Introdução ao Processamento Digital de Imagens

José Eustáquio Rangel de Queiroz [1], Herman Martins Gomes [1]

Resumo: O objetivo desse tutorial é fornecer uma visão introdutória para a área de Processamento Digital de Imagens (PDI) de modo que possa servir como base de estudo para iniciantes na área ou como referência para estudos mais avançados. O tutorial está dividido em duas partes: uma parte principal contemplando os fundamentos e uma parte complementar descrevendo aplicações. A parte de fundamentos apresenta o processo de formação de imagens, incluindo uma sucinta apresentação da estrutura do olho humano e sua analogia com uma câmera digital, bem como comentários sobre um sistema típico de PDI. O núcleo do tutorial aborda as principais operações sobre imagens, tais como, operações sobre cores, filtragem espacial, segmentação, transformações em escala e resolução, dentre outras. Na parte de aplicações, são apresentados exemplos de aplicações envolvendo segmentação de imagens, reconhecimento de palavras manuscritas e recuperação de imagens por conteúdo.

Palavras-chave: processamento digital de imagens, operações sobre imagens, aplicações de processamento de imagens

Abstract: The goal of this tutorial is to provide an introductory view of the Digital Image Processing (IP) area that can be used as a study guide for beginners or as basic reference for more advanced studies. The tutorial is divided into two parts: the main part is about the IP fundamentals and a complementary part discusses some application examples. The main part presents the image formation process, including a succinct description of the human eye structure and its relation to a digital camera, as well as comments about a typical IP system. The core of the tutorial is about image operations, such as color operations, spatial filtering, segmentation, scale and resolution transforms, among others. The applications part contains a number of examples, involving image segmentation, handwritten word recognition and content-based image retrieval.

Keywords: digital image processing, image operations, image processing applications

1 Departamento de Sistemas e Computação, UFCG, Caixa Postal 10106
{rangel,hmg}@dsc.ufcg.edu.br.br

1 Considerações Iniciais

1.1 Entendendo o Processamento Digital de Imagens

O Processamento Digital de Imagens (PDI) não é uma tarefa simples, na realidade envolve um conjunto de tarefas interconectadas (vide Fig. 1). Tudo se inicia com a captura de uma imagem, a qual, normalmente, corresponde à iluminação que é refletida na superfície dos objetos, realizada através e um sistema de aquisição. Após a captura por um processo de digitalização, uma imagem precisa ser representada de forma apropriada para tratamento computacional. Imagens podem ser representadas em duas ou mais dimensões. O primeiro passo efetivo de processamento é comumente conhecido como *pré-processamento* [1][2][3], o qual envolve passos como a filtragem de ruídos introduzidos pelos sensores e a correção de distorções geométricas causadas pelo sensor.

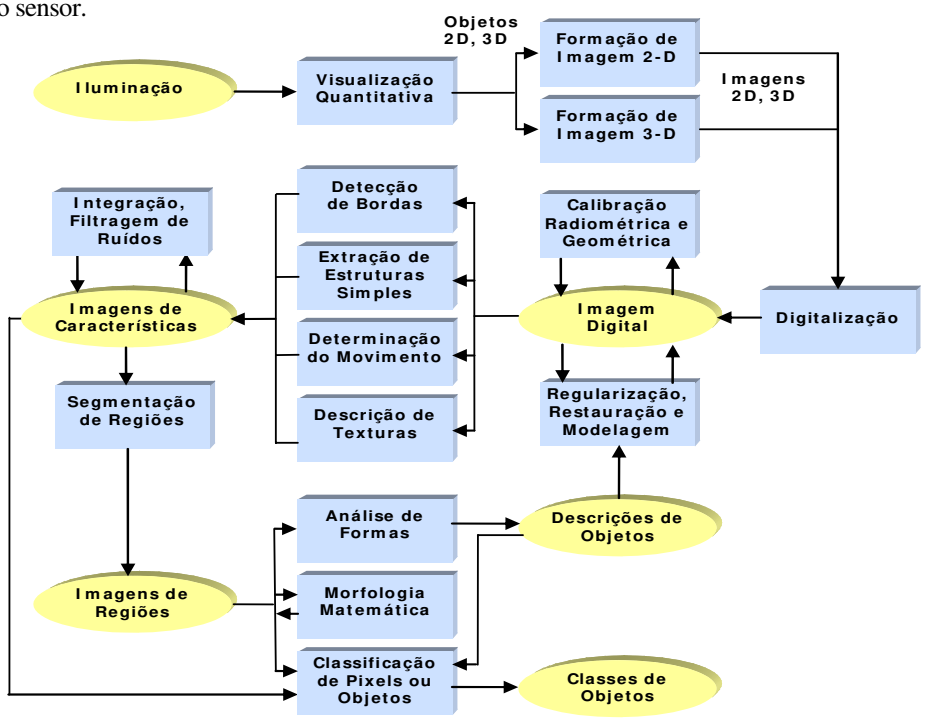


Fig. 1 - Uma hierarquia de tarefas de processamento de imagens (adaptada de [1]).

Uma cadeia maior de processos é necessária para a análise e identificação de objetos. Primeiramente, características ou atributos das imagens precisam ser extraídos, tais como as

bordas, texturas e vizinhanças. Outra característica importante é o movimento. Em seguida, objetos precisam ser separados do plano de fundo (*background*), o que significa que é necessário identificar, através de um processo de segmentação, características constantes e descontinuidades [2]. Esta tarefa pode ser simples, se os objetos são facilmente destacados da imagem de fundo, mas normalmente este não é o caso, sendo necessárias técnicas mais sofisticadas como regularização e modelagem. Essas técnicas usam várias estratégias de otimização para minimizar o desvio entre os dados de imagem e um modelo que incorpora conhecimento sobre os objetos da imagem. Essa mesma abordagem matemática pode ser utilizada para outras tarefas que envolvem restauração e reconstrução [1]. A partir da forma geométrica dos objetos, resultante da segmentação, pode-se utilizar operadores morfológicos [1][2][3] para analisar e modificar essa forma bem como extrair informações adicionais do objeto, as quais podem ser úteis na sua classificação. A classificação é considerada como uma das tarefas de mais alto nível e tem como objetivo reconhecer, verificar ou inferir a identidade dos objetos a partir das características e representações obtidas pelas etapas anteriores do processamento. Como último comentário, deve-se observar que, para problemas mais difíceis, são necessários mecanismos de retro-alimentação (*feedback*) entre as tarefas de modo a ajustar parâmetros como aquisição, iluminação, ponto de observação, para que a classificação se torne possível. Esse tipo de abordagem também é conhecido como *visão ativa* [4][5]. Em um cenário de agentes inteligentes, fala-se de *ciclos de ação-percepção*.

1.2 Relação entre Processamento de Imagens e Computação Gráfica

Em geral, autores de livros em Computação Gráfica (CG) e Processamento de Imagens (PDI) vêm tratando as duas áreas como distintas. O conhecimento em ambas as áreas tem crescido consideravelmente, o que tem permitido a resolução de problemas cada vez mais complexos. Numa visão simplificada, CG busca imagens fotos-realísticas de cenas tridimensionais geradas por computador, enquanto PDI tenta reconstruir uma cena tridimensional a partir de uma imagem real, obtida através de uma câmera. Neste sentido, PDI busca um procedimento inverso ao de CG, análise ao invés de síntese, mas ambas as áreas atuam sobre o mesmo conhecimento, o qual inclui, dentre outros aspectos, a interação entre iluminação e objetos e projeções de uma cena tri-dimensional em um plano de imagem. O cenário envolvendo todas as disciplinas que tenham algum ingrediente de processamento da informação visual, dentre as quais a CG e o PDI ocupam posição de destaque, é definido por alguns autores como Computação Visual.

1.3 Natureza Interdisciplinar do Processamento de Imagens

A área de Processamento de Imagens incorpora fundamentos de várias ciências, como Física, Computação, Matemática. Conceitos como Óptica, Física do Estado Sólido, Projeto de Circuitos, Teoria dos Grafos, Álgebra, Estatística, dentre outros, são comumente requeridos no projeto de um sistema de processamento de imagens. Existe também uma intersecção forte entre PDI e outras disciplinas como Redes Neurais, Inteligência Artificial, Percepção Visual, Ciência Cognitiva. Há igualmente um número

de disciplinas as quais, por razões históricas, se desenvolveram de forma parcialmente independente do PDI, como Fotogrametria, Sensoriamento Remoto usando imagens aéreas e de satélite, Astronomia e Imageamento Médico.

1.4 Organização do Tutorial

O tutorial está estruturado em duas partes: a primeira parte (principal) trata dos fundamentos de PDI e a segunda (complementar) apresenta exemplos de aplicações. As próximas duas seções contemplam a parte de fundamentos, incluindo o processo de formação da imagem e uma seleção de operações típicas sobre imagens. A Seção 4 apresenta alguns exemplos de aplicações. Finalmente, na Seção 5 estão as considerações finais.

2. Conceitos Fundamentais

2.1 Natureza da luz

Sendo radiação eletromagnética, a luz apresenta um comportamento ondulatório caracterizado por sua freqüência (f) e comprimento de onda (λ). A faixa do espectro eletromagnético à qual o sistema visual humano é sensível se estende aproximadamente de 400 a 770 nm e denomina-se *luz visível* [2]. Radiação eletromagnética com comprimentos de onda fora desta faixa não é percebida pelo olho humano. Dentro dessa faixa, o olho percebe comprimentos de onda diferentes como *cores* distintas, sendo que fontes de radiação com um único comprimento de onda denominam-se *monocromáticas* e a cor da radiação denomina-se *cor espectral pura* [3][4]. O *espectro eletromagnético* é a distribuição da intensidade da radiação eletromagnética com relação ao seu comprimento de onda e/ ou freqüência [6]. Na Fig. 2, apresenta-se uma síntese do espectro eletromagnético, destacando-se a faixa de luz visível.

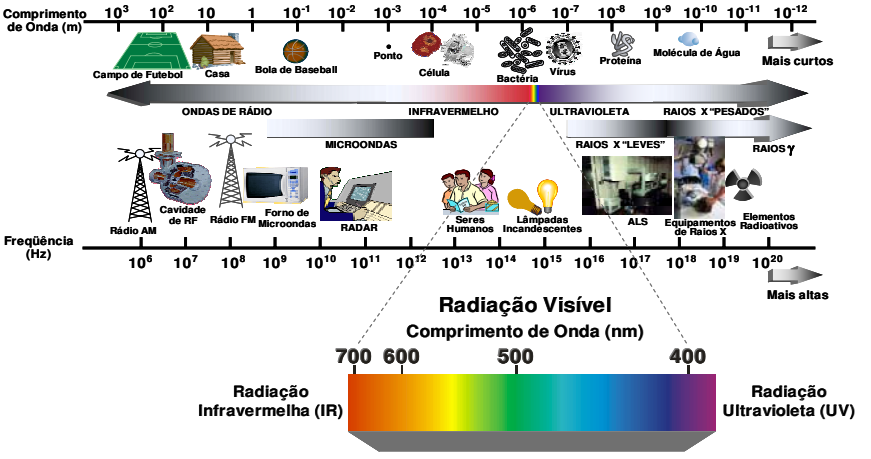


Fig. 2 - Espectro eletromagnético.

2.2 Estrutura do Olho Humano

De conformação aproximadamente esférica, o olho humano possui um diâmetro médio aproximado variando de 2 a 2,5 cm [2][3][4]. A radiação luminosa advinda de objetos do mundo real penetra no olho a partir de uma abertura frontal na íris, denominada *pupila*, e de uma lente denominada *cristalino*, atingindo então a *retina*, que constitui a camada interna posterior do globo ocular [3] (vide Fig. 3).

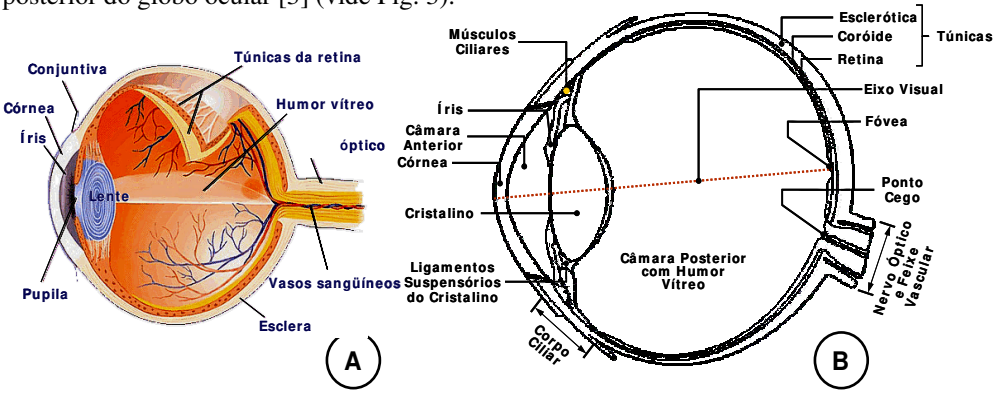


Fig. 3 - Olho humano: (A) visão geral; e (B) detalhamento dos componentes.

A focalização apropriada da cena implica a formação nítida de sua imagem invertida sobre a retina. A retina contém dois tipos de fotossensores, os *cones* (sensíveis a cores e com alta resolução, operantes apenas em cenas suficientemente iluminadas) e os *bastonetes*, (insensíveis a cores, com baixa resolução, operantes em condições de baixa luminosidade), encarregados do processo de conversão da energia luminosa em impulsos elétricos que serão transmitidos ao cérebro, para posterior interpretação. A visualização de um objeto consiste do posicionamento do olho pela estrutura muscular que o controla, implicando a projeção da imagem do objeto sobre a fóvea [3]. Em essência, toda câmara fotográfica é uma câmara escura, projetada para apreender a energia luminosa proveniente de uma cena, produzindo uma imagem adequada para propósitos os mais diversificados. Trata-se de uma extensão do olho humano, o qual compõe imagens a partir de excitação luminosa e as transmite ao cérebro sob a forma de impulsos bioelétricos. A *pálpebra* do olho tem uma função análoga àquela do *obturador* da câmara. O *diafragma* (ou *íris*) de uma câmera funciona analogamente à *íris* do olho humano, controlando a quantidade de luz que atravessa a lente. A *lente* da câmera é análoga ao conjunto formado pelo *cristalino* do olho, a *córnea* e, em menor grau, o *humor aquoso* e o *humor vítreo*. Ambos têm o propósito de focalizar a luz, de modo a tornar nítidas as imagens que se formarão invertidas no plano focal [6]. A diferença é que o cristalino se deforma para focalizar a imagem, enquanto a lente é dotada de um mecanismo manual ou automático para o ajuste da distância focal, à exceção das lentes das câmaras de *foco fixo*, projetadas para dar foco a partir de uma distância mínima (usualmente a partir de 1,5m). A *coróide* funciona como a *câmara escura* de uma câmara fotográfica. A *retina* corresponde ao *sensor* da câmara fotográfica (componente digital ou filme). A Fig. 4 ilustra essa analogia.

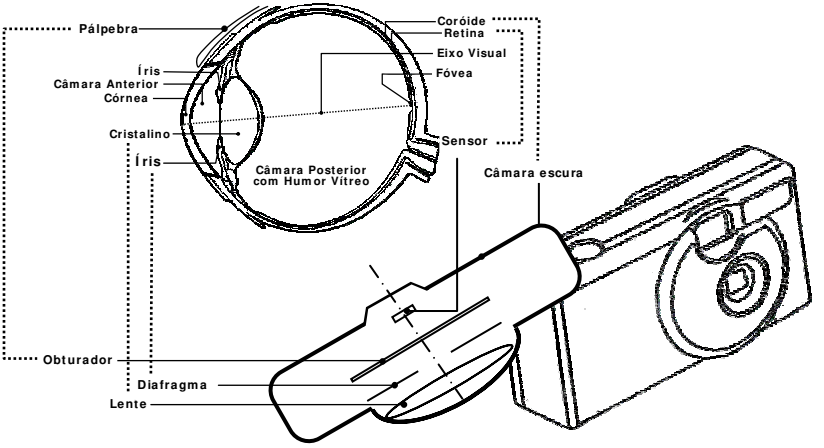
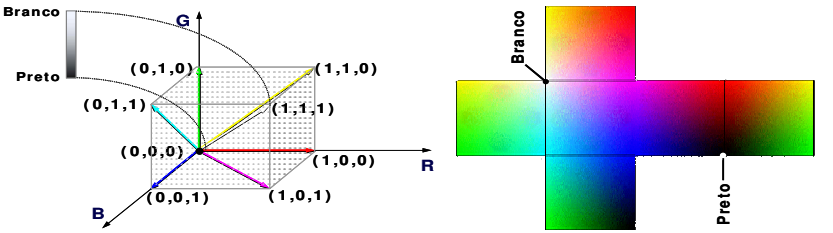


Fig. 4 – Analogia olho humano-câmara digital.

2.3 Modelos Cromáticos

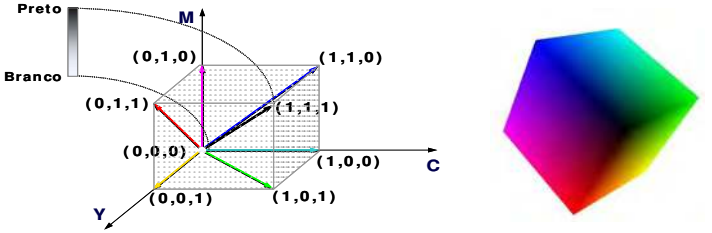
Objetos que emitem luz visível são percebidos em função da soma das cores espectrais emitidas. Tal processo de formação é denominado *aditivo*. O processo aditivo pode ser interpretado como uma combinação variável em proporção de componentes monocromáticas nas faixas espectrais associadas às sensações de cor verde, vermelho e azul, as quais são responsáveis pela formação de todas as demais sensações de cores registradas pelo olho humano. Assim, as cores verde, vermelho e azul são ditas cores *primárias*. Este processo de geração suscitou a concepção de um modelo cromático denominado *RGB* (***R**ed*, ***G**reen*, e ***B**lue*) [2][3], para o qual a *Comissão Internacional de Iluminação* (*CIE*) estabeleceu as faixas de comprimento de onda das cores primárias [7]. A combinação dessas cores, duas a duas e em igual intensidade, produz as cores *secundárias*, *Ciano*, *Magenta* e *Amarelo* (ver Fig. 5).



Primária	λ(nm)	Combinação de Primárias	Secundária Resultante
Vermelho	700,0	Vermelho + Verde	Amarelo
Verde	546,1	Vermelho + Azul	Magenta
Azul	435,8	Verde + Azul	Ciano

Fig. 5 - Modelo cromático RGB.

A cor *oposta* a uma determinada cor secundária é a cor primária que não entra em sua composição. Assim, o *verde* é oposto ao *magenta*, o *vermelho* ao *ciano* e o *azul* ao *amarelo*. A cor *branca* é gerada pela combinação balanceada de *vermelho*, *verde* e *azul*, assim como pela combinação de qualquer cor secundária com sua oposta. Objetos que não emitem radiação eletromagnética visível própria são, em contraposição, percebidos em função dos pigmentos que os compõem [3]. Assim sendo, objetos diferentemente pigmentados absorvem (ou subtraem) da radiação eletromagnética incidente uma faixa do espectro visível, refletindo o restante [6]. O processo de composição cromática pode ser interpretado como a absorção ou reflexão, em proporções variáveis, das componentes verde, vermelho e azul da radiação eletromagnética visível incidente. Tome-se como exemplo um objeto amarelo. As componentes vermelha e verde da luz branca incidente são refletidas, enquanto a componente azul é subtraída por absorção pelo objeto. Assim, a cor amarela pode ser encarada como o resultado da subtração do azul da cor branca. As cores primárias no modelo CMY são definidas em função da absorção de uma cor primária da luz branca incidente e da reflexão das demais componentes, ou seja, as cores primárias são as secundárias do modelo RGB - *Ciano*, *Magenta* e *Amarelo* (Fig. 6).



Combinação de Primárias	Secundária Resultante
Ciano + Magenta	Azul
Magenta + Amarelo	Vermelho
Amarelo + Ciano	Verde

Fig. 6 - Modelo cromático CMY.

A formação de imagens em um terminal de vídeo se dá por emissão de radiação eletromagnética visível, em um processo que integra, em diferentes proporções, as cores *verde*, *vermelha* e *azul*. Já os dispositivos de impressão coloridos (e.g. impressoras e traçadores gráficos) adotam o sistema *CMY* (***C****yan*, ***M****agenta*, ***Y****ellow*). Uma vez que os pigmentos empregados (tintas em cartuchos ou *toners*) não produzem o preto quando combinados de modo balanceado, é necessário acrescentá-lo como um quarto pigmento, o novo sistema cromático é denominado *CMYK* (***C****yan*, ***M****agenta*, ***Y****ellow*, *blac****K***). Há vários outros modelos cromáticos nos quais a caracterização da cor não se dá conforme o comportamento fisiológico da retina humana, mas sim em função de outros atributos de percepção cromática empregados por seres humanos [2][3][6]. Ao invés da caracterização da cor a partir de combinações de vermelho, verde e azul, tais modelos adotam outros atributos, tais como a *intensidade*, o *matiz* ou *tonalidade* (*hue*) e a *saturação* ou *pureza*.

2.4 Modelo de Imagem Digital

Uma imagem monocromática é uma função bidimensional contínua $f(x,y)$, na qual x e y são coordenadas espaciais e o valor de f em qualquer ponto (x,y) é proporcional à intensidade luminosa (brilho ou nível de cinza) no ponto considerado [1][2][4][6][8]. Como os computadores não são capazes de processar imagens contínuas, mas apenas *arrays* de números digitais, é necessário representar imagens como arranjos bidimensionais de pontos.

Cada ponto na grade bidimensional que representa a imagem digital é denominado *elemento de imagem* ou *pixel*. Na Fig. 7, apresenta-se a notação matricial usual para a localização de um pixel no arranjo de pixels de uma imagem bidimensional. O primeiro índice denota a posição da linha, m, na qual o pixel se encontra, enquanto o segundo, n, denota a posição da coluna. Se a imagem digital contiver M linhas e N colunas, o índice m variará de 0 a M-1, enquanto n variará de 0 a N-1. Observe-se o sentido de leitura (varredura) e a convenção usualmente adotada na representação espacial de uma imagem digital.

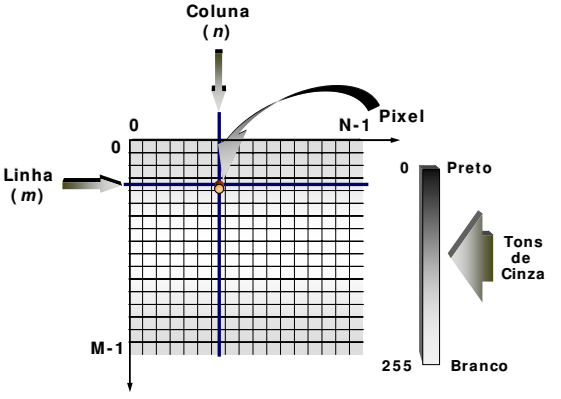


Fig. 7 – Representação de uma imagem digital bidimensional.

A intensidade luminosa no ponto (x,y) pode ser decomposta em: (i) componente de *iluminação*, $i(x,y)$, associada à quantidade de luz incidente sobre o ponto (x,y); e a componente de reflectância, $r(x,y)$, associada à quantidade de luz refletida pelo ponto (x,y) [3]. O produto de $i(x,y)$ e $r(x,y)$ resulta em:

$$f(x,y) = i(x,y).r(x,y) \tag{1}$$

na qual $0 < i(x,y) < \infty$ e $0 < r(x,y) < 1$, sendo $i(x,y)$ dependente das características da fonte de iluminação, enquanto $r(x,y)$ dependente das características das superfícies dos objetos.

Em uma imagem digital colorida no sistema RGB, um pixel pode ser visto como um vetor cujas componentes representam as intensidades de vermelho, verde e azul de sua cor. A imagem colorida pode ser vista como a composição de três imagens monocromáticas, i.e.:

$$f(x, y) = f_R(x,y) + f_G(x,y) + f_B(x,y), \tag{2}$$

na qual $f_R(x,y)$, $f_G(x,y)$, $f_B(x,y)$ representam, respectivamente, as intensidades luminosas das componentes vermelha, verde e azul da imagem, no ponto (x,y).

Na Fig. 8, são apresentados os planos monocromáticos de uma imagem e o resultado da composição dos três planos. Os mesmos conceitos formulados para uma imagem digital monocromática aplicam-se a cada plano de uma imagem colorida [3][6][8].

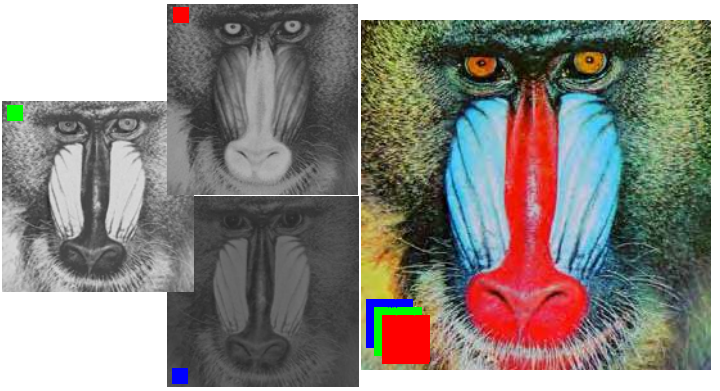


Fig. 8 – Representação de uma imagem digital bidimensional.

2.5 Amostragem e Quantização

Como já foi anteriormente mencionado, para que uma imagem possa ser armazenada e/ ou processada em um computador, torna-se necessária sua discretização tanto em nível de coordenadas espaciais quanto de valores de brilho. O processo de discretização das coordenadas espaciais denomina-se *amostragem*, enquanto a discretização dos valores de brilho denomina-se *quantização* [1][2][3][4][5][6]. Usualmente, ambos os processos são uniformes, o que implica a amostragem da imagem $f(x,y)$ em pontos igualmente espaçados, distribuídos na forma de uma matriz $M \times N$, na qual cada elemento é uma aproximação do nível de cinza da imagem no ponto amostrado para um valor no conjunto $\{0, 1, ..., L - 1\}$.

$$F \approx \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(0,M) & f(1,M) & \cdots & f(M-1,N-1) \end{bmatrix} \tag{3}$$

Costuma-se associar o limite inferior (*0*) da faixa de níveis de cinza de um pixel ao *preto* e ao limite superior (*L–1*) ao *branco*. Pixels com valores entre *0* e *L-1* serão visualizados em diferentes tons de *cinza*, os quais serão tão mais escuros quanto mais próximo de zero forem seus valores [1][2][3].

Uma vez que os processos de amostragem e quantização implicam a supressão de

informação de uma imagem analógica, seu equivalente digital é uma aproximação, cuja qualidade depende essencialmente dos valores de *M*, *N* e *L*. Usualmente, o número de valores de brilho, *L*, é associado a potências de 2:

$$L = 2^l \tag{4}$$

com $l \in \aleph$. Assim sendo, o número de bits necessário para representar uma imagem digital de dimensões *M x N* será:

$$b = M \text{ x } N \text{ x } l \tag{5}$$

Percebe-se, a partir da Eq. 5, que embora o aumento de *M, N* e *l* implique a elevação da qualidade da imagem, isto também implica o aumento do número de bits necessários para a codificação binária da imagem e, por conseguinte, o aumento do volume de dados a serem armazenados, processados e/ou transmitidos. O Quadro 1 contém o número de bytes empregado na representação de uma imagem digital monocromática para alguns valores típicos de *M* e *N*, com 2, 5 e 8 níveis de cinza.

Quadro 1 – Número de *bytes* para uma imagem monocromática.

M	*N*	Número de *Bytes* (*L*)		
		L = 2	*L* = 32	*L* = 256
480	640	38400	192000	307200
600	800	60000	300000	480000
768	1024	98304	491520	786432
1200	1600	240000	1200000	1920000

O número de amostras e o número de níveis de cinza necessários para a representação de uma imagem digital de qualidade adequada é função tanto de características da imagem, tais como suas dimensões e a complexidade dos alvos nela contidos, quanto da aplicação à qual se destina. Nas Figs. 9(A) a (D), ilustra-se a influência dos parâmetros de digitalização na qualidade visual de uma imagem monocromática.

Fig. 9 – Influência da variação do número de amostras e de níveis de quantização na qualidade de uma imagem digital: (A) 200 x 200 pixels/ 256 níveis; (B) 100 x 100 pixels/ 256 níveis; (C) 25 x 25 pixels/ 256 níveis; e (D) 200 x 200 pixels/ 2 níveis.

Em geral, costuma-se amostrar de forma idêntica os diferentes planos de uma imagem colorida [1][2]. O número de cores que um pixel pode assumir em uma imagem RGB com L_R níveis de quantização no plano R, L_G no plano G e L_B no plano B é L_R x L_G x L_G. Considerando a Eq. (6), se $l_R = log_2(L_R)$, $l_G = log_2 (L_G)$e $l_B = log_2 (L_B)$, o número de bits por pixel necessários para representar as cores será igual a $l_R + l_G + l_B$ e o número de bits necessário para representar uma imagem digital de dimensões M x N será:

$$b = M \text{ x } N \text{ x } (l_R + l_G + l_B) \tag{6}$$

Seja, por exemplo, $L_R = L_G = L_G = 2^8 = 256$ níveis de cinza possíveis em cada banda. Assim sendo, cada pixel da imagem colorida poderá assumir uma das 16.777.216 cores da paleta, uma vez que será representado por 3 x 8 = 24 bits. O Quadro 2 contém o número de bytes empregado na representação de uma imagem digital colorida para alguns valores típicos de M e N, com 2, 5 e 8 níveis de cinza.

Quadro 2 – Número de *bytes* para uma imagem colorida.

M	N	Número de *Bytes* ($L_R = L_G = L_B = L$)		
		L = 2	L = 32	L = 256
480	640	115200	576000	921600
600	800	180000	900000	1440000
768	1024	294912	1474560	2359296
1200	1600	720000	3600000	5760000

2.6 Sistema Típico para Processamento Digital de Imagens

Vários modelos de sistemas para processamento de imagens têm sido propostos e comercializados no mundo inteiro nas duas últimas décadas. Entre meados das décadas de 80 e 90, com a progressiva redução nos custos das tecnologias de *hardware*, as tendências de mercado voltaram-se para placas projetadas, segundo padrões industriais, para uso em computadores pessoais e estações de trabalho [3]. Assim, surgiram diversas empresas que se especializaram no desenvolvimento de *software* dedicado ao processamento de imagens. Nos dias atuais, o extenso uso dos sistemas para processamento de imagens desta natureza ainda é um fato, sobretudo em aplicações de sensoriamento remoto (processamento de produtos aerofotogramétricos e orbitais) [7] e imageamento biomédico (processamento de imagens geradas a partir de MR, CT, PET/ SPEC, tomografia óptica, ultra-sonografia e raios X) [8]. Todavia, tendências recentes apontam para a miniaturização e integração do *hardware* especializado para processamento de imagens a computadores de pequeno porte de uso geral.

A representação do *hardware* e o diagrama em blocos da Fig. 10 ilustram os componentes de um sistema de uso geral tipicamente utilizado para o processamento digital de imagens. O papel de cada componente será discutido, em linhas gerais, a seguir. No tocante à **aquisição** (também referida como **sensoriamento**) de imagens digitais, afiguram-se relevantes dois elementos, a saber: (i) o dispositivo físico sensível à faixa de energia irradiada pelo alvo de interesse; e (ii) o dispositivo conversor da saída do o dispositivo físico de

sensoriamento em um formato digital (usualmente referido como *digitalizador*) [1][2][3]. Tome-se como exemplo uma câmara de vídeo digital. Os sensores CCD são expostos à luz refletida pelo alvo de interesse, o feixe de radiação eletromagnética capturada é convertido em impulsos elétricos proporcionais à intensidade luminosa incidente nos diferentes pontos da superfície do sensor e, finalmente, o digitalizador converte os impulsos elétricos em dados digitais.

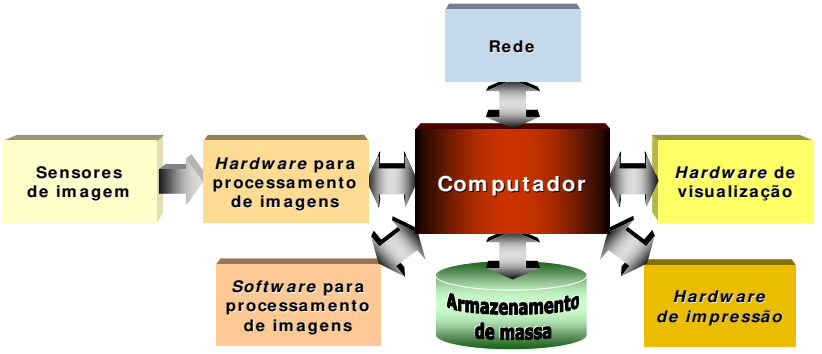


Fig. 10 - Diagrama em blocos de um sistema típico para processamento de imagens.

Em geral, o *hardware especializado* para processamento de imagens consiste de um digitalizador integrado a um *hardware* destinado à execução de outras operações primitivas, e.g. uma unidade lógico-aritmética (ULA) para a realização de operações aritméticas e lógicas em imagens inteiras, à medida que são digitalizadas. O diferencial do *hardware* desta natureza, também denominado *subsistema front-end*, é a velocidade de processamento em operações que requerem transferências rápidas de dados da entrada para a saída, e.g., digitalização e remoção de ruído em sinais de vídeo capturados a uma taxa de 30 quadros/s, tarefa que um computador típico não consegue realizar com o mesmo desempenho.

Em nível do **processamento** propriamente dito, o *computador* em um sistema para processamento de imagens é um *hardware* de uso geral que pode ser desde um PDA até um supercomputador, em função da capacidade de processamento exigida pela tarefa. Embora aplicações dedicadas possam requerer computadores especialmente projetados e configurados para atingir o grau de desempenho exigido pela tarefa de interesse, os sistemas de uso geral para processamento de imagens utilizam computadores pessoais típicos para a execução de tarefas *offline* [3].

O **armazenamento** é um dos grandes desafios para a área de processamento de imagens, uma vez que os sistemas de aquisição vêm sendo cada vez mais aprimorados para a captura de volumes de dados cada vez maiores, o que requer dispositivos com capacidades de armazenamento cada vez maiores, além de taxas de transferência de dados mais elevadas e maiores índices robustez e confiabilidade do processo de armazenamento. Costuma-se discriminar a etapa de armazenamento em três níveis, a saber: (i) **armazenamento de curta duração** (memória RAM), durante o uso temporário das imagens de interesse em diferentes etapas de processamento; (ii) **armazenamento online** ou **de massa**, típico em operações

relativamente rápidas de recuperação de imagens; e (iii) **arquivamento de imagens**, com fins ao acesso infreqüente e à recuperação quando o uso se fizer necessário [2][3][8].

No âmbito da saída do sistema de processamento de imagens, são típicas duas alternativas, a saber: (i) a **visualização** de dados; e (ii) a **impressão** de dados. A **visualização** requer tipicamente monitores de vídeo coloridos e preferencialmente de tela plana, que recebem dados de placas gráficas comerciais ou dedicadas [2][3][8]. Há circunstâncias em que se torna necessário o uso de visualizadores estéreo, e.g. em aplicações que lidam com pares estereoscópicos de produtos aerofotogramétricos [6]. No tocante à **impressão**, costuma-se utilizar diferentes dispositivos de impressão de pequeno, médio e grande porte - impressoras e/ou traçadores gráficos (*plotters*) de jato de tinta, sublimação de cera ou laser [2][3][8]. Costuma-se também incluir nesta etapa a geração de produtos em filme, que oferecem a mais alta resolução possível [6].

O *software* para processamento de imagens consiste, em geral, de módulos destinados à realização de tarefas específicas (e.g. operações de processamento radiométrico e/ou geométrico de imagens monocromáticas ou coloridas, mono ou multiespectrais). Há pacotes que incluem facilidades de integração de módulos e geração de código em uma ou mais linguagens de programação. Por fim, faz-se pertinente comentar que a conexão em rede de sistemas para processamento de imagens parece ser uma função típica nos dias atuais, uma vez que, para diversas aplicações, se faz necessária a transmissão de grandes volumes de dados. Para tais aplicações, a consideração mais relevante é a largura de faixa, uma vez que a comunicação com sites remotos via Internet pode constituir um obstáculo para a transferência eficiente de dados de imagens.

3 Operações sobre Imagens

3.1 Operações no Domínio do Espaço

As operações no domínio do espaço são caracterizadas pela manipulação direta dos pixels da imagem [1][2][3][4][5][6][8]. Pode-se representar uma operação genérica O sobre uma seqüência de n imagens, $fe_1, fe_2, ..., fe_n$ (vide Fig. 13(A), produzindo uma imagem de saída fs, i.e.:

$$fs = O(fe_1, fe_2, ..., fe_n) \tag{7}$$

Operações desta natureza são denominadas *n-árias*, uma vez que a imagem de saída resulta de uma combinação de duas ou mais imagens de entrada. Quando $n = 1$, uma operação *unária*, a partir da qual uma única imagem de entrada produz uma imagem de saída (vide Fig. 11(B)), sendo representada de forma simplificada como:

$$fs = O(fe) \tag{8}$$

As operações no domínio do espaço podem ser classificadas, no tocante ao escopo de ação, como *pontuais* (ou *ponto-a-ponto*) ou *locais* (ou *localizadas*). Nas operações *pontuais*, cada pixel da imagem de saída depende apenas do mesmo correspondente na imagem de entrada. Assim, qualquer operação pontual pode ser interpretada como um mapeamento de pixels da

imagem de entrada para a imagem de saída. A Fig. 12 ilustra genericamente uma operação pontual unária.

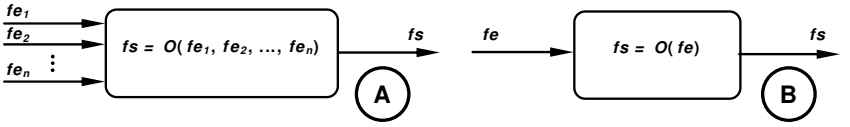


Fig. 11 – Operações no domínio do espaço: (A) *m-árias*; e (B) *unárias*.

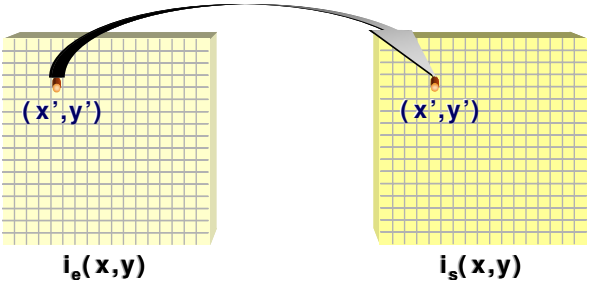


Fig. 12 – Operação pontual unária.

Cada ponto da imagem de saída, $fs(x,y)$, é obtido por: (i) uma operação O entre os pontos de coordenadas homólogas das imagens de entrada, $fe_1(x,y)$, $fe_2(x,y)$,..., $fe_n(x,y)$; ou (ii) uma transformação T do ponto de coordenadas homólogas da imagem de entrada, $fe(x,y)$. No tocante à operação O, esta pode ser qualquer operação aritmética, lógica, de comparação, etc., admitida pela natureza dos valores dos pontos das imagens. A transformação T deverá ser uma função unívoca com um domínio equivalente à faixa de valores permitidos para a imagem de entrada. Transformações dessa natureza são comumente realizadas a partir de tabelas de transformação (*LUT - Look-Up Tables*) e interpretadas a partir de diagramas como aquele ilustrado na Fig. 13.

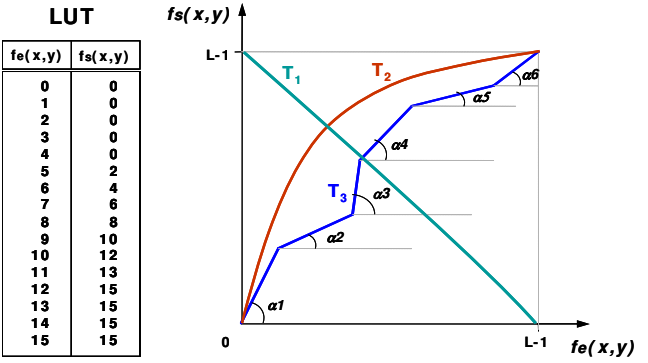


Fig. 13 – Exemplo de LUT e diagramas de transformação.

Por outro lado, nas operações *locais*, o valor de saída em uma coordenada específica depende de valores de entrada daquela coordenada e sua vizinhança [1][2][3][4]. Os tipos de vizinhos de um pixel podem ser assim definidos: (i) os vizinhos *mais próximos* de um pixel p, de coordenadas (i,j), os pixels de coordenadas $(i+1,j)$, $(i-1,j)$, $(i,j+1)$ e $(i,j-1)$; (ii) os vizinhos *mais distantes*, os pixels de coordenadas $(i-1,j-1)$, $(i-1, j+1)$, $(i+1,j-1)$ e $(i+1,j+1)$. As vizinhanças tipicamente utilizadas em operações locais estão na Fig. 14. A vizinhança *4-conectada* envolve os vizinhos mais próximos do pixel considerado, enquanto a vizinhança *8-conectada* envolve tanto os vizinhos mais próximos quanto os mais distantes do pixel considerado. É conveniente mencionar é possível processar grades de pixels hexagonais, é, que neste caso, operações locais envolverão apenas os 6 vizinhos mais próximos (vizinhança *6-conectada*).

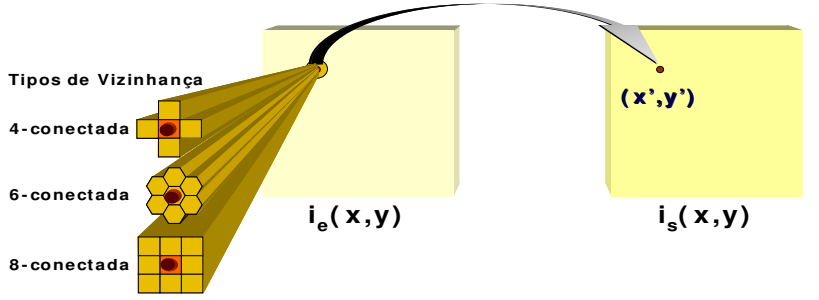


Fig. 14 – Exemplo de LUT e diagramas de transformação.

Nas subseções seguintes, são apresentadas algumas operações pontuais e locais tipicamente conduzidas no domínio do espaço.

3.2 Modificação Histogrâmica

O realce de contraste visa o melhoramento da qualidade das imagens sob o ponto de vista subjetivo do olho humano, sendo usualmente empregada como uma etapa de pré-processamento em aplicações de reconhecimento de padrões [1][2][3][6]. O *contraste* entre dois alvos de uma cena pode ser definido como a razão entre os seus níveis de cinza médios. Fundamentada neste conceito, a manipulação do contraste dos objetos presentes em uma imagem digital consiste em um remapeamento radiométrico de cada pixel da imagem, a fim de aumentar a discriminação visual entre eles. Embora a escolha do mapeamento adequado seja, em princípio, essencialmente empírica, uma análise prévia do histograma da imagem se afigura, em muitos casos, bastante útil.

O *histograma* de uma imagem traduz a distribuição estatística dos seus níveis de cinza. Trata-se, pois, de uma representação gráfica do número de *pixels* associado a cada nível de cinza presente em uma imagem, podendo também ser expressa em termos do percentual do número total de pixels na imagem [3][6]. Assim sendo, dada uma imagem digital $f(x,y)$ com M linhas e N colunas, seu histograma, $H_f(C)$, pode ser definido por:

$$H_f(C) = n_C/M.N, \tag{9}$$

sendo n_C o número de vezes em que o nível de cinza C se apresenta na imagem. A Fig. 15 ilustra alguns exemplos de histogramas.

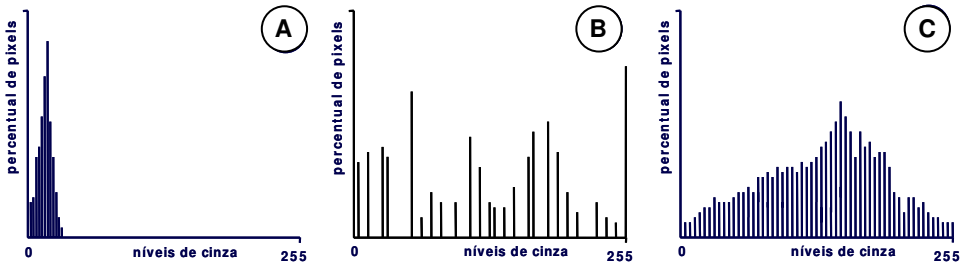


Fig. 15 – Histogramas: (A) imagem com baixo contraste; (B) imagem usando toda a faixa de tons de cinza, com dois tons de cinza dominantes; e (C) imagem usando toda a faixa de tons de cinza, com componentes ocupando a faixa de modo mais eqüidistante.

Muitas operações pontuais usam o histograma como parâmetro de decisão para fornecer resultados para o pixel da imagem processada, como se pode ver nas subseções a seguir.

3.2.1 Inversão da Escala de Cinza

A inversão da escala de cinza de uma imagem pode ter diversas aplicações. Uma delas é que, em se tratando do negativo da imagem, após o registro fotográfico a partir de uma câmera convencional, a revelação do negativo do filme produzirá uma imagem positiva, passível de uso como *slide*. Adicionalmente, o negativo de uma imagem pode possibilitar melhor discriminação de alvos em determinados tipos de imagens (e.g. imagens médicas). Na Fig. 16(A), representa-se o efeito da inversão de contraste sobre o histograma, enquanto que na Fig. 16(B), um exemplo de resultado do processo.

Fig. 16 – Inversão de contraste: (A) representação gráfica do processo; e (B) exemplo.

3.2.2 Expansão de Contraste

Iluminação deficiente no instante da aquisição da imagem, abertura insuficiente do diafragma da câmera, tempo de exposição demasiadamente curto ou problemas de natureza diversa no processo de digitalização são responsáveis pela geração de imagens de baixo contraste [3]. A redução no contraste de uma cena dificulta o discernimento de seus componentes. O propósito da expansão de contraste é redistribuir os tons de cinza dos pixels de uma imagem de modo a elevar o contraste na faixa de níveis possível (vide Fig. 17(A)). Nos casos em que a faixa de tons de cinza já se encontra totalmente utilizada, a expansão de contraste por partes, linear ou não, possibilita melhor discriminação da porção realçada da imagem, conforme ilustrado na Fig. 17(B).

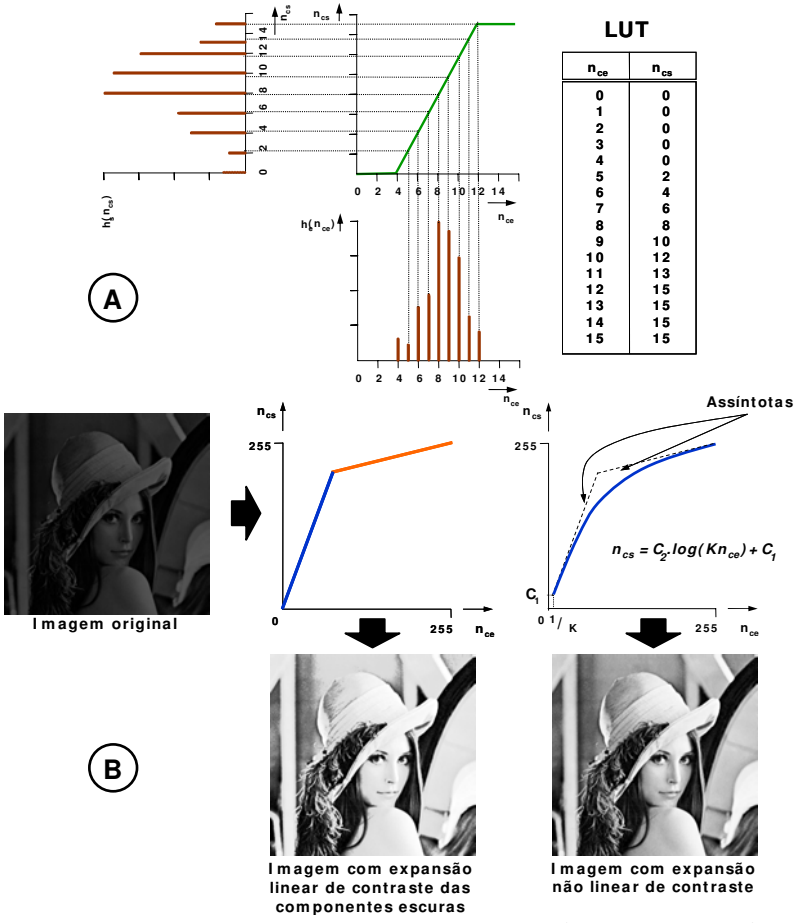


Fig. 17 – Expansão de contraste: (A) representação gráfica do processo típico; (B) Exemplos de expansão de contraste linear por partes e não linear.

3.2.3 Equalização Histogrâmica

O processo de *equalização de histograma* visa o aumento da uniformidade da distribuição de níveis de cinza de uma imagem, sendo usualmente empregado para realçar diferenças de tonalidade na imagem e resultando, em diversas aplicações, em um aumento significativo no nível de detalhes perceptíveis [1][2][3][6][8]. Um modo simples de equalizar o histograma de uma imagem de dimensões M x N com L níveis de cinza advém da transformação:

$$T(n_{ce}) = rnd[((L-1)/M.N).H_f(n_{ce})], \quad (10)$$

na qual *rnd* representa o arredondamento do resultado da expressão para o inteiro mais próximo. Na Fig. 18, exemplifica-se processo da equalização histogrâmica.

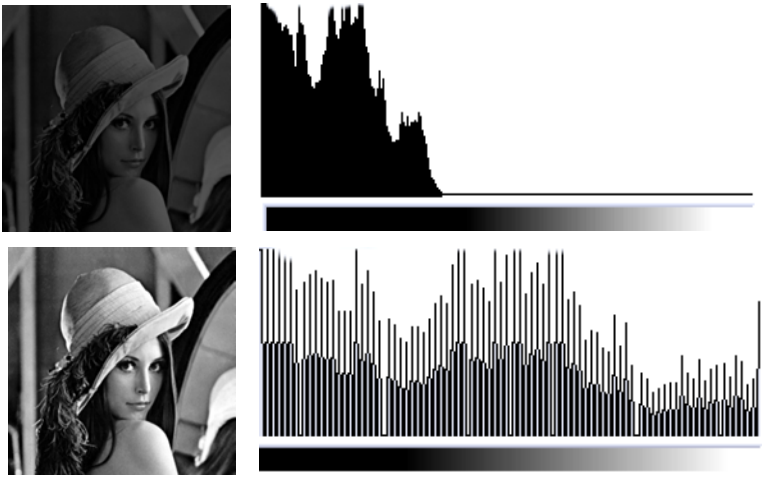


Fig.18 – Exemplo de equalização histogrâmica.

3.3 Filtragem Espacial

Imagens apresentam áreas com diferentes respostas espectrais, delimitadas por áreas geralmente estreitas denominadas *bordas*. Tais limites usualmente ocorrem entre objetos ou feições distintas presentes na imagem (e.g. regiões de um rosto, feições naturais ou artificiais em imagens multiespectrais da superfície terrestre, estruturas de um corpo em imagens médicas), podendo também representar o contato entre áreas com diferentes condições de iluminação, em função dos ângulos formados entre a radiação incidente e os planos da cena imageada. Assim sendo, as bordas representam, em imagens monocromáticas, alterações bruscas entre intervalos de níveis de cinza [3]. Sua representação gráfica é caracterizada por gradientes acentuados. Correspondem usualmente a feições de alta freqüência - limites entre áreas iluminadas e sombreadas, redes naturais (e.g. drenagem) e artificiais (e.g. de transporte), dentre outras. Em contraponto, os alvos que variam mais uniformemente com a

distância apresentam-se, em geral, sob a forma de regiões homogêneas, correspondendo a feições de baixa freqüência (áreas uniformes em imagens).

Similarmente as técnicas de manipulação de contraste, as técnicas de filtragem de uma imagem implicam transformações pixel a pixel. Todavia, diferem daquelas à medida que a alteração efetuada em um pixel da imagem filtrada depende não apenas do nível de cinza do pixel correspondente na imagem original, mas também dos valores dos níveis de cinza dos pixels situados em sua vizinhança. Sendo uma operação local, a filtragem espacial é uma transformação dependente do contexto em que se insere cada pixel considerado.

A filtragem espacial se fundamenta em uma operação de convolução de uma *máscara* (*mask*, *kernel* ou *template*) e da imagem digital considerada. A máscara é um arranjo matricial de dimensões inferiores às da imagem a ser filtrada e, em geral, quadrado, cujos valores são definidos como fatores de ponderação (pesos) a serem aplicados sobre pixels da imagem. A operação é executada progressivamente sobre os pixels da imagem, coluna a coluna, linha a linha, como ilustrado na Fig. 19.

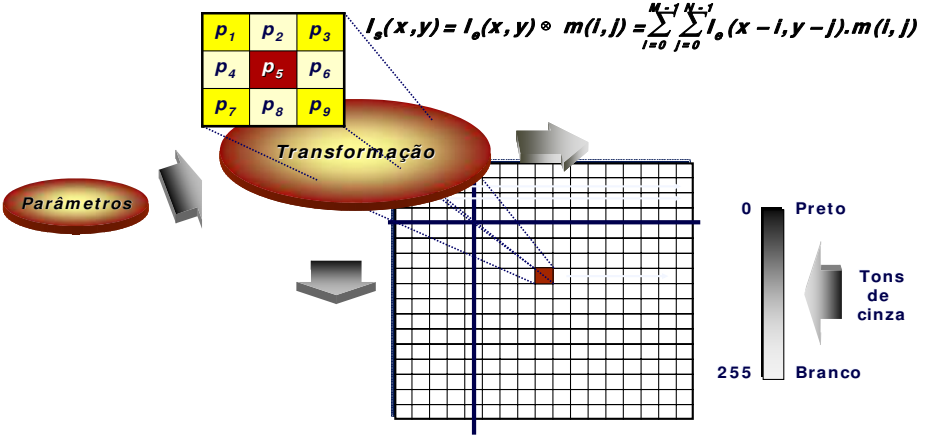


Fig. 19 – Representação gráfica do processo de filtragem espacial.

Dentre os filtros mais comuns utilizados em processamento digital de imagens encontram-se os da média, da mediana e da moda [2], todos destinados à suavização da imagem. Esses filtros atenuam variações abruptas nos níveis de cinza da imagem, o que possibilita sua aplicação à redução de ruído de origens diversas.

O filtro da *média de ordem n* produz como valor do pixel processado, a cada iteração da convolução da máscara de filtragem com a matriz de imagem original, a média aritmética dos valores dos pixels em uma vizinhança de (*i*, *j*) contendo *n* pixels. Assim sendo, a suavização produzida é função do tamanho da vizinhança considerada: quanto maiores as dimensões da máscara utilizada, mais forte será a suavização das bordas das regiões na imagem filtrada. Na Fig. 20, ilustra-se o efeito de filtragem de uma imagem ruidosa com máscaras 3x3 e 5x5.

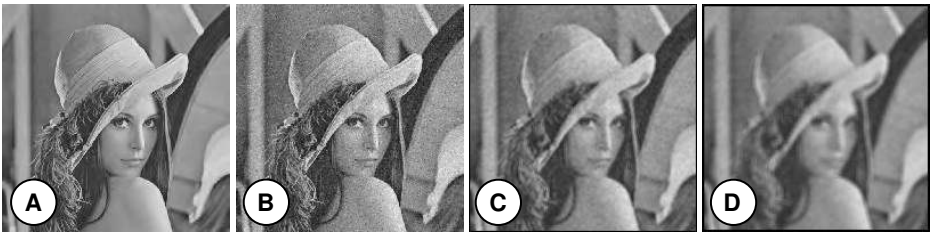


Fig. 20 – Filtro da média: (A) imagem original; (B) imagem ruidosa; (C) imagem filtrada com máscara 3x3; e (D) imagem filtrada com máscara 5x5.

Analogamente à operação da média aritmética no filtro da média, o filtro da *mediana de ordem n* produz como valor do pixel de saída a mediana dos valores dos pixels da imagem de entrada em uma vizinhança de (*i*, *j*) contendo *n* pixels. Vale ressaltar que a *mediana* de um conjunto de *n* pixels ordenados por valor é o valor do pixel na posição central da lista ordenada, se *n* for ímpar, ou a média dos valores dos dois pixels nas posições centrais da lista, se *n* for par. No caso de uma vizinhança 3 x 3 com os valores 21, 22, 17, 21, 19, 17, 21, 20, 23, após a ordenação a seqüência será 17, 17, 19, 20, 21, 21, 21, 22, 23. Deste modo, a mediana será o valor central da seqüência ordenada, i.e., o valor do quinto elemento da lista, *21*. Embora o filtro da mediana também tenda a produzir uma suavização proporcional ao tamanho da vizinhança considerada, a preservação da definição das bordas das regiões na imagem filtrada tende a ser superior do que no filtro da média. Na Fig. 21, ilustra-se uma comparação dos efeitos produzidos pelo filtro da média 3x3 (Fig. 21(C)) e da mediana 3x3 (Fig. 21(D)) sobre uma imagem ruidosa (Fig. 21(B)).

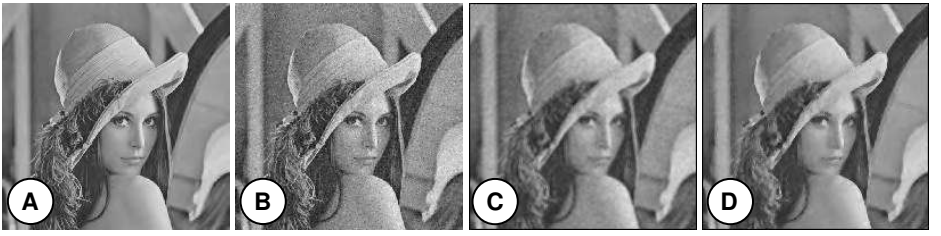


Fig. 21 – Filtros da média e mediana: (A) imagem original; (B) imagem ruidosa; (C) média 3x3; e (D) mediana 3x3.

O filtro da *moda de ordem n* produz como valor do pixel de saída a moda dos valores dos pixels da imagem de entrada em uma vizinhança de (*i*, *j*) contendo *n* pixels (a *moda* de uma série de valores é o valor mais freqüente da série). Se a seqüência contiver dois ou mais valores com a mesma freqüência de ocorrência, pode-se definir a média ou mediana dos valores em questão como valor de $g(i, j)$.

Enquanto os filtros da média, da moda e da mediana são empregados na suavização de imagens, outra categoria de filtros espaciais, tais como os operadores de gradiente [3], produzem a acentuação ou aguçamento de regiões de uma imagem nas quais ocorrem variações significativas de níveis de cinza. Define-se como *gradiente* de uma função *f*,

contínua em (i, j), o vetor:

$$G[f(i,j)] = \begin{bmatrix} \frac{\partial f}{\partial i} \\ \frac{\partial f}{\partial j} \end{bmatrix} \tag{11}$$

O vetor $\boldsymbol{G}[f(i, j)]$ aponta no sentido da maior taxa de variação de $f(i, j)$, sendo sua amplitude, $G[f(i, j)]$, dada pela expressão:

$$G[f(i,j)] = \left[\left(\frac{\partial f}{\partial i} \right)^2 + \left(\frac{\partial f}{\partial j} \right)^2 \right]^{\frac{1}{2}} \tag{12}$$

que é uma representação da taxa de variação de $f(i, j)$ por unidade de distância no sentido de **G**. A equação (11) embasa uma série de abordagens de diferenciação de imagens digitais. Uma propriedade importante da amplitude do gradiente é a sua isotropia, i.e., a independência em relação à direção do gradiente, o que possibilita a detecção de bordas independentemente da sua orientação. As desvantagens apresentadas por este operador são ser não-linear e perder a informação da direção das bordas (devido ao cálculo dos quadrados).

O cálculo do gradiente pode ser obtido através de aproximações numéricas. Na horizontal, a aproximação é dada pela diferença dos níveis de cinza de dois pixels consecutivos, i.e., $G_x = f(i, j) - f(i+1, j)$ e, similarmente, na vertical por $G_y = f(i,j+1) - f(i, j)$. A estimação do gradiente a partir de aproximações numéricas apresenta como desvantagem o cálculo da derivada horizontal e a vertical em pontos diferentes:

$$G_x = \begin{bmatrix} -1 & 1 \end{bmatrix} \quad \text{e} \quad G_y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \tag{13}$$

o que pode ser contornado a partir da utilização de janelas quadradas:

$$G_x = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{e} \quad G_y = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \tag{14}$$

Pode-se obter a 2ª derivada a partir do Laplaciano dos níveis de cinza da imagem $f(x,y)$:

$$\nabla^2 f = \frac{\partial^2 f}{\partial i^2} + \frac{\partial^2 f}{\partial j^2} \tag{15}$$

Fig. 22 – Verificação da existência de uma borda a partir do gradiente e do Laplaciano.

Além da isotropia, a 2ª derivada possibilita a preservação da informação de qual o lado mais claro/escuro da borda. Contrariamente ao gradiente, cujas amplitudes elevadas traduzem a existência de bordas, no Laplaciano são os cruzamentos por zero (alternância de sinal entre pixels adjacentes) que o fazem (vide Fig. 22).

No espaço 2-D, as aproximações numéricas resultam na seguinte janela de convolução:

$$\nabla^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{16}$$

Embora haja uma grande variedade de operadores de gradiente, serão mencionados aqui apenas os operadores de Roberts, Prewitt e Sobel. O *operador de Roberts* (*2 x 2*) executa o gradiente cruzado, i.e.,o cálculo das diferenças dos níveis de cinza é executado em uma direção rotacionada de 45º, ao invés do cálculo nas direções horizontal e vertical.

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{e} \quad G_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{17}$$

Além da diferenciação, sem o enviesamento do gradiente digital, o *operador de Prewitt* suaviza a imagem, atenuando o ruído.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad Gy = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \tag{18}$$

Similar ao operador de Prewitt, o *operador de Sobel* é difere apenas no tocante aos pesos conferidos aos vizinhos mais próximos não nulos do pixel central, apresentando sobre aquele a vantagem de produzir bordas diagonais menos atenuadas.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad Gy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{19}$$

3.4 Morfologia Matemática

Morfologia digital ou *matemática* [13] é uma modelagem destinada à descrição ou análise da forma de um objeto digital. O modelo morfológico para a análise de imagens fundamenta-se na extração de informações a partir de transformações morfológicas, nos conceitos da álgebra booleana e na teoria dos conjuntos e reticulados. O princípio de morfologia digital se embasa no fato de que a imagem é um conjunto de pontos elementares (pixels ou voxels) que formam subconjuntos elementares bi ou tridimensionais. Os subconjuntos e a inter-relação entre eles formam estruturalmente a morfologia da imagem.

As operações básicas da morfologia digital são: (i) a *erosão*, a partir da qual são removidos da imagem os pixels que não atendem a um dado padrão; e (ii) a *dilatação*, a

partir da qual uma pequena área relacionada a um pixel é alterada para um dado padrão. Todavia, dependendo do tipo de imagem sendo processada (preto e banco, tons de cinza ou colorida) a definição destas operações muda, de forma que cada tipo deve ser considerado separadamente. As demais operações e transformações baseiam-se nos operadores básicos dos conjuntos, algumas interativas, e nos dois operadores básicos da morfologia matemática.

Seja a imagem da Fig. 23, na qual há dois objetos ou conjuntos de pixels ***A*** e ***B***. Considere-se que os valores que os pixels podem assumir são binários, i.e., 0 ou 1, o que permite restringir a análise ao espaço discreto Z^2.

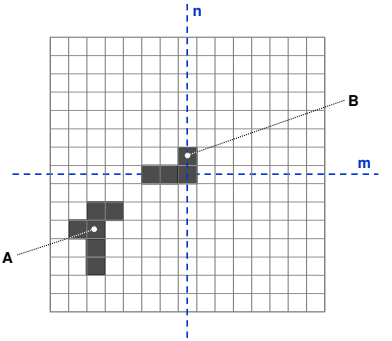


Fig. 23 - Imagem binária contendo 2 objetos, i.e.,2 conjuntos de pontos.

O objeto ***A*** consiste dos pontos α com pelo menos uma propriedade em comum, a saber:

$$Objeto\ A\text{: } A = \{\alpha \perp propriedad\ e(\alpha) = Verdade\} \tag{20}$$

Assim sendo, o objeto ***B*** **d**a Fig. 23 consistirá de {[-2,0][-1,0][0,0][0,1]}. O fundo da imagem de ***A***, denominado $\boldsymbol{A^C}$ (*complemento de* ***A***), consistirá de todos os pontos que não pertencem ao objeto ***A***:

$$Fundo: \quad A^C = \{\alpha \perp \alpha \notin A\} \tag{21}$$

As operações fundamentais associadas com um objeto são o conjunto padrão de operações: *união* ($\{\cup\}$), *interseção* ($\{\cap\}$) e *complemento* ($\{^C\}$) com *translação*. Dado um vetor $\boldsymbol{x}$ e um conjunto ***A***, a *translação*, $\boldsymbol{A + x}$, é definida como:

$$Translação\text{: } A + x = \{\alpha + x \perp \alpha \in A\Phi\} \tag{22}$$

O conjunto básico de operações de *Minkowski* [12], *adição* e *subtração*, pode ser definido em função das considerações anteriores. Dados dois conjuntos ***A*** e ***B*** contidos em um conjunto **C**, a *soma* de Minkowski de ***A*** e ***B*** é o subconjunto de ***C***, denotado $\boldsymbol{A \oplus B}$, dado por:

$$A \oplus B = \{x \in C : \exists a \in A \ \ e \ \ \exists b \in B, x = a + b\} \tag{23}$$

A *diferença* de Minkowski entre ***A*** e ***B*** é o subconjunto de ***C***, denotado $\boldsymbol{A \ominus B}$, dado por:

$$A \ominus B = \{y \in C : \forall b \in B, (\exists a \in A, y = a - b\} \tag{24}$$

Como mencionado anteriormente, as transformações singulares são realizadas através dos operadores elementares, os quais foram denominados por [10] e [11] de transformações de *dilatação* e *erosão* às quais foram incorporadas, posteriormente, mais duas transformações denominadas de *anti-dilatação* e *anti-erosão* [12]. Dilatações e erosões são usadas para a criação de transformações mais sofisticadas, as quais conduzem a vários resultados relevantes quanto à análise de imagens, dentre os quais se citam os filtros morfológicos, o preenchimento de buracos, a extração de contornos e o reconhecimento de padrões. Os operadores de dilatação e erosão invariante por translação, sobre imagens binárias, advieram originalmente das operações de adição e subtração de Minkowski, cada um dos quais pode, em geral, ser caracterizado por um subconjunto denominado *elemento estruturante*.

Via de regra, a construção de sistemas morfológicos é implementada a partir da concepção do problema e da seleção dos operadores mais adequados à solução de interesse. A adequação de operadores constitui um dos grandes problemas encontrados na especificação dos elementos estruturantes. A criação de um mecanismo capaz de encontrar os elementos estruturantes "adequados" à realização da transformação de interesse é uma possível solução para tal problema. A partir das operações básicas de Minkowski, podem-se definir as operações básicas da morfologia matemática, *dilatação* e *erosão*:

$$\textit{Dilatação}: D(A, B) = A \oplus B = A \oplus B = \{x \in E \mid Bx \cap A \neq \varnothing\} \tag{25}$$

$$\textit{Erosão}: E(A, B) = A \ominus B = \{x \in E \mid Bx \subseteq A\} \tag{26}$$

Tanto o conjunto $\boldsymbol{A}$ quanto o conjunto $\boldsymbol{B}$ podem ser considerados como sendo imagens. Todavia, $\boldsymbol{A}$ costuma ser considerado com sendo a imagem sob análise e $\boldsymbol{B}$ como o *elemento estruturante*, o qual está para a morfologia como a *máscara* (*mask*, *template* ou *kernel*) está para teoria de filtragem linear. Os *elementos estruturantes* mais comuns são os conjuntos 4-conexões e 8-conexões, $\boldsymbol{N_4}$ e $\boldsymbol{N_8}$ (Fig. 24).

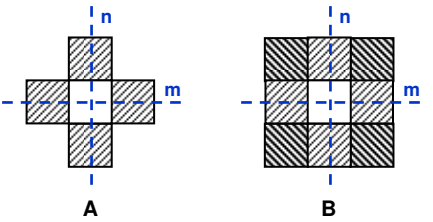


Fig. 24 - Elementos *estruturantes*: (A) padrão N_4; e (B) padrão N_8.

A *dilatação*, em geral, faz com que o objeto cresça no tamanho. Buracos menores do que o elemento estruturante são eliminados e o número de componentes pode diminuir. Por sua vez, a *erosão* reduz as dimensões do objeto. Objetos menores do que o elemento estruturante são eliminados e o número de componentes pode aumentar. O modo e a magnitude da expansão ou redução da imagem dependem necessariamente do *elemento estruturante* $\boldsymbol{B}$. A aplicação de uma transformação de dilatação ou erosão a uma imagem sem a especificação de um *elemento estruturante*, não produzirá nenhum efeito.

3.5 Segmentação

Em processos de análise de imagens, faz-se necessária a extração de medidas, características ou informação de uma dada imagem por métodos automáticos ou semi-automáticos. A primeira etapa da análise de imagem é, em geral, caracterizada por sua *segmentação* [3], que consiste na subdivisão da imagem em partes ou objetos constituintes. Algoritmos de segmentação possibilitam a identificação de diferenças entre dois ou mais objetos, assim como a discriminação das partes, tanto entre si quanto entre si e o *background*. No tocante à segmentação de imagens monocromáticas, os algoritmos fundamentam-se, em essência, na *descontinuidade* e na *similaridade* dos níveis de cinza. A fundamentação na *descontinuidade* consiste no particionamento da imagem em zonas caracterizadas por mudanças bruscas dos níveis de cinza. O interesse recai usualmente na detecção de pontos isolados, de linhas e de bordas da imagem. Por outro lado, a fundamentação na *similaridade* consiste na limiarização e no crescimento de regiões.

3.5.1 Limiarização (*Thresholding*)

Limiarização é uma abordagem para a segmentação fundamentada na análise da similaridade de níveis de cinza, de modo a extrair objetos de interesse mediante a definição de um limiar *T* que separa os agrupamentos de níveis de cinza da imagem. Uma das dificuldades do processo reside na determinação do valor mais adequado de limiarização, i.e., do ponto de separação dos pixels da imagem considerada. Através da análise do histograma da imagem, é possível estabelecer um valor para *T* na região do *vale* situado entre picos que caracterizam regiões de interesse na imagem. Há diversas variantes de limiarização. A mais simples delas é a técnica do particionamento do histograma da imagem por um limiar único *T*. A segmentação se dá varrendo-se a imagem, pixel a pixel, e rotulando-se cada pixel como sendo do objeto ou do fundo, em função da relação entre o valor do pixel e o valor do limiar. O sucesso deste método depende inteiramente de quão bem definidas estão as massas de pixels no histograma da imagem a ser segmentada.

3.5.2 Segmentação orientada a regiões

A segmentação *orientada a regiões* se fundamenta na similaridade dos níveis de cinza da imagem. O *crescimento de regiões* é um procedimento que agrupa pixels ou sub-regiões de uma imagem em regiões maiores. A variante mais simples da segmentação orientada a regiões é a *agregação de pixels,* que se fundamenta na definição de uma *semente*, i.e., um conjunto de pontos similares em valor de cinza, a partir do qual as regiões crescem com a agregação de cada pixel à semente à qual estes apresentem propriedades similares (e.g. nível de cinza, textura ou cor). A técnica apresenta algumas dificuldades fundamentais, se afigurando como problemas imediatos (i) a seleção de sementes que representem adequadamente as regiões de interesse; e (ii) a seleção de propriedades apropriadas para a inclusão de pontos nas diferentes regiões, durante o processo de crescimento. A disponibilidade da informação apropriada possibilita, em cada pixel, o cálculo do mesmo

conjunto de propriedades que será usado para atribuir os pixels às diferentes regiões pré-definidas, durante o processo de crescimento. Caso o resultado de tal cálculo implique agrupamentos de valores das propriedades, os pixels cujas propriedades se localizarem mais perto do centróide desses agrupamentos poderão ser usados como sementes.

3.5.3 Segmentação Baseada em Bordas

A detecção de bordas, anteriormente discutida, possibilita a análise de descontinuidades nos níveis de cinza de uma imagem. As bordas na imagem de interesse caracterizam os contornos dos objetos nela presentes, sendo bastante úteis para a segmentação e identificação de objetos na cena. Pontos de borda podem ser entendidos como as posições dos pixels com variações abruptas de níveis de cinza. Os pontos de borda caracterizam as transições entre objetos diferentes. Várias técnicas de segmentação baseiam-se na detecção de bordas, sendo as mais simples aquelas nas quais as bordas são detectadas pelos operadores de gradiente (e.g. Sobel, Roberts, Laplaciano), seguida de um processo de limiarização.

3.6 Extração de Características e Reconhecimento

A próxima tarefa após a segmentação é o reconhecimento dos objetos ou regiões resultantes. O objetivo do reconhecimento de padrões é identificar objetos na cena a partir de um conjunto de medições. Cada objeto é um padrão e os valores medidos são as características desse padrão. Um conjunto de objetos similares, com uma ou mais características semelhantes, é considerado como pertencente à mesma classe de padrões. Há diversos tipos de características, cada uma das quais é obtida a partir de uma técnica específica. Além disso, características de ordem mais alta advêm da combinação de características mais simples, e.g. cada letra do alfabeto é composta por um conjunto de características como linhas verticais, horizontais e inclinadas, bem como segmentos curvilíneos. Enquanto a letra **A** pode ser descrita por duas linhas inclinadas e outra horizontal, a letra **B** pode ser descrita por uma linha vertical e 2 curvilíneas conectadas em pontos específicos. Outras características relevantes para um objeto 2D ou 3D são a área, volume, perímetro, superfície, dentre outras, as quais podem ser medidas a partir da contagem de pixels.

Analogamente, a forma de um objeto pode ser descrita em termos de suas bordas. Outros atributos mais específicos para a forma podem ser obtidos através de invariantes de momentos, descritores de Fourier, eixos medianos dos objetos, dentre outros [1][2][3][14]. Para realizar o reconhecimento de objeto, existe uma grande variedade de técnicas de classificação. Uma representação geral para o processo de classificação é ilustrada na Fig. 25.

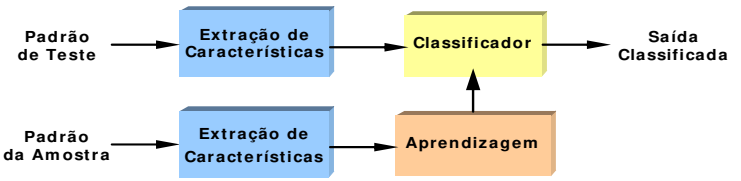


Fig. 25 - Representação geral para o processo de classificação.

As técnicas de reconhecimento de padrões podem ser divididas em 2 tipos principais: classificação baseada em aprendizagem *supervisionada* e *não-supervisionada*. Por sua vez, os algoritmos de classificação supervisionada subdividem-se em *paramétricos* e *não-paramétricos*. O classificador *paramétrico* é treinado com uma grande quantidade de amostras rotuladas (conjunto de treinamento, padrões cujas classes se conhecem *a priori*) para que possa estimar os parâmetros estatísticos de cada classe de padrão (e.g. média, variância). Exemplos de classificadores supervisionados são os de *distância mínima* e o de *máxima verossimilhança*. Na *classificação não-paramétrica*, os parâmetros estimados do conjunto de treinamento não são levados em consideração. Um exemplo de classificador não paramétrico é o dos *K-vizinhos mais próximos*. Na classificação *não supervisionada*, o classificador particiona o conjunto de dados de entrada a partir de algum critério de similaridade, resultando em um conjunto de *clusters* ou grupos, cada um dos quais normalmente associado a uma classe. Na área de reconhecimento de objetos, destacam-se os algoritmos e técnicas baseadas em redes neurais [15] (com variantes tanto para classificação supervisionada como para classificação não-supervisionada). Outro importante exemplo são os classificadores bayesianos [3]. Uma visão mais aprofundada da área de classificação de padrões pode ser encontrada em [16].

4 Exemplos de Aplicações

O objetivo desta seção é fornecer exemplos que abordem alguns dos conceitos e operações apresentadas nas seções anteriores. Com o fim de promover a disseminação da área no Brasil, os exemplos de aplicações apresentados a seguir foram selecionados dos anais do principal evento nacional da área, o *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*.

4.1 Segmentação de imagens

Conforme visto na Seção 3.5, a segmentação de imagens tem como principal objetivo a separação de objetos de interesse do *background* da imagem. Na segmentação por limiarização, a escolha de um limiar normalmente depende de características intrínsecas da imagem, e.g. entropia e outras estatísticas, não levando usualmente em conta a percepção humana do processo de segmentação. Numa abordagem alternativa para realizar a limiarização de imagens em tons de cinza, foi proposto em [17] um método de modelagem perceptiva que aprende a decisão humana na limiarização através de uma rede de funções de base radial (RBFN), uma máquina de aprendizagem que permite aproximar a função que mapeia características globais da imagem (e.g. desvio padrão dos tons de cinza) em limiares escolhidos por humanos. A partir de imagens de treinamento, o usuário seleciona o limiar (nível de cinza) que melhor separa os pixels do *background* daqueles do objeto. As decisões são armazenadas em uma tabela de 2 colunas, a primeira coluna armazena o limiar escolhido e a outra armazena uma característica global da imagem. Essa tabela é então utilizada para o treinamento da RBFN. Como resultado da comparação da modelagem perceptiva com três outros métodos automáticos de segmentação por limiarização, verificou-se que as respostas humanas possuíam alta correlação com alguns dos métodos automáticos avaliados, demonstrando a viabilidade da abordagem proposta.

Uma abordagem para a segmentação dos blocos de endereço em envelopes postais, baseada em histogramas 2D e a operação morfológica de *watershed*, foi apresentada por [18]. Considerem-se uma imagem digital $F = [f(x,y)]$ e sua versão $G = [g(x,y)]$ filtrada através de um filtro de média ou, conforme proposto pelos autores, filtrada através de uma reconstrução morfológica. Ambas as imagens possuem dimensões $M \times N$ e com $[0,...,L\text{-}1]$ tons de cinza. O histograma 2D $C = [c_{ij}]$ de dimensões $L \times L$ é computado a partir de quaisquer pares de pixels $f(x,y)$ e $g(x,y)$ que possuem os tons de cinza iguais a i e j, respectivamente, podendo ser formalizado como segue: $c_{ij} = \#\{(f(x,y), g(x,y)) \mid f(x,y) = i, f(x,y) = j \}$, em que o operador # denota a cardinalidade do conjunto operando.

A partir do histograma 2-D é realizado um processo de agrupamento das regiões da imagem que correspondem a 3 classes: blocos de endereço e carimbos postais, selos e *background* do envelope. Esse agrupamento é conduzido através da operação de *watershed*. Uma vez que o histograma 2-D pode ser visto como uma topografia na qual os maiores valores correspondem a picos e os menores a vales, se uma gota de água é depositada em um ponto qualquer da região do histograma, ela irá escoar para um vale (ponto de mínimo local). A área da *watershed* associada ao mínimo M é definida como sendo o agrupamento de todos esses pontos de mínimo cuja elevação tem valor igual a M. O ponto de encontro entre duas áreas de *watershed* gera uma borda que é o resultado final da operação. A complexidade das imagens de envelope gerou histogramas 2-D com um número muito elevado de vales (tipicamente mais de 25). Para evitar uma supersegmentação da imagem, realizou-se uma seqüência de erosões morfológicas de modo a reduzir o número de regiões para apenas 3, correspondendo exatamente ao número de classes a serem segmentadas. Uma avaliação experimental demonstrou que o bloco de endereço com carimbos postais foi segmentado corretamente em 75% dos casos, o que demonstra uma significativa robustez.

4.2 Reconhecimento de Manuscritos

Diferentemente da escrita mecânica, na qual há uma grande regularidade na forma, intensidade e posicionamento das palavras e caracteres, apesar das diferentes fontes e estilos, os manuscritos apresentam enorme variação em todos esses aspectos, além de serem dependentes do autor. Aplicações típicas envolvem a verificação de assinaturas e o reconhecimento de textos manuscritos na forma de caracteres isolados e palavras inteiras, dentre outros.

No trabalho de [19], foi apresentada uma avaliação de duas abordagens para o reconhecimento de palavras isoladas dos meses do ano: uma baseada em Redes Neurais (RN) e a outra baseada em Modelos de Markov Escondidos (MME). O primeiro estágio do processamento consistiu de 3 etapas: (i) correção da inclinação geral dos caracteres; (ii) detecção e correção de inclinações na linha de base da palavra inteira; e (iii) utilização de um filtro para atenuação de imperfeições e falhas nos manuscritos. O estágio seguinte foi específico para cada classificador utilizado. Para o classificador neural, dividiu-se a imagem dos manuscritos em 8 sub-regiões fixas, correspondendo a aproximadamente o número médio de letras no conjunto de palavras a serem reconhecidas (meses do ano). Para cada uma das sub-regiões, um total de 10 características perceptivas foi extraído, produzindo um padrão com 80 características para cada palavra. As

características perceptivas foram obtidas a partir da análise direcional dos pixels e incluíram as posições e tamanho das linhas ascendentes, descendentes e *loops* fechados, além dos ângulos das concavidades e uma estimativa para o tamanho da palavra. A ausência de uma essas características numa sub-região particular foi indicada pelo valor 1. Para o classificador de Modelos de Markov Escondidos, a partir do histograma de projeção horizontal dos pixels da imagem do manuscrito, três zonas foram definidas: ascendente, corpo e descendente. Um processo de segmentação variável, dependente das transições escuro-claro presentes na linha central da palavra, é aplicado. Em seguida, para cada segmento, foram identificadas características perceptivas e características baseadas em deficiências na concavidade/ convexidade dos traços encontrados nesses segmentos. Na avaliação experimental, 3600 imagens de manuscritos contendo os meses do ano foram utilizadas para treinamento, 1200 para teste e 1200 para validação dos classificadores. Como resultado, verificou-se que a melhor taxa de reconhecimento ocorreu para o classificador neural utilizando características perceptivas (81,8%), enquanto foi possível obter uma taxa de reconhecimento muito superior combinando 3 classificadores (um baseado em Modelos de Markov Escondidos e 2 baseados em Redes Neurais), com um resultado de 90.4% de correta classificação.

4.3 Classificação e Recuperação de Imagens por Conteúdo

A classificação e recuperação de imagens por conteúdo têm forte relação com as áreas de sistemas de informação e banco de dados. Uma consulta tradicional a um banco de dados normalmente envolve a utilização de chaves textuais ou numéricas como parte de expressões relacionais e lógicas. O próximo passo lógico é justamente incluir campos e operações (e.g. classificação, segmentação, etc.) sobre imagens na consulta. Atualmente existem vários sistemas de banco de dados comerciais (e.g. Oracle) e não-comercias (e.g. Postgres) que permitem algumas funcionalidades envolvendo imagens.

Um sistema para a classificação de imagens coletadas da *Web* em duas classes semânticas, gráficos e fotografias, foi apresentado por [20]. O sistema utilizou um método de classificação baseado em árvores de decisão (ID3, um algoritmo de indução de árvores de decisão a partir de exemplos, popular na área de IA). Foi identificado um conjunto de características adequadas à separação entre as duas classes semânticas escolhidas. Características marcantes de fotografias identificadas no trabalho foram: (i) existências de objetos reais com uma tendência a texturas e ausência de regiões com cores constantes; (ii) pequenas diferenças na proporção (altura x largura); (iii) poucas ocorrências de regiões com alta saturação de cores; e (iv) presença de um grande número de cores utilizadas. As características identificadas como marcantes de gráficos foram: (i) presença de objetos artificiais com bordas bem definidas bem como a presença de regiões cobertas com cores saturadas; e (ii) grandes diferenças na proporção e tendência a serem menores em tamanho do que fotografias. Assim, foram definidas métricas sobre o número de cores, a cor predominante, o vizinho mais distante, a saturação, o histograma de cores, o histograma do vizinho mais distante, a proporção das dimensões e a menor dimensão.

As duas primeiras métricas, diretas, não serão mencionadas neste texto. A métrica do

vizinho mais distante é baseada nas transições entre cores. Para dois *pixels* $p1$ e $p2$, de cores $(r1,g1,b1)$ e $(r2,g2,b2)$, foi definida uma medida de distância d como sendo: $d = |r1 - r2| + |g1 - g2| + |b1 - b2|$. Considerando que cada componente de cor varia de 0 a 255, então d varia de 0 a 765. A partir de uma vizinhança de 4 *pixels* (acima, abaixo, esquerda e direita), um vizinho $p2$ de $p1$ é considerado como sendo o vizinho mais distante se a medida d para $p2$ for a maior de todas as distâncias dentro da vizinhança. A métrica de saturação de um *pixel* $p = (r,g,b)$ é definida como $|m - n|$, em que m e n são os valores mínimo e máximo entre os valores de r,g e b, respectivamente. A métrica do histograma de cores é definida a partir da correlação entre o histograma de uma imagem t de teste e os histogramas médios para um conjunto de referência f de fotografias e outro conjunto de referência g para gráficos. Supondo $a = C(H_t, H_f)$ e $b = C(H_t, H_g)$, em que C é a correlação (produto interno) entre dois histogramas, a métrica do histograma de cores foi definida como $s = b / (a + b)$. Claramente, a medida que a aumenta, s também aumenta, e, à medida em que b aumenta, s diminui. Assim, espera-se que fotografias tenham uma resposta maior em s quando comparadas a gráficos. A métrica do histograma do vizinho mais distante baseou-se nas mesmas premissas da métrica do vizinho mais distante, mas fornece uma forma diferente de testar a imagem. A métrica da proporção é definida como m / l, em que m é o valor máximo entre a altura e a largura da imagem e l é o valor mínimo. Finalmente, a métrica da menor dimensão é simplesmente o valor de l.

Na fase experimental, foram definidos dois conjuntos de treinamento, contendo gráficos e fotografias nos formatos e imagem GIF (3058 gráficos e 1350 fotografias) e JPEG (1434 gráficos e 4763 fotografias). Para cada conjunto de treinamento, foram extraídas as métricas discutidas acima e cada vetor de características de uma dada imagem recebeu um rótulo (gráfico ou fotografia) através de inspeção visual da imagem. A aplicação do algoritmo ID3 gerou uma árvore de decisão para a classificação de cada conjunto. As taxas médias de classificação correta em imagens de teste, não utilizadas durante o treinamento, corresponderam a 97,3% para imagens GIF e 93,9% para imagens JPEG, com desvios padrão de 1,6 e 2,6, respectivamente.

5 Considerações Finais

O presente tutorial forneceu uma visão geral da área de PDI, tendo como um dos objetivos despertar, por parte de alunos brasileiros de nível técnico e superior, o interesse pela área. Outro objetivo foi o de permitir uma reciclagem ou um primeiro contato de profissionais dos diferentes setores da economia, cujas atividades envolvam alguma informação baseada em imagens. Por se tratar de uma área bastante ampla, não foi possível incluir todos os possíveis tópicos relevantes, mas procurou-se fornecer um mínimo de detalhes associados a cada etapa de processamento em um sistema típico de PDI, da aquisição à classificação. Para aqueles interessados em se aprofundar nos tópicos pouco explorados, e.g. segmentação, extração de características e classificação, ou em outros tópicos igualmente importantes que não puderam ser incluídos neste documento por restrições de espaço, e.g. transformações geométricas, representação no domínio da freqüência (transformada de Fourier e *Wavelets*), técnicas de compressão, dentre muitos outros, poderão fazê-lo consultando as referências apresentadas a seguir.

Referências

[1] JÄHNE, B. *Digital Image Processing*. Springer-Verlag, 2002.

[2] ACHARYA, T., RAY, A. K. *Image Processing- Principles and Applications*. John Wiley & Sons, Inc. 2005.

[3] GONZALEZ, R., WOODS, P. *Digital Image Processing*. Prentice Hall, 2002, 2nd ed.

[4] FORSYTH, D., PONCE, J. *Computer Vision: A modern approach*. Prentice Hall, 2001.

[5] JÄHNE, B., HAUSSECKER, H. (Eds.) *Handbook of Computer Vision and Applications*. Academic Press, 2000.

[6] RENCZ, A. N., RYERSON. R. A. (Eds.) *Manual of Remote Sensing, Remote Sensing for the Earth Sciences*. John Wiley & Sons, Inc. 1999, 3rd ed.

[7] HANSEN, C. D., JOHNSON, C. R. *Visualization Handbook*. Elsevier, 2005.

[8] RUSS, J. C. *The image processing handbook*. CRC Press LLC, 2000 3rd ed.

[9] BANKMAN, I. (Ed.) *Handbook of Medical Imaging: Processing and Analysi*s. Academic Press. 2000.

[10] MATHERON, G. *Random sets and integrated geometry*. Wiley, 1975.

[11] SERRA, J. Introduction to mathematical morphology, *Computer Vision, Graphics and Image Processing*, 35(3):283–305, September 1986.

[12] SERRA, J. *Image analysis and mathematical morphology*. Academic Press, London, 1988.

[13] DOUGHERTY, E. R., LOTUFO, R. A. *Hands-on Morphological Image Processing*, SPIE Press, Bellingham, 2003, 1st ed.

[14] CESAR JR, R. M., COSTA, L. F. *Shape Analysis and Classification Theory and Practice*. CRC Press, 2001.

[15] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall. 1998. 2nd ed.

[16] DUDA, R. O. *Pattern Classification*, John Wiley & Sons, Inc., 2000, 2nd ed.

[17] LOPES, L. M., CONSULARO, L. A. A RBFN Perceptive Model for Image Thresholding, *Proc. of SIBGRAPI*, pp 225-232, 2005.

[18] YONEKURA, E., FACON, A. J. 2-D Histogram-based Segmentation of Postal Envelopes, *Proc. of SIBGRAPI*, pp 247-251, 2003.

[19] OLIVEIRA JR., J. J., CARVALHO, J. M., FREITAS, C. O. A., SABOURIN, R. Evaluating NN and HMM Classifiers for Handwritten Work Recognition, *Proc. of SIBGRAPI*, pp 210-217, 2002.

[20] OLIVEIRA, C. J. S., ARAÚJO, A. A., SEVERIANO JR, C. A.., GOMES, D. R. "Classifying Images Collected on the World Wide Web", *Proc. of SIBGRAPI*, pp 327-334, 2002.