

Processamento Digital de Imagens

Israel Andrade Esquef^a Márcio Portes de Albuquerque^b
Marcelo Portes de Albuquerque^b

^a*Universidade Estadual do Norte Fluminense - UENF*

^b*Centro Brasileiro de Pesquisas Físicas - CBPF*

Resumo

As imagens são produzidas por uma variedade de dispositivos físicos, tais como câmeras e vídeo câmeras, equipamentos de radiografia, microscópios eletrônicos, magnéticos e de força atômica, radares, equipamento de ultra-som, entre vários outros. A produção e utilização de imagens podem ter diversos objetivos, que vão do puro entretenimento até aplicações militares, médicas ou tecnológicas. O objetivo da análise de imagens, seja por um observador humano ou por uma máquina, é extrair informações úteis e relevantes para cada aplicação desejada.

Em geral, a imagem pura, recém adquirida pelo dispositivo de captura, necessita de transformações e realces que a torne mais adequada para que se possa extrair o conteúdo de informação desejada com maior eficiência. O Processamento Digital de Imagens (PDI) é uma área da eletrônica/teoria de sinais em que imagens são convertidas em matrizes de números inteiros, sendo que cada elemento desta matriz é composta por um elemento fundamental: o *pixel* (uma abreviação de *picture element*). A partir desta matriz de pixels que representa a imagem, diversos tipos de processamento digital podem ser implementados por algoritmos computacionais. A aplicação destes algoritmos realizam as transformações necessárias para que se possa, por exemplo, obter uma imagem com os realces pretendidos ou extrair atributos ou informações pertinentes. O PDI será considerado neste trabalho como a união das áreas de processamento de imagem e visão computacional.

1 Representação da Imagem Digital

Uma imagem monocromática é uma função bidimensional $f(x, y)$ da intensidade luminosa, onde x e y denotam coordenadas espaciais, que por convenção: $x = [1, 2, \dots, M]$ e $y = [1, 2, \dots, N]$. O valor de f no ponto (x, y) é proporcional ao brilho (ou nível de cinza) da imagem neste ponto, como ilustrado na figura 1. Esta figura ¹ apresenta uma região em destaque em que se pode observar os pixels e os níveis de cinza ou níveis de luminância de cada um deles.

Pixel e Conectividade

Um pixel é o elemento básico em uma imagem. A forma mais comum para o pixel é a forma retangular ou quadrada. O pixel é também um elemento de

¹ A imagem “Goldhill” é freqüentemente utilizada para testes e demonstrações em PDI



Figura 1. Imagem monocromática “Goldhill” com destaque para uma região de 17 x 17 pixels

dimensões finitas na representação de uma imagem digital. Frequentemente, a organização de uma imagem sob a forma de uma matriz de pixels é feita em uma simetria quadrada. Isto se deve a facilidade de implementação eletrônica, seja dos sistemas de aquisição seja dos sistemas de visualização de imagens. Este tipo de organização provoca o aparecimento de dois problemas importantes nas técnicas de processamento. Em primeiro lugar um pixel não apresenta as mesmas propriedades em todas as direções, isto é, ele é anisotrópico. Esta propriedade faz com que um pixel tenha quatro vizinhos de borda e quatro vizinhos de diagonal, ilustrado na figura 2. Esta propriedade obriga que seja definido o tipo de conectividade que será utilizada, ou B4 (considerando apenas os vizinhos de borda) ou B8 (considerando os vizinhos de borda e os de diagonal). O segundo problema é consequência direta do primeiro, ou seja, as distâncias entre um ponto e seus vizinhos não é a mesma para qualquer tipo de vizinho. Será igual a 1 para vizinhos de borda e $\sqrt{2}$ para aqueles na diagonal.

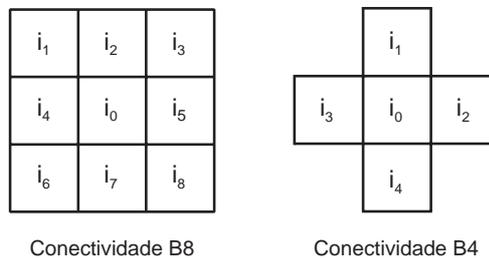


Figura 2. Ilustração de tipos de conectividade dos pixels vizinhos ao pixel central i_0 . Conectividade B8 apresenta 8 vizinhos, sendo 4 de bordas e 4 diagonais. Conectividade B4 apresenta apenas os pixels de borda.

2 Etapas do Processamento de Imagens

Um sistema de processamento de imagens é constituído de diversas etapas, tais como: formação e aquisição da imagem, digitalização, pré-processamento, segmentação, pós-processamento, extração de atributos, classificação e reconhecimento, como ilustra a figura 3. A seguir, serão apresentadas breves descrições de cada uma destas etapas.

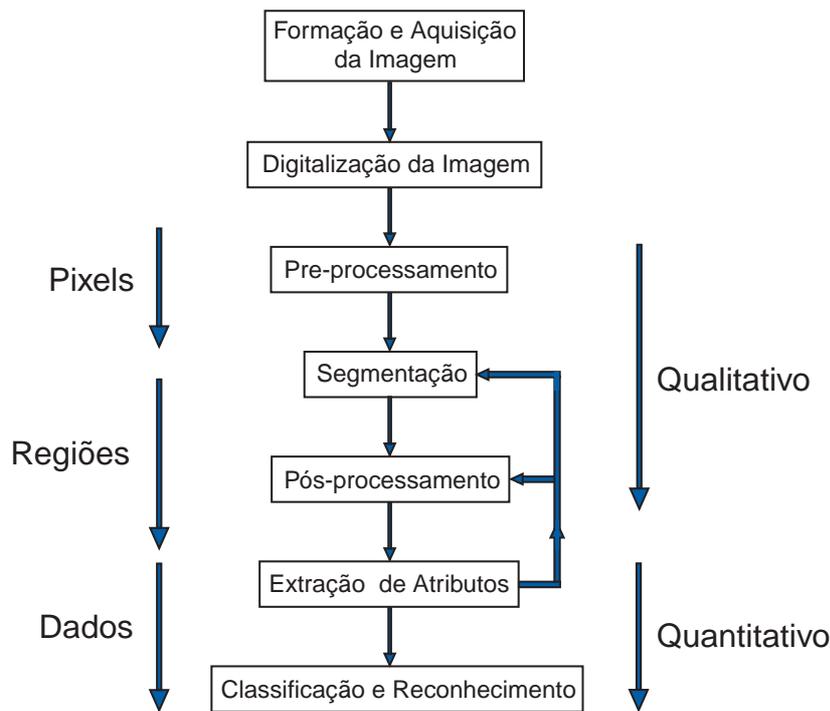


Figura 3. *Etapas de um sistema de PDI.*

2.1 Aquisição de Imagens Digitais

Dois elementos são necessários para a aquisição digital de imagens. O primeiro é um dispositivo físico que deve ser sensível ao espectro de energia eletromagnético, como por exemplo ao espectro de raio-x, luz ultravioleta, visível, ou infravermelha. Este dispositivo transdutor deve produzir em sua saída um sinal elétrico proporcional ao nível de energia percebido. O segundo, chamado digitalizador, é um dispositivo que converte o sinal elétrico analógico produzido na saída do sensor em um sinal digital.

Em laboratórios de física experimental diversos processos levam a uma organização bi-dimensional de dados, buscando uma representação dos mesmos sob a forma de imagens. A grande vantagem é que esta representação permite

que ao observarmos todo o conjunto de dados, nós possamos bucar correlações espaciais entre eles.

2.2 Técnicas de Pré-processamento

As técnicas de pré-processamento têm a função de melhorar a qualidade da imagem. Estas técnicas envolvem duas categorias principais: métodos que operam no domínio espacial e métodos que operam no domínio da frequência. Técnicas de processamento no domínio espacial baseiam-se em filtros que manipulam o plano da imagem, enquanto que as técnicas de processamento no domínio da frequência se baseiam em filtros que agem sobre o espectro da imagem. É comum para realçar determinadas características de uma imagem, combinar vários métodos que estejam baseados nestas duas categorias. A figura 4 ilustra um pré-processamento simples: a aplicação de um filtro mediana, para redução de ruído e em seguida um filtro passa-altas, para realce nos contornos ou bordas dos objetos na imagem.

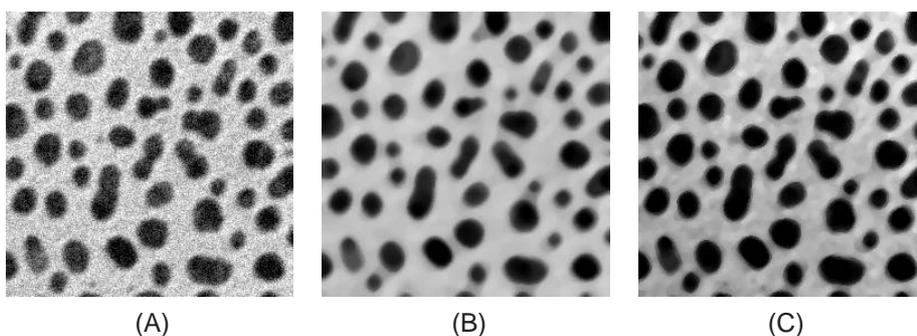


Figura 4. *Exemplo de um pre-processamento simples: (A) Imagem original corrompida com ruído gaussiano, (B) Imagem após a aplicação de um filtro mediana para redução do ruído, e (C) Imagem final, após a aplicação de um filtro passa-altas para realce dos contornos.*

2.2.1 O Histograma de Luminância

O histograma de uma imagem digital com k níveis de cinza é definido por uma função discreta

$$p(k) = \frac{n_k}{n} \quad (1)$$

em que o argumento k representa os níveis de luminância discretos, n_k representa o número de pixels na imagem com intensidade k e n é o número total de pixels da imagem, ou seja, $n = M \times N$. De forma simplificada, podemos afirmar que o histograma de luminância de uma imagem representa a contagem

dos níveis de cinza da imagem, podendo informar a distribuição dos pixels dentro dos k níveis possíveis. O histograma pode ser considerado como uma função distribuição de probabilidades, obedecendo aos axiomas e teoremas da teoria de probabilidades, *i.e.* que $\sum_k p(k) = 1$.

O histograma da imagem digital é uma ferramenta bastante útil na etapa de pre-processamento, fornece uma visão estatística sobre a distribuição dos pixels, sobre o contraste da imagem e os níveis de iluminação. Além disso, o histograma é bastante utilizado na etapa de segmentação, principalmente em técnicas que se utilizam da similaridade entre os pixels. O histograma é utilizado com frequência como sendo uma distribuição estatística dos pixels (“luminância”) na imagem, como por exemplo no caso das técnicas que o utilizam para calcular a entropia da imagem.

Na figura 5 são apresentadas duas imagens e seus histogramas, sendo que a imagem (A) é uma imagem de baixo contraste, enquanto a imagem (B) possui um maior contraste.

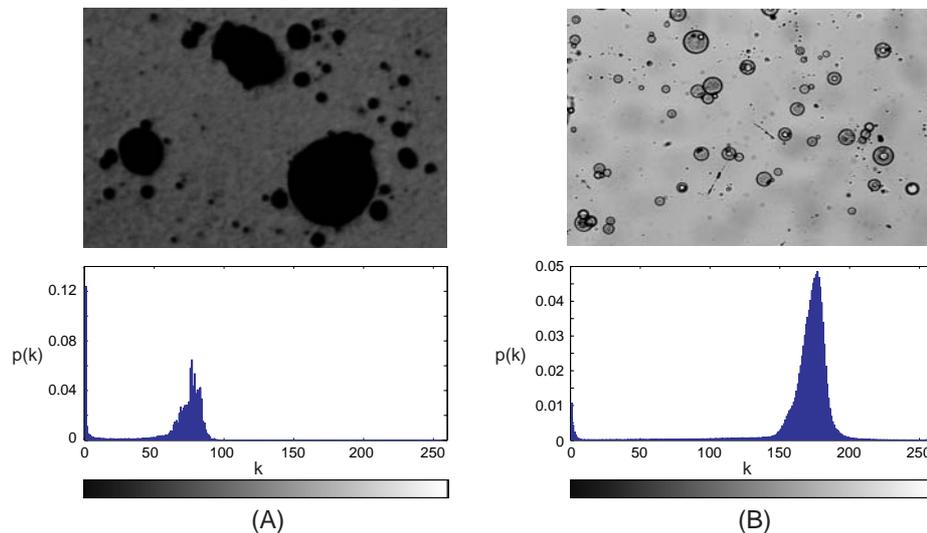


Figura 5. *Exemplo de imagens com histogramas diferenciados. (A) Imagem de baixo contraste e seu histograma de luminância. (B) Imagem de alto contraste e seu histograma de luminância.*

A figura 6 apresenta outras duas imagens com histogramas de tipos bimodal e multimodal. O histograma bimodal é classificado desta forma devido a presença de dois picos, ou duas regiões de luminância com maior incidência de pixels da imagem. Este é o caso típico de imagens que apresentam objetos e fundo de maneira bem definida, como a imagem (A) da figura 6. O histograma multimodal apresenta os pixels distribuídos em mais de duas regiões de níveis de cinza, o que se pode perceber através dos picos no histograma em torno destes valores de luminância. Este é o caso da imagem (B) da figura 6, que apresenta três picos bem definidos.

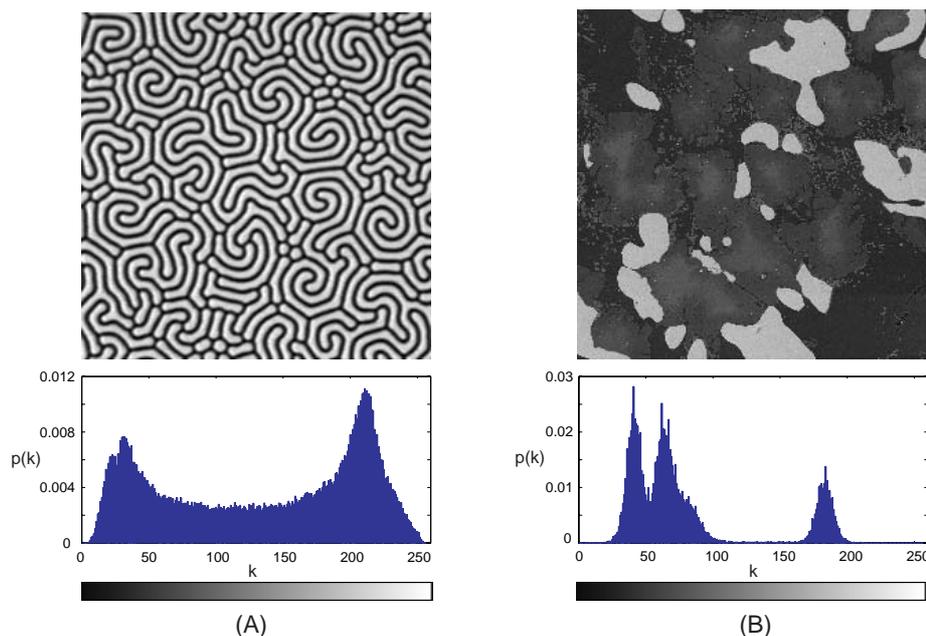


Figura 6. *Exemplo de imagens com histogramas bimodal e multimodal. A imagem (A) possui um histograma bimodal típico, com a presença de dois picos bem definidos. A imagem (B) possui um histograma do tipo multimodal, apresentando mais de duas regiões de concentração dos pixels da imagem.*

2.3 Segmentação

Segmentar uma imagem significa, de modo simplificado, separar a imagem como um todo nas partes que a constituem e que se diferenciam entre si. É usual denominar “objetos” da imagem os grupos de pixels de interesse, ou que fornecem alguma informação para o PDI. Da mesma forma, a denominação “fundo” da imagem é utilizada para o grupo de pixels que podem ser desprezados ou que não têm utilidade no PDI. Essas denominações “objeto” e “fundo” possuem uma conotação bastante subjetiva, podendo se referir a grupos de pixels que formam determinadas regiões na imagem sem que representem um objeto, de modo literal, presente na imagem processada.

A segmentação é considerada, dentre todas as etapas do processamento de imagens, a etapa mais crítica do tratamento da informação. É na etapa de segmentação que são definidas as regiões de interesse para processamento e análise posteriores. Como consequência deste fato, quaisquer erros ou distorções presentes nesta etapa se refletem nas demais etapas, de forma a produzir ao final do processo resultados não desejados que podem contribuir de forma negativa para a eficiência de todo o processamento.

Deve ser ressaltado que não existe um modelo formal para a segmentação de

imagens. A segmentação é um processo empírico e adaptativo, procurando sempre se adequar às características particulares de cada tipo de imagem e aos objetivos que se pretende alcançar. Apesar de existir uma grande diversidade de técnicas de segmentação de imagens, ainda assim existe atualmente, um grande interesse no estudo e desenvolvimento de novas técnicas.

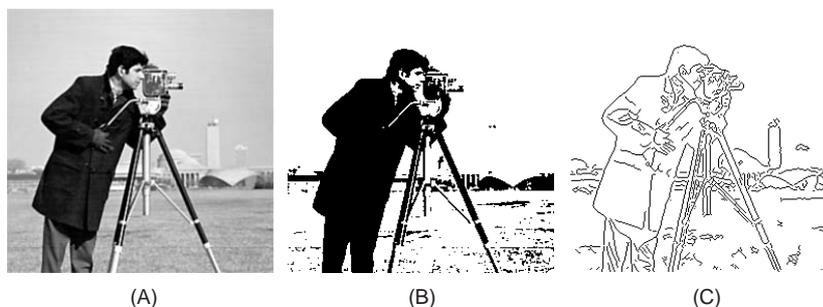


Figura 7. Duas abordagens para segmentação. (A) Imagem original em níveis de cinza. (B) Imagem segmentada através de uma binarização. (C) Imagem segmentada por detecção de bordas.

De um modo geral, as técnicas de segmentação utilizam duas abordagens principais: a similaridade entre os pixels e a descontinuidade entre eles. Sem dúvida, a técnica baseada em similaridade mais utilizada é a chamada *binarização*. A binarização de imagens ou *image thresholding* é uma técnica eficiente e simples do ponto de vista computacional, sendo portanto largamente utilizada em sistemas de visão computacional. Este tipo de segmentação é utilizado quando as amplitudes dos níveis de cinza são suficientes para caracterizar os “objetos” presentes na imagem. Na binarização, um nível de cinza é considerado como um limiar de separação entre os pixels que compõem os objetos e o fundo. Nesta técnica, se obtém como saída do sistema uma imagem binária, i.e., uma imagem com apenas dois níveis de luminância: preto e branco. A determinação deste limiar de modo otimizado para segmentação da imagem é o objetivo principal dos diversos métodos de binarização existentes.

As técnicas baseadas em descontinuidade entre os pixels procuram determinar variações abruptas do nível de luminância entre pixels vizinhos. Estas variações, em geral, permitem detectar o grupo de pixels que delimitam os contornos ou bordas dos objetos na imagem. A técnica de segmentação baseada em descontinuidade mais utilizada é a chamada detecção de bordas. A figura 7 apresenta dois exemplos de segmentação, uma binarização e uma detecção de bordas. A imagem (A) é a imagem original em níveis de cinza, a imagem (B) foi segmentada por binarização e a imagem (C) foi segmentada por detecção de bordas.

2.4 Pós-processamento

O pós-processamento geralmente é a etapa que sucede a segmentação. É nesta etapa que os principais defeitos ou imperfeições da segmentação são devidamente corrigidos. Normalmente, estes defeitos da segmentação são corrigidos através de técnicas de Morfologia Matemática, com a aplicação em seqüência de filtros morfológicos que realizam uma análise quantitativa dos pixels da imagem.

2.4.1 Operações Morfológicas Básicas

A Morfologia Matemática (MM) é uma das grandes áreas do Processamento Digital de Imagens. Todos os métodos descritos pela MM são fundamentalmente baseados em duas linhas: os operadores booleanos de conjuntos (união, interseção, complemento etc.) e a noção de forma básica, chamado de “elemento estruturante”. As operações são realizadas sempre entre a imagem e o elemento estruturante. A forma do elemento estruturante é função do tratamento desejado e do tipo de conectividade adotada (B4 ou B8).

Dois operadores básicos são utilizados na maior parte das técnicas de MM: a erosão e a dilatação. Consideremos por exemplo, um objeto X como um grupo de pixels x delimitado por uma linha tracejada, como mostrado na figura 8. A operação de erosão consiste em eliminar do conjunto X os pixels x em função do elemento estruturante B , tal que:

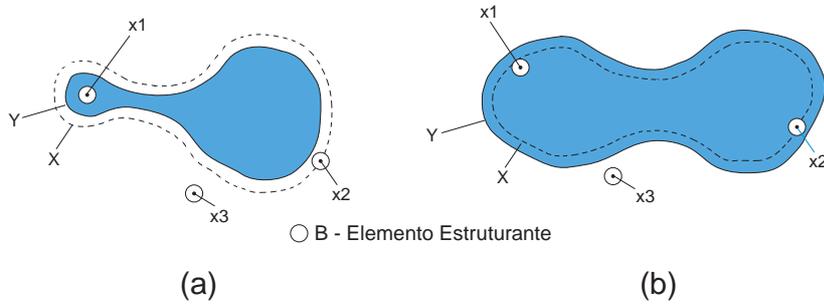


Figura 8. Operação de erosão (a) e dilatação (b) aplicada a um elemento estruturante B . Obtemos a forma Y . x_1 é um ponto de X que também pertence a Y . Em (a) x_2 pertence a X , mas não a Y devido a operação de erosão. Por outro lado, em (b) x_2 passa a pertencer a Y devido a dilatação. x_3 não pertence a ambos X e Y .

$$Y = E^B(X) \rightarrow Y = \{x/B(x) \subset X\} \quad (2)$$

onde $B(x)$ é o elemento estruturante centrado no pixel x . Na prática este procedimento corresponde a construir um novo conjunto de pontos Y , a partir

do conjunto X , tal que o elemento estruturante esteja inserido totalmente em X . A figura 8(a) apresenta o resultado da operação de erosão para 3 pontos distintos da imagem, \mathbf{x}_1 , \mathbf{x}_2 e \mathbf{x}_3 . O elemento estruturante \mathbf{B} (representado por um círculo), centrado em \mathbf{x}_1 , está totalmente inserido em X . O ponto \mathbf{x}_1 pertencerá a forma resultante Y . O ponto \mathbf{x}_2 , na borda de X , e \mathbf{x}_3 fora de X não farão parte de Y .

A operação dual da Erosão é a Dilatação. Esta operação consiste em dilatar o objeto X com o elemento estruturante B tal que:

$$Y = D^B(X) \rightarrow Y = (E^B(X^c))^c = \{x/B(x) \cap \subset X \neq \emptyset\} \quad (3)$$

em que c representa o complemento da operação booleana. A figura 8(b) apresenta o resultado da operação de dilatação para 3 pontos da imagem.

A operação de erosão permite separar objetos que se tocam. Ao contrario, a operação de dilatação permite preencher furos no interior de um objeto ou mesmo ligá-los. Este resultado dependerá da forma do elemento estruturante. Como as duas operações são iterativas é possível realizar uma seqüência de N operações de erosão e dilatação sucessivas ou mesmo alternadas. A operação de abertura, uma erosão seguida de uma dilatação, permite eliminar pequenas partículas na imagem (partículas do tamanho do elemento estruturante) e suavizar o contorno dos objetos. Inversamente, a operação de fechamento, uma dilatação seguida de uma erosão, permite fechar canais estreitos que separam objetos, ou suprimir os pequenos furos no seu interior.

2.5 Extração de Atributos

A etapa final de um sistema de processamento de imagens é aquela em que se extrai as informações úteis da imagem processada. Quando o objetivo do processamento é obter informações numéricas, realiza-se a extração de atributos da imagem.

2.5.1 Rotulação ou Labelização

A etapa chamada Labelização ou Rotulação é uma etapa intermediária na extração de atributos. Após a etapa de segmentação obtemos uma imagem onde as regiões correspondentes aos “objetos” estão separadas daquelas correspondentes ao “fundo” da imagem. Neste ponto do sistema de processamento, as regiões de interesse estão contiguamente agrupadas por pixels que se tocam. O próximo passo é dar um rótulo (ou *label*) para cada um desses grupos de pixels. Esta identificação permitirá posteriormente parametrizar os objetos

segmentados calculando para cada região de pixels contíguos um parâmetro específico, como área ou perímetro por exemplo. A figura 9 apresenta um exemplo desta técnica para uma imagem constituída de células bem delimitadas entre si. O processo de segmentação separa as regiões pertencentes às células daquelas pertencentes às regiões entre células (fundo), criando um delimitador entre elas. A etapa de “labelização” cria um rótulo que identifica cada uma dessas regiões para que os processos seguintes de tratamento da informação sejam concentrados em cada uma das regiões quer receberem um rótulo.

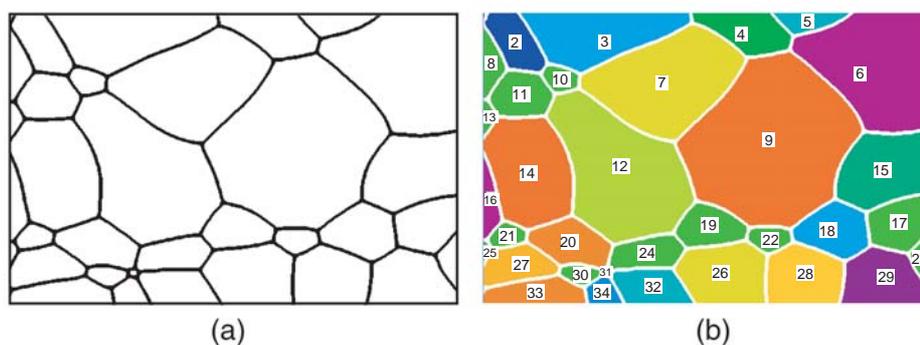


Figura 9. Imagem ‘Labelizada’: (a) Imagem original composta por regiões contíguas de pixels. (b) Imagem final após o processo de rotulação. As cores são utilizadas para auxiliar na visualização das regiões.

2.5.2 Atributos da Imagem

Existem basicamente duas classes de medidas: *i*) atributos da imagem como um todo (*field features*), por exemplo número de objetos, área total de objetos, etc. e *ii*) atributos de região (*region features*) que se referem aos objetos independentemente, por exemplo área, perímetro, forma, etc. Os atributos de região podem ser muito sofisticados, permitindo uma nova separação dos objetos em classes de similaridades, em função dos parâmetros medidos. A figura 10 apresenta os principais atributos de região que podem ser extraídos de uma imagem, após as etapas de segmentação e pós-processamento.

2.6 Classificação e Reconhecimento

O objetivo do reconhecimento é realizar, de forma automática, a “identificação” dos objetos segmentados na imagem. Existem duas etapas no processo de classificação de formas: o aprendizado e o reconhecimento propriamente dito. Na maior parte dos sistemas de reconhecimento de formas, os parâmetros provenientes da etapa de extração de atributos são utilizados para construir um espaço de medida à N dimensões. Os sistemas de aprendizado irão definir uma função discriminante que separe eficientemente todas as formas representadas neste espaço de medida.

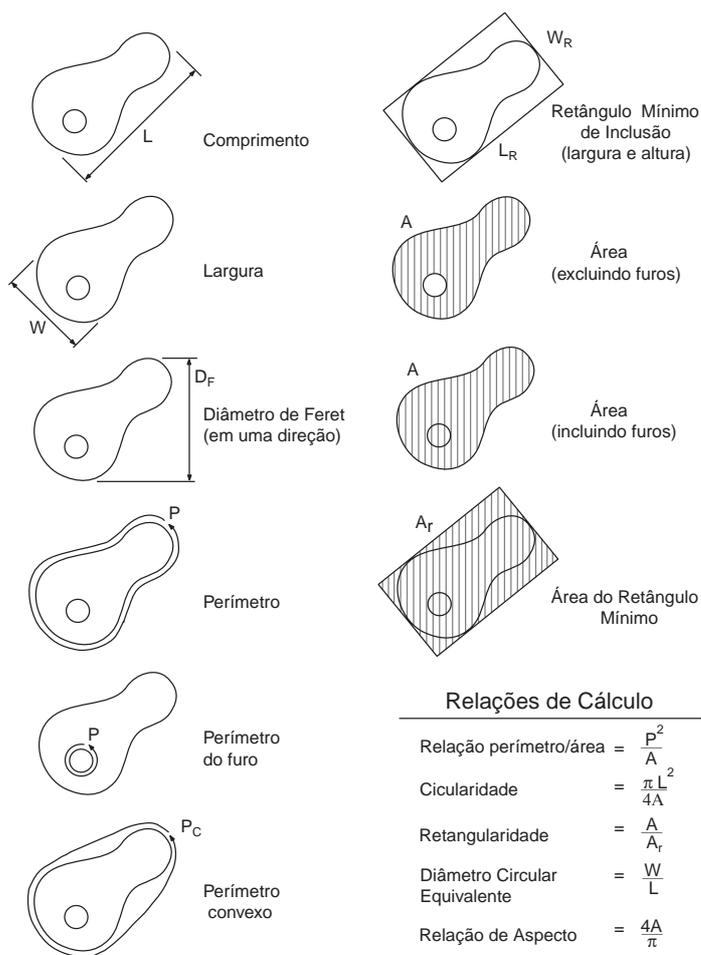


Figura 10. Principais atributos de região, ou seja, dos objetos independentes presentes na imagem.

O espaço de medidas é um espaço de dimensão N onde cada dimensão corresponde a um parâmetro (atributo). A escolha e a qualidade dos parâmetros para construir este espaço é essencial para a realização de um bom processo de reconhecimento de forma. A escolha de um grande número de parâmetros leva a um espaço de grande dimensão e a uma difícil etapa de aprendizado. Um pequeno espaço de medidas pode acarretar em uma baixa caracterização da forma e a muitos erros no processo de reconhecimento. Em alguns casos pode ser interessante normalizar cada eixo para construir um espaço que seja bem distribuído e facilite o processo de classificação.

Podemos dividir o processo de aprendizado em dois tipos: os métodos supervisionados e não supervisionados. No método supervisionado, o classificador, em sua fase de aprendizado, recebe informações de como as classes devem ser identificadas. Por exemplo, em um sistema de reconhecimento de caracteres, existirá classes independentes para todas as letras do alfabeto. Por exemplo a classe das letras “**A**” será representada pela ocorrência da letra A em suas diferentes variações.

O aprendizado consiste então em apresentar ao sistema uma série de objetos “teste” no qual suas classes são previamente conhecidas e definir uma função que separe todas as classes entre si.

Podemos dizer que o sistema de aprendizado supervisionado age sob a supervisão de um outro sistema de reconhecimento (externo por exemplo) que identificou anteriormente os objetos testes e permitirá a construção correta de seu espaço de medida e sua função discriminante. Durante este processo devemos modificar os parâmetros que compõem o espaço de medida e permitir um melhor ajuste da função discriminante, objetivando sempre que o sistema possa realizar com mais eficiência o processo de classificação. Ao final, é possível determinar a função discriminante responsável pela separação das diversas classes. Este processo pode ser lento e de elevado custo computacional. Somente então os objetos desconhecidos serão fornecidos a este classificador, na fase de reconhecimento.

No caso em que a classificação não é supervisionada, o classificador receberá os objetos desconhecidos e, a partir da medida dos diferentes parâmetros (atributos dos objetos presentes na imagem), ele tentará alocá-los em diferentes classes. A identificação de classes é usualmente realizada a partir da identificação de agrupamentos em “clusters” de objetos no espaço de medidas.

Existem atualmente vários métodos de reconhecimento de formas. Entre eles podemos citar os métodos baseados em propriedades estatísticas dos objetos (classificador de Bayes), os métodos baseados em técnicas de distância entre os objetos na imagem e suas formas padrões (como as redes neurais artificiais) ou ainda a descrição da forma através de um dicionário ou uma linguagem básica. Neste último caso é definido uma seqüência de elementos básicos que representem as formas dos objetos. Em seguida é construída uma linguagem e formada uma gramática. O reconhecimento é neste caso um reconhecimento sintático baseado nesta gramática [1]. O reconhecimento por meio de redes neurais artificiais é um método bastante atraente, pois consegue dar um maior grau de liberdade a função discriminante. Nesta tese apresentamos de forma introdutória, no capítulo 6, um método de classificação de formas baseado na técnica de entropia cruzada entre objetos da imagem e padrões previamente definidos.

Referências

- [1] Gonzalez, R. C. and Woods, R. E., *Digital Image Processing*, Addison Wesley, 1993.