribui uma unidade u desconhecida com medidas ou características d à categoria c_j onde d_j é o padrão mais próximo de d , por uma determinada função de distância.

280 - FUNÇÃO DISCRIMINANTE (DISCRIMINANT FUNCTION):

Uma função discriminante $f_i(d)$ é uma função escalar, cujo domínio é normalmente o espaço das medidas e cuja amplitude normalmente são os números reais. Quando $f_i(d) \geq f_k(d)$, para $k = 1, 2, \dots, K$, então a regra de decisão atribui a i -ésima categoria à unidade que originou o padrão d .

281 - FUNÇÃO DISCRIMINANTE LINEAR (LINEAR DISCRIMINANT FUNCTION):

Uma função discriminante linear f é uma função da forma $f(d) = \sum_{j=1}^n a_j \delta_j + a_0$ onde $d = (\delta_1, \delta_2, \dots, \delta_n)$ representa o padrão de medida.

282 - FUNÇÃO DISCRIMINANTE QUADRÁTICA (QUADRATIC DISCRIMINANT FUNCTION):

Uma função discriminante quadrática f é uma função da forma $f(d) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \delta_i \delta_j + \sum_{j=1}^n a_j \delta_j + a_0$ onde $d = (\delta_1, \delta_2, \dots, \delta_n)$ representa o padrão de medida.

283 - FRONTEIRA DE DECISÃO (DECISION BOUNDARY): Uma fronteira de decisão entre a i -ésima e a k -ésima categorias é um subconjunto H de padrões no espaço de medidas M , definido por:

$$H = \{D \in M \mid f_i(d) = f_k(d)\} \quad (5.42)$$

onde f_i e f_k são funções discriminantes para a i -ésima e a k -ésima categorias.

284 - FRONTEIRA DE DECISÃO HIPERPLANAR (HYPERPLANE DECISION BOUNDARY):

Uma fronteira de decisão hiperplanar é um nome especial dado às fronteiras de decisão formadas pelo uso de funções discriminante lineares.

285 - REGRA DE DECISÃO LINEAR (LINEAR DECISION RULE):

Uma regra de decisão linear é uma regra de decisão simples do reconhecimento de padrões estatístico, que normalmente trata as unidades independentemente e faz as atribuições

de categorias usando funções discriminantes lineares. As fronteiras de decisão obtidas com regras de decisão lineares são hiperplanos.

286 - PROBLEMA DA DISCRIMINAÇÃO DE PADRÕES (PATTERN DISCRIMINATION PROBLEM): O problema de discriminação de padrões diz respeito a como construir a regra de decisão que atribui uma unidade a uma determinada categoria com respeito ao(s) padrão(ões) de medida na sequência de dados ou com respeito ao(s) padrão(ões) de características na sequência de dados.

287 - IDENTIFICAÇÃO DE PADRÕES (PATTERN IDENTIFICATION): A identificação de padrões é o processo no qual uma regra de decisão é aplicada. Se $S_u = \langle u_1, u_2, \dots, u_j \rangle$ é a sequência de unidades a serem observadas e identificadas, e se $S_d = \langle d_1, d_2, \dots, d_j \rangle$ é a sequência de padrões correspondentes, então o processo de identificação produz uma sequência de identificação de categorias $S_c = \langle c_1, c_2, \dots, c_j \rangle$ onde c_i é a categoria em C à qual a regra de decisão atribui a unidade u_i com base nos j padrões em S_d . Algumas vezes a identificação de padrões também é chamada de classificação de padrões ou classificação.

288 - PERCEPTRON ou REDE NEURAL (PERCEPTRON or NEURAL NETWORK): Um perceptron ou rede neural é uma rede interconectada de elementos não lineares ou unidades processadoras capaz de aprender e se auto organizar. A resposta de um elemento ou unidade processadora é uma função monotônica não linear da soma ponderada das entradas para a unidade processadora. Os pesos, chamados de pesos sinápticos são modificados por um algoritmo de aprendizado ou reforço. Funções não lineares típicas são: $sgn(x)$, $(\frac{1}{1+e^{-x}})$ e $tanh(x)$. Quando cada unidade processadora contribui com um componente no vetor de saída, e suas entradas são selecionadas dos componentes dos vetor de entrada, o perceptron é chamado de perceptron simples. Unidades processadoras cujas saídas influenciam apenas indiretamente os componentes do vetor de saída são chamadas de unidades ocultas.

289 - REFORÇO DE CORREÇÃO DE ERROS (ERROR CORRECTIVE REINFORCEMENT): Um reforço de correção de erros para um perceptron é um algoritmo de treinamento no qual a mudança nos pesos sinápticos é uma função do grau que a saída de uma unidade processadora não é o que é esperado que seja. Algoritmos de reforço na correção do erro também são chamados de algoritmos de **(retropropagação do erro)**.

290 - PERCEPTRON PAREADO AVANTE (FORWARD COUPLED PERCEPTRON): Um perceptron "pareado avante" é um perceptron no qual as unidades processadoras estão organizadas em camadas. As entradas para as unidades processadoras na camada n vêm das saídas das unidades processadoras da camada anterior à camada n . Perceptrons com uma única camada podem criar superfícies de decisão lineares. Perceptrons com duas camadas podem criar regiões de decisão convexas. Perceptron com três camadas podem criar regiões de decisão de forma quase arbitrária.

291 - PERCEPTRON PAREADO EM SERIE (SERIES COUPLED PERCEPTRON): Um perceptron "pareado em série" é um perceptron "pareado avante" no qual as entradas para o centro de processamento na camada n vêm das saídas das unidades processadoras na camada $n - 1$.

292 - PERCEPTRON COM REALIMENTAÇÃO (BACK COUPLED PERCEPTRON): Um perceptron com realimentação é um perceptron no qual as

conexões não são avante. Ou seja, existe algum processador na camada n cuja saída é a entrada para um processador em uma camada anterior à camada n .

293 - GRUPO (**CLUSTER**): Um grupo (cluster) é um grupo homogêneo de unidades que são muito "parecidas" umas com as outras. "A semelhança" entre as unidades é determinada normalmente pela associação, similaridade ou distância entre os padrões de medida associados com as unidades.

294 - FUNÇÃO DE ATRIBUIÇÃO DE GRUPO (**CLUSTER ASSIGNMENT FUNCTION**): Uma função de atribuição de grupo é uma função que atribui cada unidade observada a um grupo de acordo com o(s) padrão(ões) de medida na seqüência de dados ou com base nas suas características correspondentes. Algumas vezes as unidades são tratadas independentemente. Neste caso a função de atribuição de grupo pode ser considerada como uma transformação do espaço de medidas para o conjunto de grupos (clusters).

295 - CLASSIFICAÇÃO DE PADRÕES (**PATTERN CLASSIFICATION**): O problema de classificação de padrões diz respeito a como construir a função de atribuição de grupos (clustering), a qual agrupa unidades similares. Classificação de padrões é sinônimo de taxonomia numérica ou agrupamento (clustering).

296 - IDENTIFICAÇÃO DE GRUPOS (**CLUSTER IDENTIFICATION**): O processo de identificação de grupo (cluster) é o processo no qual a função de atribuição de grupo é aplicada a uma seqüência de unidades levando a uma seqüência de identificação de grupo.

297 - ERRO DE IDENTIFICAÇÃO ou ERRO DE TIPO *I* (**MISIDENTIFICATION** or **TYPE I ERROR**): Um erro de identificação ou de detecção (má identificação ou má detecção), ou erro do tipo *I*, ocorre para a categoria c_i se uma unidade cuja verdadeira categoria é a categoria c_i , é atribuída pela regra de decisão a uma categoria c_k , com $k \neq i$. Um erro de identificação é também chamado de erro de omissão.

298 - IDENTIFICAÇÃO FALSA ou ERRO DE TIPO *II* (**FALSE IDENTIFICATION** or **TYPE II ERROR**): Uma identificação falsa, ou alarme falso, ou erro do tipo *II*, ocorre para uma categoria c_i se uma unidade cuja verdadeira categoria é a categoria c_k , com $k \neq i$, é atribuída pela regra de decisão para a categoria c_i . Um erro de falsa identificação também é chamado de erro de comissão.

299 - SEQÜÊNCIA DE PREDIÇÃO (**PREDICTION SEQUENCE**): Uma seqüência de predição, ou seqüência de teste, ou uma seqüência de generalização é um conjunto de duas seqüências: (1) seqüência de dados (cuja identificação da verdadeira categoria correspondente pode ser considerada desconhecida da regra de decisão) e (2) uma seqüência de identificação de categoria correspondente determinada pela atribuição da regra de decisão. Fazendo-se a comparação da seqüência de identificação gerada pela atribuição da regra de decisão com verdadeira seqüência de identificação, a taxa de má identificação e a taxa de falsa identificação para cada categoria podem ser estimadas.

300 - MATRIZ DE CONFUSÃO (**CONFUSION MATRIX**): Uma matriz de confusão ou tabela de contingência é um vetor de probabilidades cujas colunas e linhas são designadas similarmente nomes de categorias e que indicam a probabilidade de identificação correta para cada categoria, bem como as probabilidades dos erros de tipo *I* e *II*. Para o elemento (i, k) , P_{ik} é a probabilidade de uma unidade

da categoria c_i , seja atribuída pela regra de decisão para a categoria c_k .

301 - Uma unidade é dita detectada se a regra de decisão é capaz de atribuí-la como pertencente a apenas um dado subconjunto A de categorias do conjunto C de categorias. Detectar uma unidade não implica que a regra de decisão é capaz de identificar a unidade como pertencente especificamente a uma categoria em particular.

302 - Uma unidade é dita ser reconhecida, identificada, classificada, categorizada ou ordenada se a regra de decisão é capaz de atribuí-la a alguma categoria do conjunto de categorias. Em algumas aplicações, pode haver uma distinção definida entre reconhecer e identificar. Nestas aplicações, para que uma unidade seja reconhecida, a regra de decisão deve ser capaz de atribuí-la a um tipo de categoria, sendo o tipo formado por muitas subcategorias. Para que uma unidade seja identificada, a regra de decisão deve atribuí-la não apenas a um tipo de categoria, mas também a uma subcategoria do tipo de categoria. Por exemplo, uma pequena área de terra que pode ser reconhecida como contendo árvores pode ser especificamente identificada como contendo macieiras.

303 - Uma unidade é dita ser localizada se coordenadas específicas podem ser dadas para a sua localização física.

304 - EXACTIDÃO (**ACCURACY**): A exatidão se refere ao grau de proximidade que uma estimativa possui com o real valor que ela está estimando.

305 - PRECISÃO (**PRECISION**): A precisão se refere ao grau de proximidade que uma estimativa possui com o seu valor esperado.

306 - A característica operacional do receptor ou a curva operacional do receptor de um classificador de padrões é a função de sua taxa de erro por má detecção contra sua taxa de alarmes falsos.

307 - O método de avaliação de um classificador, deixe k de fora, divide o conjunto de dados de treinamento em L subconjuntos mutuamente exclusivos, cada um contendo k elementos. O classificador é sucessivamente treinado usando $L - 1$ dos subconjuntos e testado no L -ésimo subconjunto. A avaliação é feita na performance acumulada dos testes dos experimentos, onde em cada experimento k padrões são omitidos do conjunto de treinamento e usados no conjunto de teste. As estimativas de performance obtidas usando-se o método deixe k de fora são imparciais. Contudo, para valores pequenos de k , as estimativas terão grande variância.

308 - O método de resubstituição para avaliar um classificador usa o mesmo conjunto para treinar e testar o classificador. As estimativas de performance obtidas usando o método da resubstituição são sempre muito tendenciosas.

Bibliografia

- [Abu89] A.S. Abutaleb. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Graphics and Image Processing*, 47(1):22–32, july 1989.
- [Ber86] J. Bernsen. Dynamic thresholding of gray-level images. *Proc. Eighth Int'l Conf. Conf. on Pattern Recognition, Paris, France*, pages 1251–1255, 1986.
- [CW94] E. Casey and S. Wong. Adaptative thresholding methodology for images based on intensity gradient information. *Computer Vision, Graphics and Image Processing*, 51:238–247, 1994.
- [FF87] M.A Fischler and O. Firschein. *Intelligence: the Eye, the Brain and the Computer*. Addison Wesley, 1987.
- [HD88] Q.Z Haralick and P.E Danielson. On minimum error thresholding and its implementations. *Pattern Recognition Letters*, 7(4):201–206, april 1988.
- [JB82] G. Johannsen and J. Bille. A threshold selection method using information measures. *Proceedings, 6th Int. Conf. Pattern Recognition, Munich, Germany*, pages 140–143, 1982.
- [Kat65] Y.H. Katz. Pattern recognition od meteorological satellite cloud photography. *Proc. 3rd Symp. Remote Sensing Environment*, pages 173–214, 1965.
- [KZ93] G.H. Kamel and J. Zhao. *Analyzing Oriented Patterns*. IEEE Computer Society Press, 1993.
- [Mas75] D. Mason. Measurement of c-bands in human chromosomes. *Computers Biology and Medicine*, 5:179–201, 1975.
- [MG89] C.H. Mitchell and L. Gillies. *Fundamentals of Digital Image Processing*. Prentice-Hall Information and System Sciencies Series, Englewood Cliffs, NJ, 1989.
- [Pun81] T. Pun. Entropic thresholding, the new approach. *Computer Graphics and Image Processing*, 16:210–239, 1981.
- [RRK84] S.S. Reddi, S.F. Rudin, and H.R Keshavan. An optimal multiple threshold scheme for image segmentation. *IEEE Trans. Systems, Man and Cybernetics*, CMC-14(4):661–665, july-august 1984.
- [RT83] A. Rosenfeld and P. De La Torre. Histogram concavity analysis as an aid in threshold selection. *IEEE Trans. Systems, Man and Cybernetics*, CMC-13:231–235, 1983.

- [SSW88] P.K. Sahoo, S. Soltani, and A.K.C. Wong. A survey of thresholding techniques. *Computer Graphics and Image Processing*, 41:233–260, 1988.
- [Tsa85] W.H. Tsai. Moment-preserving thresholding: A new approach. *Computer Graphics and Image Processing*, 29(3):377–393, march 1985.
- [Wat74] S. Watanabe. A automated apparatus for cancer processing cybest. *Computer Graphics and Image Processing*, 3(4):350–358, december 1974.
- [WNR74] J.S. Weszka, R.N. Nagel, and A. Rosenfeld. A threshold selection technique. *IEEE Trans. Computer*, C-23:1322–1326, 1974.
- [Won78] K. Wong. *Image Segmentation by texture*. IEEE Computer Society Press, 1978.
- [WR83] J.M. White and G.D. Rohrer. Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM J. Research and Development*, 27(4):400–411, july 1983.

Índice

- área, 50
- área do operador de interesse, 59
- ótica inversa, 65
- abertura, 46
- acutance, 34
- afinamento, 47
- agudização, 21
- ajuste iterativo de curvas, 53
- alongamento, 52
- alongamento de contraste, 41
- análise
 - de área, 60
 - de assinaturas, 52
 - de componentes conectados, 52
 - de sombreamento, 63
 - dinâmica, 64
- aspecto de visão, 65
- backlighting, 37
- Bayes, 72
- binarização, 7, 41
 - multimodal, 41
- borda, 23, 49
 - conexão de, 45
 - detecção, 9
 - detecção de, 45
 - detecção por passagem por zero, 12
 - imagem de, 45
 - oclusa, 45
 - perseguição de, 25
 - simples, 49
- bottom up, 66
- brilho, 19
- buraco, 49
- busca
 - de contorno, 50
 - de região, 57
- código de cadeia, 53
- calibragem visual, 66
- campo de movimento, 63
- característica, 24
 - intrínseca, 64
 - tonal discreta, 56
- centróide, 50
- classes de padrões, 69
- classificação de padrões, 74
- classificação, 21
- classificador, 69
- cluster, 74
- clustering, 74
- compacidade, 51
- componente conectado, 49, 52
- compressão, 60
- concurva, 50
- cone generalizado, 54
- conexão de bordas, 45
- congruência, 60
- conjunto
 - de intervalos, 69
 - multi imagens, 35
- contorno, 49
- contraste, 34
- convolução, 43
- coordenadas
 - HS, 38
 - rgb, 37
 - YIQ, 38
- cor, 13
- correlação, 43
- crescimento de região, 57
- curva piramidal, 53
- descrição
 - 2D, 50
 - de superfície de borda, 54
 - de textura, 57
 - de textura estrutural, 57
 - estrutural, 58
- detecção
 - de bordas, 45
 - de Duda, 11
 - de Hueckel, 11
 - de mudança, 42
- dilatação, 46
- discriminação de padrões, 73
- disparidade, 16, 62

- distância
 - City-Block, 48
 - Euclidiana, 48
 - principal, 39
 - quadrada, 48
- divisão-fusão, 26
- edgel, 35
- eixo
 - ótico, 39
 - epipolar, 62
 - simétrico, 52
- elemento estruturante, 46
- emparelhamento
 - de imagens, 58
 - de padrões, 57
 - de pontos característicos, 58
 - estéreo, 62
 - relacional, 59
 - simbólico, 58
- epipólo, 62
- erosão, 46
- erro de identificação, 74
- esboço
 - $2^{1/2}D$, 64
 - primitivo, 55
- esfera Gaussiana, 48
- espaço
 - de características, 69
 - de medição, 69
 - do Gradiente, 48
- espessamento, 47
- estéreo fotométrico, 63
- estereoscopia, 16
- estratégia
 - de controle de quadro-negro, 67
 - de controle hierárquico, 67
- estrutura
 - a partir do movimento, 64
 - espacial escalar, 46
- extração de características, 69
- feature, 69
- fechamento, 47
- filtro
 - da média, 8, 20
 - da mediana, 21, 45
 - de emparelhamento, 58
 - espacial, 43
 - espacial linear, 43
 - Gaussiano, 43
 - linear, 24
 - minmax, 21
 - não linear, 24
 - passa-alto, 44
 - passa-baixo, 44
 - passagem de banda, 44
 - por convolução, 43
- fixação visual, 66
- fluxo ótico, 5, 16, 63
- fluxo luminoso, 36
- foco
 - de contração, 63
 - de expansão, 63
- foco de expansão, 17
- forma
 - a partir da textura, 64
 - a partir do contorno, 64
 - de sombreado, 63
- fotogrametria, 38
 - analítica, 38
 - digital, 38
- Fourier
 - análise de, 14
 - descritores, 53
 - transformada, 44
- fractal, 14
- fronteira
 - de decisão hiperplanar, 72
 - oclusa, 62
- frontlighting, 37
- função
 - de atribuição de grupo, 74
 - discriminante, 72
 - discriminante linear, 72
- fusão de regiões, 57
- geometria
 - construtiva, 54
 - digital, 48
- gradiente, 10, 22, 23
 - Sobel, 10
- grafo de aspecto, 65
- ground truth, 71
- hipótese de compacidade, 70
- histograma, 8, 22, 28, 36
 - de diferença dos níveis de cinza, 56
- identificação de padrões, 73
- identificação falsa, 74
- iluminância, 36
- iluminação, 8, 19, 36
 - ambiente, 37
- imagem, 33
 - aquisição, 5
 - binária, 36

- de bordas, 45
 - de vídeo, 33
- digital, 35
- em níveis de cinza, 36
- espacial escalar, 46
- Gaussiana, 48
- intrínseca, 64
- multi temporal, 36
- multi-spectral, 36
- piramidal, 44
- range, 35
- simbólica, 36
- inspeção visual automática, 65
- intensidade radiante, 37
- irradiância, 37
-
- julgamento de reserva, 70
-
- kernel, 43
-
- Laplaciano, 19, 21, 24, 29
- light striping, 61
- likelihood, 72
- limiarização, 7, 26
 - bimodal, 41
 - global, 28
 - local adaptativa, 30
 - multimodal, 32, 41
- linha epipolar, 62
- local feature focus method, 59
- localização, 59
- luz estruturada, 61
-
- mapa
 - de orientação, 35
 - de profundidade, 35
 - de refletância, 37
 - refletância, 62
- mascaragem, 42
- matching, 57, 62
- matriz
 - de co-ocorrência, 14, 56
 - de confusão, 74
 - do segundo momento, 51
- medidas estatísticas de textura, 56
- modelo de facetas, 55
- momento, 50
 - central, 50
 - central normalizado, 50
 - de rotação, 50
- morfologia matemática, 46
- mundo de blocos, 65
-
- número de Euler, 51
- nível de cinza, 34
-
- navegação passiva, 64
-
- octree, 54
- operador
 - de bordas, 44
 - de Imagem, 40
 - de interesse, 59
 - morfológico, 46
 - pontual, 41
- orientação
 - absoluta, 39
 - externa, 38
 - interna, 39
 - relativa, 38
-
- padrão, 69
 - de característica, 69
 - de medição, 68
 - de referência, 70
- par de imagens estéreas, 61
- paralaxe, 40
- pares de pontos correspondentes, 62
- passagem por zero, 12
- pattern recognition, 68
- perímetro, 50
- perador de vizinhança, 42
- perceptron, 73
 - backward, 74
 - forward, 73
 - series, 73
- perseguição de borda, 25, 49
- pixel, 6, 34
 - de borda externa, 49
 - de borda interna, 49
 - extremo, 51
 - interno, 49
- plano epipolar, 62
- polígono perimetral mínimo, 50
- ponto
 - de fuga, 39
 - morto, 40
 - principal, 39
- preenchimento de contorno, 57
- preprocessamento, 19, 60
- probabilidade condicional, 71
- procedimento
 - de treinamento, 71
 - por correção, 71
 - por janela, 71
 - robusto de visão, 66
- processamento, 42
- processamento interativo , 61
- produto cartesiano, 68
- projecção

- de uma imagem binária, 53
 - paralela, 40
 - perspectiva, 39
- quadtree, 53
- quantização, 41
- radiância, 37
- raio epipolar, 62
- rangel, 35
- raster scan, 35
- realce, 61
- reconhecimento
 - de padrões, 68
 - de padrões pictoriais, 68
- reconhecimento de padrões, 18
- reconstrução, 60
- reconstrução de superfície, 64
- rede neural, 73
- refletância, 19, 37
- reforço de correção de erros, 73
- região, 48
 - conectada, 49
 - convexa, 49
 - de classe, 70
 - homogênea, 26
- registration, 60
- regra de decisão, 70
 - Bayesiana, 72
 - composta, 70
 - do vizinho mais próximo, 72
 - hierárquica, 70
 - linear, 73
 - não paramétrica, 72
 - por semelhança, 72
 - simples, 70
- reinforcement, 73
- relação de emparelhamento, 59
- relaxação, 59
- resolução, 34
- restauração, 21, 60
- retângulo delimitador, 51
- retas anti-paralelas, 45
- ruído sal e pimenta, 20
- run length encoding, 54
- screening, 60
- segmentação, 21, 25, 57
- segmento de reta, 49
- seleção de características, 69
- separabilidade, 70
- seqüência
 - de dados, 70
 - de predição, 74
- de treinamento, 71
- sistema
 - de inspeção flexível, 66
 - de visão por máquina, 65
- smoothing, 20
- stereopsis, 61
- strip tree, 55
- suavização, 7, 8, 20, 45
 - em modelos, 8
 - geométrica, 9
- subimagem, 48
- superquadrática, 55
- superfície Luminosa, 37
- texel, 56
- textura, 13, 15, 56
- thresholding, 41
- top down, 67
- transformada
 - de distância, 48
 - de Fourier, 44
 - de Hough, 25, 47
 - de Radon, 47
- triangulação, 62
- unidade, 68
- vetor n-tupla de medição, 68
- visão, 61
 - 3D , 65
 - computacional, 1, 66
 - computacional baseada em con-
hecimento, 67
 - computacional baseada em mod-
elos, 67
- vizinhança 4-8, 49
- voxel, 34