

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

LUCIANO SILVA DA SILVA

**Uma Proposta para a Representação
Geométrica de Imagens com Aplicação em
Segmentação e Compressão**

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Prof. Dr. Jacob Scharcanski
Orientador

Porto Alegre, abril de 2005

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Silva, Luciano Silva da

Uma Proposta para a Representação Geométrica de Imagens com Aplicação em Segmentação e Compressão / Luciano Silva da Silva. – Porto Alegre: PPGC da UFRGS, 2005.

63 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2005. Orientador: Jacob Scharcanski.

1. Representação de Imagens. 2. Segmentação de Imagens. 3. Compressão de Imagens. 4. Modelagem de Malhas. I. Scharcanski, Jacob. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Vice-Reitor: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora de Pós-Graduação: Prof^a. Valquiria Linck Bassani

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Flávio Rech Wagner

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	5
LISTA DE FIGURAS	6
LISTA DE TABELAS	8
RESUMO	9
ABSTRACT	10
1 INTRODUÇÃO	11
2 REPRESENTAÇÃO GEOMÉTRICA EM IMAGENS: CONCEITOS E ESTADO DA ARTE	13
2.1 Introdução	13
2.2 Triangulação de Delaunay e Triangulação de Delaunay com Restrições	15
2.3 Malhas Regulares	17
2.4 Malhas Irregulares	17
3 SEGMENTAÇÃO DE IMAGENS: CONCEITOS E ESTADO DA ARTE	20
3.1 Introdução	20
3.2 Revisão de Técnicas de Segmentação de Imagens	21
3.2.1 Técnicas Orientadas à Análise de Feições	22
3.2.2 Técnicas Baseadas em Malhas	24
4 COMPRESSÃO DE IMAGENS: CONCEITOS E ESTADO DA ARTE	26
4.1 Introdução	26
4.2 Fundamentos	27
4.2.1 Redundância de Dados	27
4.2.2 Codificador e Decodificador	29
4.2.3 Critérios de Fidelidade	29
4.3 Compressão sem perdas	31
4.4 Compressão com perdas	34
5 PROPOSTAS PARA A REPRESENTAÇÃO GEOMÉTRICA DE IMAGENS	35
5.1 Modelagem de Imagens por Malhas	35
5.1.1 Aproximação de Bordas	35
5.1.2 Aproximação de Regiões	38
5.1.3 Seleção dos Vértices	40

5.2	Um Método para a Segmentação de Imagens usando Representação Geométrica	43
5.2.1	Subdivisão Triangular	44
5.2.2	Clustering	44
5.3	Um Método para a Compressão de Imagens Usando Representação Geométrica	46
5.3.1	Subdivisão Triangular	47
5.3.2	Esquema de Codificação	48
6	RESULTADOS EXPERIMENTAIS	49
6.1	Segmentação de Imagens	49
6.2	Compressão de Imagens sem Perdas	53
6.3	Compressão de Imagens com Perdas	53
7	CONCLUSÕES	57
7.1	Trabalhos Futuros	58
	REFERÊNCIAS	59

LISTA DE ABREVIATURAS E SIGLAS

DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DHT	Discrete Hartley Transform
DPCM	Differential Pulse Code Modulation
DWT	Discrete Wavelet Transform
FCM	Fuzzy c-means
JBIG	Joint Bi-level Image Experts Group
JPEG	Joint Photographic Experts Group
KL	Karhunen-Loève
LG	Line Graph
LoG	Laplacian of Gaussian
LZW	Lempel-Ziv-Welch
MRF	Markov Random Fields
RAG	Region Adjacency Graph
WHT	Walsh-Hadamard Transform

LISTA DE FIGURAS

Figura 2.1:	Triangulações para um pequeno conjunto de pontos (a) apenas triângulos de Delaunay; (b) nem todos são triângulos de Delaunay.	16
Figura 2.2:	Exemplos de duas triangulações de Delaunay válidas para o mesmo conjunto de pontos.	16
Figura 4.1:	Diagramas em blocos: (a) codificador e (b) decodificador.	29
Figura 4.2:	Modelo de codificação preditiva: (a) codificador e (b) decodificador. .	32
Figura 5.1:	Aproximação de regiões por detecção de bordas (a) original; (b) borda simples (detector de primeira ordem); (c) borda dupla (detector de segunda ordem); (d) aproximação com base no máximo do gradiente; (e) aproximação com base nos máximos absolutos do laplaciano. . . .	37
Figura 5.2:	Exemplo de uma borda conexa com uma bifurcação (a) o pixel com mais de dois vizinhos é localizado; (b) a borda é quebrada em duas. .	37
Figura 5.3:	Aproximação de uma borda por segmentos de retas (a) inicialização; (b) o segmento é dividido em dois no ponto mais distante; (c) final do processo.	38
Figura 5.4:	Processo de criação de uma Triangulação de Delaunay com Restrições (a) imagem original; (b) mapa de bordas; (c) aproximação por segmentos de retas; (d) Triangulação de Delaunay com Restrições correspondente.	39
Figura 5.5:	Inicialização do processo de inserção de vértices.	42
Figura 6.1:	Teste de segmentação com e sem refinamento de regiões. As linhas do topo para baixo mostram: Imagens Originais, mapas de bordas, triangulações sem refinamento, aproximações sem refinamento, triangulações com refinamento e aproximações com refinamento.	50
Figura 6.2:	Comparação do método proposto com o <i>mean-shift</i> . (a,d) imagens originais; (b) imagem segmentada pelo método proposto com 26 cores e 2070 regiões conexas; (c) imagem segmentada pelo método <i>mean-shift</i> com 2399 cores e 2407 regiões conexas; (e) imagem segmentada pelo método proposto com 53 cores e 347 regiões conexas; (f) imagem segmentada pelo método <i>mean-shift</i> com 318 cores e 319 regiões conexas.	51

Figura 6.3:	Método de segmentação proposto aplicado à quantização de paletas de cores. (a,d,g) imagens originais; (b) resultado do método proposto por Cheng e Sun, com 14 cores e PSNR = 20,98dB; (c) resultado obtido a partir do método proposto, com 14 cores e PSNR = 23,97dB; (e) resultado do método proposto por Cheng e Sun, com 8 cores e PSNR = 20,72dB; (f) resultado obtido a partir do método proposto, com 8 cores e PSNR = 22,75dB; (h) resultado do método proposto por Cheng e Sun, com 9 cores e PSNR = 22,99dB; (i) resultado obtido a partir do método proposto, com 9 cores e PSNR = 23,49dB.	52
Figura 6.4:	Mamografias digitalizadas de 16 bits.	54
Figura 6.5:	Compressão com perdas. (a) imagens originais; (b) imagens comprimidas pelo método proposto e (c) imagens comprimidas pelo método JPEG.	56

LISTA DE TABELAS

Tabela 4.1:	Código de Gray de 3 bits	32
Tabela 6.1:	Taxas de compressão em bits/pixel para mamografias de 16 bits. . . .	53
Tabela 6.2:	Resultados da compressão com perdas	55

RESUMO

O presente trabalho descreve uma proposta para a representação geométrica de imagens. Através da subdivisão espacial adaptativa de uma imagem em triângulos, uma representação simplificada da estrutura da imagem pode ser obtida. Demonstramos que a representação gerada é adequada para aplicações como a segmentação e a compressão de imagens. O método de segmentação de imagens desenvolvido faz uso deste tipo de representação para obter resultados robustos e compactos, comparados a outros métodos existentes na literatura, e adequado para aplicações como a detecção, descrição e codificação de objetos. Utilizando uma representação geométrica semelhante a métodos de modelagem de superfícies, criamos um novo método de compressão de imagens que apresenta vantagens em relação a outros métodos existentes, em especial na compressão de imagens sem perdas.

Palavras-chave: Representação de Imagens, Segmentação de Imagens, Compressão de Imagens, Modelagem de Malhas.

A New Approach for Geometric Representation of Images Applied to Segmentation and Compression

ABSTRACT

In the present work, a geometric image representation approach is described. A simplified representation of the image structure has been obtained through the adaptative spatial image subdivision in triangles. We demonstrate that the resulting representation is suitable for applications such as image segmentation and compression. This type of representation enables to obtain more compact and robust segmentation results than those obtained through other methods described in the literature. Besides it is applicable in object detection, description and coding techniques. A new image compression method is also created by using a geometric representation similar to surface modelling techniques. This new approach presents advantages when compared to existing ones, especially regarding to lossless image compression.

Keywords: Image Representation, Image Segmentation, Image Compression, Mesh Modelling.

1 INTRODUÇÃO

Com os avanços das diversas áreas do processamento de imagens, a necessidade de novas formas de representação se faz cada vez mais importante. Devido à proliferação do conteúdo multimídia nos últimos anos, usuários desejam que o conteúdo possa ser acessado de acordo com suas próprias preferências. Além do mais, aparelhos portáteis têm apresentado um crescimento de popularidade e de recursos, entusiasmando as pessoas em relação ao acesso de conteúdo multimídia a partir destes aparelhos. A capacidade dos diversos dispositivos de visualização de imagens varia muito em termos de conectividade de rede, velocidade de processamento, limitações de visualização e capacidade de decodificação (TSENG; LIN; SMITH, 2004). A existência de clientes tão heterogêneos somada à grande variedade de fontes de dados, fazem com que a implementação de sistemas universalmente compatíveis que se encaixem em diversos ambientes de uso se torne um grande desafio.

Ainda, o advento da World Wide Web (WWW), bem como os avanços nas tecnologias computacionais resultaram em uma explosão de imagens sendo geradas, armazenadas e transmitidas. No entanto, os mecanismos de busca atualmente permitem que apenas buscas textuais sejam feitas. A identificação de informação visual é uma tarefa difícil e normalmente não existem descrições reconhecíveis deste tipo de material. Isto faz com que seja necessário o desenvolvimento de sistemas de recuperação baseados em conteúdo que permitam aos usuários procurar informações diretamente através do conteúdo semântico das imagens (HSIEH; GRIMSON, 2003).

Para alcançarmos estes objetivos, primeiramente devemos prover meios de diminuir a lacuna semântica existente entre o significado procurado pelos usuários e a forma com que os dados são organizados e armazenados. A representação de imagens e vídeos baseada em objetos é uma alternativa adequada para solucionar este desafio e a atenção da comunidade científica para este tema vem crescendo significativamente nos últimos anos (DIMITROVA et al., 2002; GEVERS, 2002). Padrões para a codificação de estruturas de alto nível semântico já existem (NACK; LINDSEY, 2002a,b; MARTÍNEZ; KOENEN; PEREIRA, 2002; MARTÍNEZ, 2002; TSENG; LIN; SMITH, 2004); no entanto, não há unanimidade em relação a métodos que façam a ligação entre os aspectos de baixo nível e os de alto nível em imagens ou vídeos. Isto se deve, em grande parte, à independência dos aspectos de processamento de baixo nível, tais como a segmentação de imagens, em relação aos aspectos semânticos de alto nível. A falta de robustez dos métodos de baixo nível existentes é uma das principais causas para esta lacuna semântica, no sentido de que

pequenas alterações em parâmetros internos dos métodos ou fatores externos (ruído, iluminação) geram grandes mudanças no resultado final, no que diz respeito ao significado de alto nível percebido pelos usuários.

Os primeiros métodos de codificação sem perdas de imagens eram simples aplicações de métodos da teoria de informação para sinais 1D (HUFFMAN, 1952; GOLOMB, 1966), em versões 2D. Métodos de compressão com perdas baseados em transformadas se tornaram populares, por apresentarem maior compactação da informação visual em suas representações (PENNEBAKER; MITCHELL, 1993). Atualmente, melhorias consideráveis no desempenho tanto da compressão com perdas quanto da compressão sem perdas estão sendo alcançadas a partir de métodos que incluem informação de contexto e características singulares das imagens visando a redução da quantidade de dados necessária para a reconstrução da imagem, como é o caso dos métodos baseados em segmentação (SHEN; RANGAYYAN, 1997; RATAKONDA; AHUJA, 2002), para compressão sem perdas, e os métodos baseados em modelagem de malhas (DEMARET; ROBERT, 2000; ZHAOYANG et al., 2001), para compressão com perdas.

O presente trabalho descreve uma abordagem de representação de imagens baseada na modelagem de malhas triangulares. A malha é construída a partir de uma seleção adaptativa dos vértices com base em informação local, segundo diferentes critérios. É utilizada a Triangulação de Delaunay (BERG et al., 1997), bem como a Triangulação de Delaunay com Restrições (CHEW, 1987) para a formação dos triângulos. Quando a Triangulação de Delaunay com Restrições é considerada, as bordas são respeitadas durante a construção da malha. São apresentados novos métodos de segmentação e de compressão de imagens, utilizando a representação por malhas triangulares.

O método de segmentação desenvolvido pode ser dividido em duas etapas. Na primeira, é feito um mapeamento da representação baseada em pixels para uma representação baseada em triângulos homogêneos. Em seguida, um método de classificação de triângulos é aplicado, agrupando triângulos semelhantes em classes. Os triângulos pertencentes a um mesmo grupo são unidos, gerando um conjunto de regiões poligonais como resultado.

As malhas triangulares também são utilizadas para gerar compressão com perdas e compressão sem perdas. Neste caso, a imagem é vista como uma superfície a ser aproximada por uma malha. Para uma compressão com perdas, apenas os parâmetros para a construção da malha são codificados. Para a compressão sem perdas, o erro de aproximação é codificado adequadamente e armazenado junto à malha.

O presente trabalho está organizado da seguinte forma: no capítulo 2 são discutidas formas de representação geométrica de imagens. O capítulo 3 apresenta uma introdução à segmentação de imagens, bem como uma revisão de métodos de segmentação encontrados na literatura. No capítulo 4 são apresentados conceitos e uma revisão de métodos da literatura para a compressão de imagens com perdas e sem perdas. Uma nova proposta para a representação geométrica de imagens, com sua aplicação em segmentação e compressão de imagens é apresentada no capítulo 5. Resultados experimentais dos métodos propostos para a segmentação e a compressão de imagens, bem como comparações com métodos da literatura são apresentados no capítulo 6. No capítulo 7 são apresentadas as conclusões finais e perspectivas de trabalhos futuros.

2 REPRESENTAÇÃO GEOMÉTRICA EM IMAGENS: CONCEITOS E ESTADO DA ARTE

2.1 Introdução

O método mais comum e mais simples de representação de imagens é a partir de um vetor bi-dimensional $I(i, j)$ de elementos chamados *pixels*. Cada pixel p de uma imagem carrega informação de cor individual para a posição em que o pixel se encontra na imagem, sendo que esta informação de cor pode ser simplesmente um valor escalar relativo à luminância da imagem no ponto p (imagens em tons de cinza), um vetor tri-dimensional que define um ponto em algum espaço de cores qualquer (imagens a cores), ou ainda um vetor n -dimensional que representa a energia eletro-magnética que incide no ponto p em diferentes bandas do espectro (imagens multi-espectrais). Cada pixel abrange uma área igual da imagem e são igualmente espaçados entre si ¹. Este tipo de representação não está relacionada a qualquer característica particular de uma imagem e não inclui qualquer informação semântica. Por isso, seu uso não é apropriado em tarefas de análise e processamento de imagens de alto nível.

Muitas formas de representação foram propostas até hoje para facilitar a execução de tarefas de alto-nível. Os chamados códigos de cadeia (SONKA; HLAVAC; BOYLE, 1999) são comumente usados para a descrição de bordas de objetos, ou outras linhas em uma imagem com um pixel de largura. A borda é definida pelas coordenadas de um pixel de referência e uma sequência de símbolos correspondentes a linhas de uma unidade de comprimento em várias direções pré-definidas. Ou seja, os códigos de cadeia são de natureza relativa; os dados são expressos com relação a algum ponto de referência. Se é necessário extrair informação local de um código de cadeia, então é necessário procurar por toda a cadeia sistematicamente. Por exemplo, se desejarmos saber se uma determinada borda vira 90° à esquerda em algum lugar, simplesmente devemos procurar por um par de símbolos na cadeia correspondentes à mudança de 90° à esquerda. Por outro lado, questões sobre a forma da borda próximo a um determinado pixel (i_0, j_0) não são triviais.

Estruturas de dados topológicas descrevem a imagem como um conjunto de elementos e suas relações. Estas relações são geralmente representadas utilizando-se grafos. Um grafo de adjacência de regiões, ou RAG (*Region Adjacency Graph*) (PAVLIDIS, 1980), é

¹Em alguns casos, os pixels não correspondem a áreas iguais da imagem, ou ainda, não são igualmente espaçados entre si. No entanto, estes são casos particulares e pouco comuns, e por isso não serão considerados neste trabalho

uma estrutura utilizada para representar a disposição espacial das regiões de uma imagem. Um RAG é um grafo em que cada nodo representa uma região e uma aresta entre dois nodos indica que as regiões correspondentes são vizinhas. Os grafos de adjacência provêm uma visão espacial da imagem, facilitando a execução de tarefas que envolvem a análise da disposição espacial das regiões na imagem. Em (TRÉMEAU; COLANTONI, 2000) é proposto um método de segmentação de imagens baseado em crescimento de regiões e *watershed* que faz uso de RAG's. Os autores ainda propuseram uma nova estrutura para a representação de adjacência entre regiões chamada de *line graph* (LG). LG's são estruturas duais aos RAG's, sendo que os nodos de um LG correspondem às arestas de um RAG (ou seja, representam fronteiras entre duas regiões vizinhas) e as arestas de um LG ligam os pares de nodos (fronteiras) que compartilham uma mesma região da imagem. Assim, os LG's mantêm informação de adjacência de segunda ordem. Em (HSIEH; GRIMSON, 2003), é proposto um método de descrição do relacionamento espacial de regiões em uma imagem visando a recuperação de informações. Os autores apresentam um método que calcula medidas de similaridade de imagens com base nas informações de relacionamento extraídas a partir da segmentação das imagens. As estruturas de dados topológicas oferecem vantagens na execução de tarefas de alto nível semântico; no entanto, para que seu funcionamento seja satisfatório, é necessário prover meios de unir aspectos de baixo nível com a semântica desejada, o que pressupõe o uso de um método de segmentação adequado.

Estruturas de dados hierárquicas são bastante úteis em tarefas de processamento de imagens, pois permitem a criação de algoritmos que decidam uma estratégia de processamento com base em quantidades relativamente pequenas de dados, além de facilitar a paralelização do processamento. Assim, é possível trabalhar em resoluções maiores apenas nas partes da imagem onde isto se faz necessário, usando conhecimento ao invés de força bruta para facilitar e acelerar o processo. As pirâmides estão entre as estruturas hierárquicas mais simples. Uma pirâmide é uma sequência $\{M_L, M_{L-1}, \dots, M_0\}$ de imagens, onde M_L tem as mesmas dimensões e elementos que a imagem original, e M_{i-1} é derivada de M_i reduzindo a resolução pela metade. Quando as pirâmides são criadas, é comum trabalhar com matrizes quadradas de dimensões iguais a potências de 2. Dessa forma, M_0 corresponde a somente um pixel. As pirâmides são utilizadas quando existe a necessidade de se trabalhar com uma imagem em diferentes resoluções simultaneamente. Uma imagem que é um grau de resolução menor em uma pirâmide contém quatro vezes menos dados, de forma que seu processamento é aproximadamente quatro vezes menor. Em alguns casos, ao invés de manter representações de imagens inteiras em uma mesma resolução, pode ser útil representar diferentes partes da imagem em diferentes resoluções. As pirâmides irregulares (IP; LAM, 1996) provêm representações em que a escala é ajustada localmente.

Salembier e Garrido (SALEMBIER; GARRIDO, 2000) propuseram uma forma de representação hierárquica em que as imagens são representadas por árvores de particionamento binário. Cada nodo da árvore corresponde a uma região conexa da imagem, sendo que as folhas correspondem aos pixels da imagem e a raiz corresponde à imagem inteira. O método é inicializado pelas folhas, onde cada pixel corresponde a uma única região. A partir de um método iterativo de segmentação por aglomeração, a cada passo o

par de regiões vizinhas mais semelhantes da imagem é combinado em uma única região, armazenada na árvore como o nodo pai das duas regiões similares, até chegar ao nodo raiz. Esta forma de representação é adequada para a execução de tarefas como filtragem morfológica, análise de imagens e recuperação de informações.

A modelagem de imagens através de uma malha envolve o particionamento do domínio da imagem em um conjunto de pedaços (geralmente poligonais) não-sobrepostos chamados de *elementos de malha*. A função imagem é geralmente determinada (ou aproximada) em cada elemento por interpolação. A modelagem de imagens a partir de malhas provê uma representação compacta e eficiente da imagem sendo uma ferramenta eficaz em aplicações como compressão de imagens, segmentação de imagens, detecção, rastreamento e descrição de objetos, estimativa de movimento em vídeos, análise de elementos finitos, processamento de imagens através de manipulação geométrica, análise de imagens médicas, entre outras. O método de representação proposto neste trabalho é baseado em modelagem de malhas, portanto, será dada uma atenção especial a este tipo de representação.

Conforme a aplicação, diferentes tipos de malhas podem ser utilizados. As malhas podem ser divididas em regulares e irregulares, de acordo com a distribuição espacial dos vértices. As seções 2.3 e 2.4 tratam de métodos baseados em malhas regulares e malhas irregulares, respectivamente. Muitos dos métodos de representação baseados em malhas existentes utilizam a Triangulação de Delaunay ou a Triangulação de Delaunay com Restrições para gerarem suas malhas. Uma atenção especial para estas estruturas é dada na seção 2.2.

2.2 Triangulação de Delaunay e Triangulação de Delaunay com Restrições

Dado um conjunto de pontos $P = \{p_1, p_2, \dots, p_n\}$ no plano, várias triangulações podem ser geradas. Para problemas envolvendo aproximações de superfícies, a qualidade da aproximação depende do tipo de triangulação escolhida. Para problemas envolvendo interpolação, uma boa triangulação é aquela que produz triângulos os mais equiláteros possíveis. Isto porque com triângulos muito alongados e com ângulos muito agudos, a interpolação em pontos internos aos triângulos pode ser influenciada por vértices muito distantes entre si, geralmente degradando a qualidade da aproximação. A Triangulação de Delaunay apresenta a propriedade de minimização dos máximos ângulos internos dos triângulos, e por isso é utilizada neste trabalho.

Outra propriedade da Triangulação de Delaunay, e que é importante no processo de atualização da triangulação, é a seguinte: seja $P = \{p_1, p_2, \dots, p_n\}$ um conjunto de pontos no plano, e seja $T = \{t_1, t_2, \dots, t_m\}$ um conjunto de triângulos cujos vértices sejam os pontos de P . Um triângulo t_i qualquer é dito triângulo de Delaunay se e somente se o círculo que circunscreve este triângulo não contiver nenhum ponto de P em seu interior. A triangulação T é dita Triangulação de Delaunay se e somente se todos os triângulos $t_i \in T$ são triângulos de Delaunay (Figura 2.1).

Além do mais, a Triangulação de Delaunay é única para um dado conjunto de pontos P desde que não existam quatro pontos co-circulares em P . A Figura 2.2 mostra que para

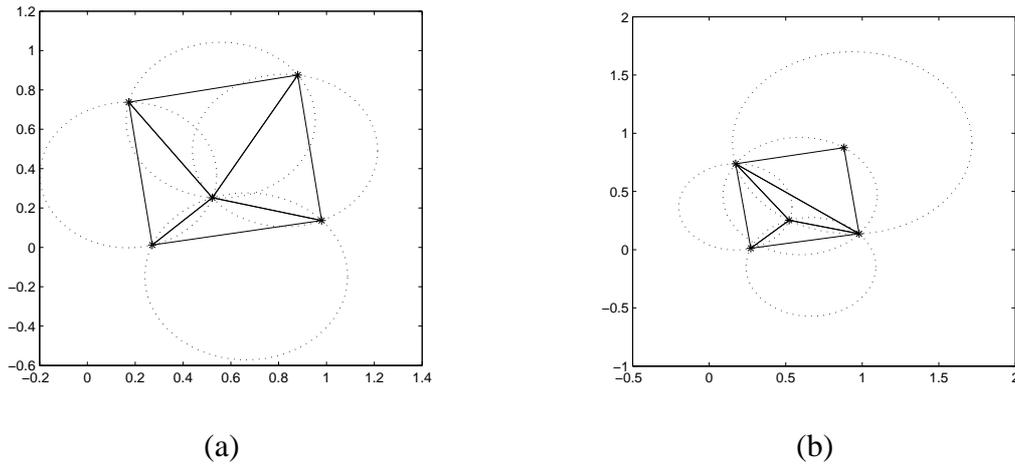


Figura 2.1: Triangulações para um pequeno conjunto de pontos (a) apenas triângulos de Delaunay; (b) nem todos são triângulos de Delaunay.

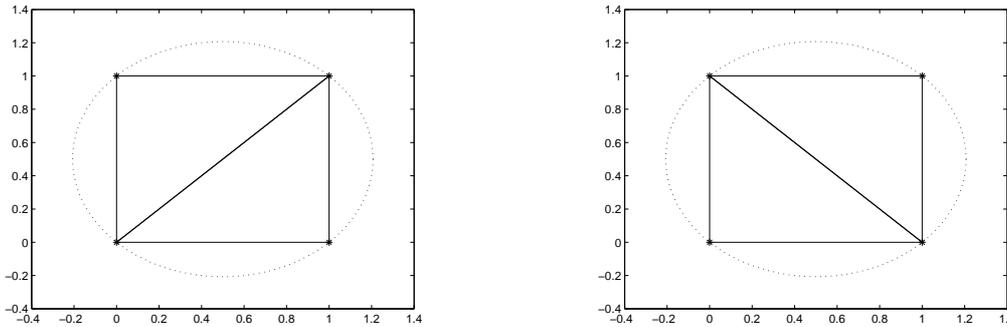


Figura 2.2: Exemplos de duas triangulações de Delaunay válidas para o mesmo conjunto de pontos.

quatro pontos co-circulares, existem duas maneiras de triangulá-los. No entanto, podemos definir uma determinada ordenação do conjunto de pontos, como por exemplo, de cima para baixo e da esquerda para a direita (com relação ao domínio da imagem), e toda a vez que quatro pontos coplanares forem encontrados, selecionamos a triangulação em que os três primeiros dos quatro pontos formam um triângulo. Assim garantimos que, dado um conjunto de pontos, uma triangulação é sempre única, para efeito de reconstrução posterior.

A propriedade de unicidade da Triangulação de Delaunay é especialmente útil ao utilizarmos a representação proposta para a codificação de imagens, já que apenas as posições e os valores dos vértices devem ser armazenados para que a reconstrução da triangulação possa ser feita depois, não sendo necessário codificar qualquer informação de adjacência dos vértices.

Em alguns casos, é desejável que na triangulação resultante algumas conexões entre vértices sejam preservadas (por exemplo, arestas localizadas nas bordas da imagem). Neste caso, não é garantido que uma Triangulação de Delaunay válida possa ser gerada. No entanto, podemos gerar uma triangulação que seja a mais semelhante possível a uma

Triangulação de Delaunay. Seja P um conjunto de pontos, e seja $A = \{a_1, a_2, \dots, a_n\}$ um conjunto de arestas ligando pontos de P , a chamada Triangulação de Delaunay com Restrições é uma triangulação em que todas as arestas de A são mantidas e que os máximos ângulos dos triângulos resultantes sejam minimizados.

Não serão apresentadas aqui os métodos utilizados para a criação e atualização das triangulações. Na literatura é possível encontrar diversos métodos para a criação e atualização eficiente de Triangulações de Delaunay e Triangulações de Delaunay com Restrições (BERG et al., 1997; CHEW, 1987; KOHOUT; KOLINGEROVÁ, 2003; SHEWCHUK, 2003). O método utilizado neste trabalho está descrito em (SHEWCHUK, 2002).

2.3 Malhas Regulares

As malhas regulares envolvem o particionamento do domínio da imagem em elementos de malha poligonais com topologia uniforme.

As malhas regulares normalmente são utilizadas na representação de imagens como estruturas de subdivisão hierárquica, onde a resolução da malha varia adaptativamente em diferentes regiões da imagem, conforme estão dispostas as regiões de interesse. As malhas regulares são comumente aplicadas na análise de imagens, estimativa de movimento, casamento de regiões, análise de elementos finitos, etc.

Uma abordagem muito utilizada na representação de imagens através de malhas é partir de uma malha regular e, adaptativamente, ajustar a malha a características locais da imagem. Esta abordagem foi utilizada em (SALEMBIER; MARTÍ; PARDÀS, 1996) em um esquema de codificação de vídeos com altas taxas de compressão, para a representação das regiões e estimativa de movimento. Na abordagem apresentada, o método parte de uma malha retangular regular e os vértices são então deslocados para pontos com altos valores de gradientes. Este tipo de abordagem, embora consiga atingir taxas muito altas de compressão, não é capaz de representar adequadamente pequenos detalhes da imagem, deteriorando a qualidade da imagem percebida.

Em (HSU; LIU; CHEN, 2001) é proposto um método de codificação de vídeos com altas taxas de compressão que utiliza malhas retangulares regulares para a compensação de movimento entre quadros consecutivos e para a modelagem de regiões a partir de interpolação adaptativa. São utilizados vários padrões de interpolação pré-definidos, e a partir de informação de bordas e movimento extraída localmente, são definidos quais os padrões devem ser utilizados em cada elemento de malha de forma a minimizar os artefatos retangulares resultantes.

2.4 Malhas Irregulares

Ao contrário da representação por meio de malhas regulares, as malhas irregulares podem conter elementos de tamanhos e formas variados. Embora costumem ser computacionalmente mais custosas do que as técnicas de divisão por malhas regulares, as malhas irregulares fornecem uma representação mais compacta da disposição espacial da imagem, sendo mais comumente utilizadas em tarefas onde a codificação eficiente é desejada.

Uma das grandes vantagens que a representação baseada em malhas provê é a facilidade em tratar o problema da compensação de movimento, utilizando apenas os vértices da malha como referência. Um método para a construção de malhas triangulares irregulares, visando a compensação de movimento de objetos em um vídeo, é apresentado em (KONYHA et al., 2003). A eficiência é um ponto forte deste método, onde a seleção dos vértices é feita de forma rápida e adaptativa utilizando informação de um gradiente espaço-temporal. O conjunto de vértices é inicializado com os quatro cantos da imagem (ou quatro pontos na fronteira de um objeto, caso o objetivo seja a representação de objetos individuais) e a inserção de vértices ocorre de forma adaptativa. Nas regiões com valores mais altos de gradiente, são escolhidos mais pontos. Cada vez que um ponto é inserido no conjunto de vértices, o triângulo que contém o ponto é localizado e cada aresta do triângulo é ligada ao novo ponto formando três novos triângulos. Ao final da inserção dos vértices, um processo de otimização da malha é utilizado para evitar triângulos muito alongados, através do giro (*flip*) de arestas.

Uma questão essencial na construção de malhas irregulares é a escolha dos pontos que formam os vértices da malha. A escolha de todos os pontos da malha em apenas um passo é uma tarefa difícil e, dependendo da abordagem utilizada, nem sempre tem solução ótima. Por outro lado, métodos iterativos de seleção de pontos podem ser muito custosos computacionalmente. Em (YANG; WERNICK; BRANKOV, 2003) é proposto um método de representação de imagens em tons de cinza através da Triangulação de Delaunay. Segundo os autores, os diferentes elementos de malha são aproximados por interpolação linear, e a seleção dos vértices da triangulação é feita em apenas um passo, fazendo com que a concentração de pontos seja maior nas regiões da imagem que apresentam maiores frequências e a concentração é menor em regiões suaves e homogêneas da imagem. Para isso, o algoritmo clássico de difusão de erro Floyd-Steinberg é aplicado no gradiente da imagem original, gerando mais pontos onde o gradiente é maior e menos pontos onde é menor. Esta técnica é muito útil para aplicações em que o tempo de processamento é um fator crítico. No entanto, embora seja fácil controlar a quantidade de informação codificada (número de vértices), é difícil controlar a qualidade da codificação gerada; isto é, a qualidade da aproximação em relação à imagem original. Esta mesma técnica, adaptada a imagens vetoriais (a cores ou multi-espectrais) é apresentada em (BRANKOV; YANG; WERNICK, 2003).

Outro método baseado na Triangulação de Delaunay e na interpolação linear dos elementos de malha é apresentado em (ZHAOYANG et al., 2001). Os autores propuseram duas técnicas iterativas de seleção de vértices. A primeira visa minimizar a diferença absoluta entre a imagem original e a aproximada (interpolada a partir dos vértices) escolhendo a cada passo o ponto que gera o maior erro em toda a imagem. A segunda técnica tenta minimizar a variância da cor dos pixels dentro de um mesmo elemento de malha, escolhendo a cada passo o centróide do triângulo que apresenta a maior variância. Este método é bastante semelhante ao apresentado em (DUDON; AVARO; ROUX, 1997), que visa estimativa de movimento em vídeos. Neste último, os critérios para seleção dos vértices são a qualidade da aproximação e o movimento entre dois quadros consecutivos; assim, são escolhidos pontos que ou apresentam movimento de um quadro para outro, ou geram um grande erro de aproximação por interpolação.

Beek et al. (BEEK et al., 1999) propuseram um método de representação hierárquica baseado na Triangulação de Delaunay. O resultado é uma hierarquia de malhas M_l , com níveis $l = 0, 1, \dots, L$, onde L denota a escala mais grossa da triangulação (menos vértices) e M_0 denota a malha em resolução completa. A construção desta hierarquia parte da escala mais fina (M_0), com o número máximo de vértices e, a cada passo, os vértices menos significativos (localizados em regiões de poucos detalhes da imagem) são retirados da malha, simplificando-a cada vez mais. O processo de codificação da hierarquia se dá na ordem inversa; ou seja, primeiramente são codificados os vértices da escala L e, em seguida são adicionados os vértices da escala $L - 1, L - 2, L - 3$, e assim por diante, juntamente com rótulos indicando em qual escala o vértice em questão foi codificado. Esta representação provê uma forma de transmissão progressiva da geometria de objetos. No entanto, para casos onde a escalabilidade não é necessária, este método se mostra demasiadamente custoso.

3 SEGMENTAÇÃO DE IMAGENS: CONCEITOS E ESTADO DA ARTE

3.1 Introdução

Segmentação de imagens é o processo de agrupamento de pixels de uma dada imagem em regiões homogêneas em relação a certos atributos, e com um possível conteúdo semântico. Normalmente, é o primeiro passo em tarefas de análise de imagens e reconhecimento de padrões.

Não é eficiente operar diretamente no nível de representação baseada em pixels de uma imagem para tarefas de alto-nível, tais como a detecção de objetos em uma cena. Regiões podem ser as unidades básicas em uma representação eficiente da cena. Representações precisas de regiões de uma cena provêm vantagens (LUO; GUO, 2003) como:

1. Um número tratável de regiões de forma a tornar tarefas de análise de alto-nível possíveis e eficientes;
2. Medidas precisas de certas propriedades, como tamanho e forma;
3. Uma descrição precisa da localização bem como o relacionamento espacial entre os objetos;
4. Uma medida precisa de atributos de cor e textura dos objetos.

A segmentação de imagens pode ser definida formalmente como (PAL; PAL, 1993): seja $P()$ um predicado de homogeneidade definido em grupos de pixels conexos, então segmentação é uma partição do conjunto F em sub-conjuntos conexos ou regiões (S_1, S_2, \dots, S_n) , tais que:

$$\bigcup_{i=1}^n S_i = F, \text{ com } S_i \cap S_j = \emptyset \text{ (} i \neq j \text{)}$$

O predicado de homogeneidade $P(S_i)$ é verdadeiro para todas as regiões S_i , e $P(S_i \cup S_j)$ é falso quando $i \neq j$ e S_i e S_j são vizinhos.

Porém, podemos ter várias definições diferentes do que pode ser caracterizado como uma região de uma imagem segmentada. Diferentes métodos recaem em diferentes definições. De acordo com os métodos de segmentação existentes, podemos ter quatro definições principais de regiões (SKARBEEK; KOSCHAN, 1994):

1. Região é um componente conexo de um conjunto de pixels especificados por uma função de agrupamento de classe definida em um espaço de cores;
2. Região é um conjunto conexo de pixels no qual a condição de uniformidade é satisfeita;
3. Região é um conjunto conexo de pixels limitados por pixels de bordas que criam contornos;
4. Região corresponde a uma superfície ou um objeto de material homogêneo, ou seja, regiões representam materiais conexos na cena.

De acordo com Fu e Mui (FU; MUI, 1981), o problema da segmentação de imagens é basicamente uma forma de percepção psicofísica e, com isso, não é suscetível a uma solução puramente analítica. Os métodos de segmentação automática atualmente estão muito longe da perfeição. Primeiramente, a segmentação executada por um observador humano envolve o reconhecimento de objetos baseado em modelos gravados de objetos familiares na mente. Ainda, variações de cor e iluminação causam grandes problemas para algoritmos automáticos, mas são efetivamente descontados por humanos. Existem centenas de técnicas propostas na literatura, mas não existe um único método que pode ser considerado bom para todos os tipos de imagens.

Apesar de os métodos de segmentação propostos até o momento estarem muito longe da perfeição, ou ainda, muito longe de gerarem resultados similares aos da segmentação subjetiva executada automaticamente por um humano, deve-se ter em mente que a segmentação normalmente não é o objetivo do processo, mas apenas um passo necessário para servir tarefas de alto-nível na compreensão de imagens. E o propósito da tarefa em questão pesará em relação ao que é considerado um bom método de segmentação.

3.2 Revisão de Técnicas de Segmentação de Imagens

Atualmente, muitas das técnicas de segmentação de imagens a cores existentes são meramente extensões de métodos de segmentação de imagens monocromáticas, implementados em diferentes espaços de cores. Para uma revisão de diferentes espaços de cores existentes, e suas propriedades particulares comumente exploradas pelos métodos de segmentação, ver a Ref. (CHENG et al., 2001). Muitas das técnicas de segmentação de imagens monocromáticas podem ser estendidas para imagens a cores, tais como limitação de histogramas, agrupamento, crescimento de regiões, detecção de bordas, entre outras. Estas técnicas podem ser aplicadas diretamente sobre cada um dos componentes do espaço de cores, e os resultados podem ser combinados de alguma forma para se obter o resultado final da segmentação (KURUGOLLU; SANKUR; HARMANCI, 2001). No entanto, um dos problemas é como empregar a informação de cor com um todo para cada pixel. Quando a cor é projetada em três componentes, a informação de cor se torna dispersa de forma que a imagem colorida se torna simplesmente uma imagem multiespectral e a informação de cor que um humano percebe é perdida.

Nesta seção, uma revisão de algumas técnicas notáveis de segmentação de imagens é apresentada. A revisão bibliográfica que segue não é de forma alguma exaustiva, e serve

principalmente para identificar os temas comuns entre as diversas abordagens, destacando suas vantagens e limitações, fixando o contexto que será apresentado nos capítulos seguintes.

Não é fácil categorizar as diversas técnicas de segmentação de imagens existentes na literatura, nem mesmo existe um esquema de categorização amplamente aceito disponível. Para o propósito desta dissertação, as técnicas de segmentação são divididas em técnicas baseadas na análise de feições (orientadas a regiões, orientadas a bordas, *clustering*, limiarização de histogramas, etc.) e técnicas baseadas em malhas.

3.2.1 Técnicas Orientadas à Análise de Feições

Limiarização de histogramas é uma das técnicas mais amplamente utilizadas para segmentação de imagens monocromáticas. A técnica se baseia na hipótese de que as diferentes regiões de uma imagem são compostas por diferentes faixas de níveis de cinza, de forma que o histograma pode ser separado em um certo número de picos, cada um correspondente a uma região, e existe um valor limiar que corresponde a um vale entre dois picos adjacentes. Para imagens coloridas, a situação é diferente pelos múltiplos atributos de cores. A limiarização de histogramas divide cada uma das componentes do espaço de cores. Existe uma limitação ao dividirmos múltiplas dimensões, pois esta é uma técnica para imagens monocromáticas. Por exemplo, em duas dimensões, só é possível formar agrupamentos retangulares. Como a informação de cor é representada por três coordenadas no espaço de cores, representar a o histograma de uma imagem colorida em um vetor tri-dimensional e selecionar limiares neste histograma não é uma tarefa simples (HARALICK; SHAPIRO, 1985), e detectar agrupamentos de pontos neste espaço normalmente é muito custoso computacionalmente.

Para obter a máxima informação entre modos (regiões com alta densidade) e vales (regiões com baixa densidade), Guo et al (GUO; YU; MA, 1998) adotou um método de limiarização baseado na entropia. A procura pelos modos é decidida por uma estimativa da função densidade de probabilidade multi-modal.

Geralmente, duas ou mais feições características formam um espaço de feições e assume-se que cada classe de regiões forme um agrupamento separado neste espaço. A razão de se utilizar múltiplas feições características para efetuar a segmentação de imagens é que, às vezes, existem problemas que não são solucionáveis com uma feição, mas são solucionáveis com múltiplas feições. As feições características podem ser qualquer feição que possa ser utilizada para o problema da segmentação, tais como valor do nível de cinza de imagens multi-espectrais, histograma de níveis de cinza, média, desvio padrão, textura, etc. Para imagens a cores, o espaço de cores é um espaço de feições natural, e aplicar a abordagem do agrupamento para imagens a cores é uma idéia óbvia, pois as cores tendem a formar agrupamentos no espaço de cores. O agrupamento (*clustering*) tem sido utilizado como uma importante técnica no reconhecimento de padrões por muitos anos. O maior problema que esta técnica sofre é como determinar o número de classes em um esquema de classificação não-supervisionado. Também é crítica a seleção do espaço de cores a ser utilizado.

Em (KURUGOLLU; SANKUR; HARMANCI, 2001) é proposto um método de segmentação de imagens a cores baseado no particionamento de histogramas. Neste caso,

são utilizados subconjuntos das bandas originais. Assim, para o espaço RGB , são computados os histogramas bidimensionais dos subconjuntos RG , RB e GB , e os mesmos segmentados, gerando três mapas de segmentação. Estes mapas são então fundidos em um único, que representa o resultado final da segmentação. Os autores apresentam resultados da segmentação aplicada em subconjuntos de bandas, e resultados do método adaptado aos casos de histogramas 1D e 3D. Os resultados demonstram que a utilização de subconjuntos de bandas apresenta resultados muito parecidos aos obtidos utilizando-se histogramas tridimensionais, sendo bastante superior ao método utilizado em histogramas 1D. Já o tempo de execução do método proposto (histogramas 2D) é apenas duas vezes maior que o método aplicado em histogramas 1D, enquanto o tempo de execução do método aplicado em histogramas 3D é aproximadamente 400 vezes maior que quando aplicado em histogramas 2D. Sendo assim, o método proposto apresenta grande aplicabilidade em casos onde é desejável uma boa relação entre desempenho e qualidade da segmentação. No entanto, o seu método de detecção de picos apresenta a limitação de não ser adaptativo à imagem, devendo ser conhecido *a priori* o número de regiões da imagem.

Em (CHENG; SUN, 2000) é proposto um método de segmentação do espaço de cores, baseado na detecção de picos de histogramas. A idéia de histogramas é generalizada para o domínio da homogeneidade. Na primeira fase do método, regiões uniformes são identificadas através de um método de particionamento de histogramas de homogeneidade. Dada uma imagem cuja componente de luminância possui n tons de cinza, o histograma de homogeneidade é um contador de n pilhas, onde a pilha i armazena a soma do valor de homogeneidade de todos os pixels da imagem com valor de luminância igual a i . Logo após, cada região gerada pelo passo anterior é dividida em sub-regiões de acordo com a análise do matiz dos pixels. Finalmente, a fusão de regiões semelhantes no espaço $L^*a^*b^*$ é efetuada, para evitar a sobre-segmentação. Este método apresenta resultados extremamente compactos no que diz respeito ao número de classes de cor resultantes, sendo uma boa alternativa para a quantização de imagens, onde o número de cores resultantes é selecionado adaptativamente, minimizando a distorção visual percebida. No entanto, seu algoritmo de detecção de picos é constituído de várias heurísticas que o tornam pouco robusto.

O método descrito em (COMANICIU; MEER, 2002) que combina a análise de feições tanto no espaço de cores como no domínio espacial é considerado estado da arte em segmentação de imagens atualmente. É baseado no procedimento *mean shift*, proposto em 1975 por Fukunaga e Hostetler (FUKUNAGA; HOSTETLER, 1975) e esquecido durante muito tempo. Este método é fundamentado na busca pelos modos (regiões de alta densidade) de uma distribuição em um espaço de feições. A busca é efetuada calculando-se para um determinado ponto do espaço de feições uma estimativa do gradiente da densidade no ponto. Assim, o processo iterativamente caminha na direção positiva do gradiente de densidade, até que se alcance um ponto de gradiente nulo, o que indicará a existência de um modo da distribuição. O método é aplicado em toda a imagem e, para cada ponto, é memorizado para qual o ponto de máxima densidade o mesmo converge. Assim, o cluster correspondente a um determinado máximo local será o conjunto de pontos que convergem para ele utilizando-se o *mean-shift*. Após algumas adaptações em relação ao

método original, os autores chegaram a uma forma de aplicação do método em que só é utilizado um parâmetro que define a resolução da análise do espaço de feições. Os autores demonstram que a utilização de um *kernel* normal (gaussiano) para o cálculo da estimativa do gradiente, leva a uma propriedade interessante: o ângulo entre dois vetores *mean shift* consecutivos é sempre menor do que 90° . Isso faz com que durante a busca pelo máximo local, o método realize uma trajetória suave, evitando o zig-zag obtido pela aplicação de outros métodos iterativos. No entanto, em muitos experimentos, é utilizado um *kernel* uniforme, no qual a convergência é garantida em tempo finito, embora os testes quase sempre confirmam que a qualidade dos resultados é melhorada utilizando-se um *kernel* normal. Os autores propõem que o espaço de feições utilizados pelo método seja p -dimensional, sendo $p = d + 2$, onde d é a dimensão do espaço de cores utilizado e p incluirá as dimensões espaciais. A métrica Euclidiana é utilizada tanto no domínio espacial como no espaço de cores. Em (CHRISTOUDIAS; GEORGESCU; MEER, 2002) são propostas melhorias em relação ao descrito em (COMANICIU; MEER, 2002), sendo que a principal modificação é a utilização de informação de bordas para contornar a subsegmentação obtida apenas com o procedimento *mean shift*.

Em (LUO; GUO, 2003) é proposto um método de agrupamento perceptual de regiões de propósito geral, sem utilizar nenhum modelo específico de descrição de objetos. Para tanto, é proposto um modelo probabilístico para o problema da segmentação definindo as regiões como *Markov Random Fields* (MRF). Uma coleção de funções de energia são utilizadas para caracterizar propriedades desejáveis de uma região única, e propriedades de pares de regiões passíveis de serem fundidas. As propriedades de uma região única incluem a sua área, a sua convexidade, a sua compactação e variância de cor. As propriedades de pares de regiões incluem diferença da média das cores, intensidade da borda que separa as duas regiões, variância de cor da área que limita as duas regiões, e a continuidade do contorno das duas regiões.

3.2.2 Técnicas Baseadas em Malhas

Métodos de segmentação de imagens baseados em malhas irregulares não são muito populares em aplicações envolvendo imagens estáticas. Por outro lado, vários métodos de codificação de vídeos se utilizam deste artifício para gerarem representações eficientes de objetos em movimento. Em (GOKÇETEKIN et al., 2000) é proposto um método de segmentação de objetos em vídeos onde no primeiro quadro (ou nos quadros-chave) é feita uma seleção de pontos baseada em conteúdo, e logo após uma malha é gerada com base na Triangulação de Delaunay destes pontos. A cada novo quadro da sequência, é estimado o movimento individual em cada vértice da malha. São consideradas 'regiões de interesse' os triângulos onde pelo menos dois de seus vértices possuem movimentos diferentes, indicando que são áreas de fronteiras entre objetos em movimento relativo. Estas regiões de interesse são refinadas, com a inserção de novos vértices em seus interiores. A malha é atualizada nas regiões de interesse e os movimentos dos novos vértices são calculados a fim de determinar com exatidão as bordas dos objetos.

Hermes e Buhmann (HERMES; BUHMANN, 2003) propuseram um método de segmentação de imagens composto de duas etapas: uma etapa de divisão da imagem original em triângulos, e uma segunda etapa de fusão de triângulos adjacentes, gerando regiões

poligonais. Durante a fase de divisão da imagem, novos vértices são inseridos nas regiões mais ricas em detalhes da imagem, até que seja alcançado um ponto em que mais inserções serviriam apenas para representar o ruído. Para isto são utilizados intervalos de confiança estimados pelo *teroma* de Sanov.

Conforme foi visto no capítulo 2, representações de imagens através de malhas são comumente utilizadas para aproximar regiões através de interpolação linear, tal como é feito com superfícies 3D. Em (RAZDAN; BAE, 2003) é proposto um método de segmentação de malhas 3D baseado em feições (cor, por exemplo) e curvatura da superfície. Dada uma triangulação qualquer, este método subdivide triângulos cujos vértices diferem com relação às suas feições e que formem uma curvatura demasiadamente grande em relação a seus vizinhos.

Malhas triangulares regulares são utilizadas em (VALETTE; MAGNIN; PROST, 2001) para a segmentação de objetos em vídeos. Para segmentar corretamente os contornos dos objetos, a malha se deforma para coincidir seus vértices com as regiões de alto valor de gradiente espaço-temporal. Além disso, é possível refinar a malha a partir de uma estrutura de divisão baseada em *quad-trees*. Tanto a deformação quanto o refinamento da malha são executados de forma a maximizar uma função global de energia, desenvolvida para preencher três requisitos: (1) nenhum elemento de malha deve conter partes de mais de um objeto em seu interior; (2) as arestas da malha devem estar posicionadas preferencialmente em regiões de gradiente intenso e (3) deve haver um bom casamento ao se projetar elementos de malha do quadro atual no anterior.

4 COMPRESSÃO DE IMAGENS: CONCEITOS E ESTADO DA ARTE

4.1 Introdução

A compressão de imagens trata o problema da redução da quantidade de dados necessária para representar uma imagem digital. Tal problema é agravado substancialmente quando se trabalha com imagens em movimento (vídeo digital) e em ambos os casos requer a ação de algoritmos e técnicas que permitam a redução da quantidade de dados necessária para representar uma imagem ou sequência de imagens. Aplicações de compressão de dados estão principalmente em transmissão e armazenamento de informação. A compressão de dados também é útil no desenvolvimento de algoritmos rápidos onde o número de operações necessárias para implementar um determinado algoritmo é reduzido ao se trabalhar com os dados comprimidos (JAIN, 1989).

A base do processo de redução é a remoção de dados redundantes. Do ponto de vista matemático, isto corresponde a transformar uma matriz de pixels de duas dimensões num conjunto de dados estatisticamente descorrelacionado (GONZALEZ; WOODS, 1992). A transformação é aplicada antes do armazenamento ou transmissão da imagem. Posteriormente, a imagem comprimida é descomprimida para reconstruir a imagem original ou uma aproximação dela. O interesse pelas técnicas de compressão de imagens começou há quase meio século (na época, utilizando técnicas analógicas) e é cada vez maior, graças à popularização da multimídia. As técnicas de compressão de imagens costumam ser divididas em dois grandes grupos: aquelas nas quais toda a informação original é preservada (chamadas de técnicas de compressão sem perdas) e aquelas em que ocorre uma perda tolerável de informação, com o objetivo de atingir maiores taxas de compressão (técnicas com perdas).

O restante deste capítulo está organizado da seguinte forma: a seção 4.2 concentra-se em explicar o conceito de redundância e como ela pode ser explorada pelos algoritmos de compressão de imagens, além de apresentar os critérios de fidelidade de imagens utilizados no restante deste trabalho. A seção 4.3 apresenta uma revisão de métodos de compressão sem perdas encontrados na literatura, enquanto uma revisão de métodos de compressão com perdas é apresentada na seção 4.4.

4.2 Fundamentos

4.2.1 Redundância de Dados

A redundância de dados é um aspecto importante no estudo de compressão de imagens digitais. Para quantificá-la matematicamente, suponhamos que n_1 e n_2 representam o número de unidades portadoras de informação em dois conjuntos de dados que representam a mesma informação. Neste caso, a redundância relativa (R_D) do primeiro conjunto de dados (aquele representado por n_1) poderá ser definida como

$$R_D = 1 - \frac{1}{C_R} \quad (4.1)$$

onde o parâmetro C_R , comumente chamado de razão (ou taxa) de compressão, é

$$C_R = \frac{n_1}{n_2} \quad (4.2)$$

Quando $n_2 = n_1$, $C_R = 1$ e $R_D = 0$, podemos concluir que o primeiro conjunto de dados não contém nenhum dado redundante em relação ao segundo. Quando $n_2 \ll n_1$, $C_R \rightarrow \infty$ e $R_D \rightarrow 1$, haverá significativa compressão de dados altamente redundantes. Finalmente, no caso em que $n_2 \gg n_1$, $C_R \rightarrow 0$ e $R_D \rightarrow -\infty$, podemos concluir que o segundo conjunto de dados contém muito mais dados do que o primeiro, representando, obviamente, o caso da expansão de dados, normalmente indesejado. Em geral, C_R e R_D situam-se nos intervalos abertos $(0, \infty)$ e $(-\infty, 1)$, respectivamente.

Na compressão de imagens digitais, três tipos de redundância de dados podem ser identificadas e exploradas: redundância de codificação, redundância interpixel, e redundância psicovisual (GONZALEZ; WOODS, 1992). A compressão de dados é efetivamente obtida quando uma ou mais dessas redundâncias são reduzidas ou eliminadas. Geralmente, os métodos de compressão sem perdas obtêm ganhos nas taxas de compressão ao eliminar a redundância interpixel, enquanto a redundância psicovisual é normalmente abordada nos métodos de compressão com perdas. A redundância de codificação é utilizada em todos os tipos de métodos.

Redundância de Codificação

Consideremos que uma variável aleatória discreta r_k no intervalo $[0, 1]$ representa os níveis de cinza de uma imagem e que cada r_k ocorre com probabilidade $p_r(r_k)$. Temos que

$$p_r(r_k) = \frac{n_k}{n} \quad k = 0, 1, \dots, L - 1 \quad (4.3)$$

onde L é o número dos níveis de cinza, n_k é o número de vezes em que o nível de cinza k aparece na imagem, e n é o número total de pixels na imagem. Se o número de bits utilizado para representar cada valor r_k é $l(r_k)$, a quantidade média de bits exigida para representar cada pixel é

$$L_{avg} = \sum_{k=0}^{L-1} l(r_k) p_r(r_k) \quad (4.4)$$

Em resumo, o comprimento médio das palavras-código atribuídas aos diversos valores de tom de cinza é calculado através da soma do produto do número de bits utilizados, para

representar cada nível de cinza, pela probabilidade de ocorrência daquele nível. Assim, o número total de bits exigido para codificar uma imagem de dimensões $M * N$ é MNL_{avg} . A representação dos níveis de cinza de uma imagem com um código binário natural de m bits reduz o lado direito da equação 4.4 para m bits. Em outras palavras, $L_{avg} = m$ quando $l(r_k)$ for substituído por m na equação 4.4.

Redundância Interpixel

Existem situações em que uma imagem apresenta pixels fortemente correlacionados, sendo estas correlações decorrentes de uma relação estrutural ou geométrica entre os objetos que a compõem. Pelo fato de o valor de qualquer pixel de uma imagem poder ser razoavelmente predito a partir dos valores de seus vizinhos, a informação contida por pixels individuais é relativamente pequena. A maior parte da contribuição visual de um simples pixel para uma imagem é redundante; ela poderia ter sido predita com base nos valores de seus vizinhos. As expressões 'redundância espacial', 'redundância geométrica', e 'redundância entre quadros (*interframe*)' são utilizadas para indicar estas dependências.

A fim de reduzir as redundâncias interpixel em uma imagem, o arranjo bidimensional de pixels, normalmente utilizado para a visualização e interpretação, deve ser transformado em um formato mais eficiente (mas, geralmente, não visualizável), por exemplo, utilizando as diferenças entre os pixels adjacentes para representar uma imagem. Transformações capazes de remover a redundância interpixel são conhecidas como mapeamentos. Estes mapeamentos são ditos reversíveis se os elementos da imagem original puderem ser reconstruídos a partir do conjunto de dados transformados (sem perdas), e irreversíveis caso contrário (com perdas).

Redundância Psicovisual

Existem inúmeras experiências capazes de comprovar o fato de que o olho humano não responde com igual sensibilidade a toda informação visual que recebe (GONZALEZ; WOODS, 1992). Certas informações possuem menor importância relativa do que outras, no processo visual normal. Estas informações menos importantes podem ser consideradas redundantes do ponto de vista psicovisual e, portanto, podem ser eliminadas sem prejudicar significativamente a qualidade da imagem percebida pelo sistema visual humano.

A redundância psicovisual é fundamentalmente diferente das redundâncias anteriormente discutidas. Ao contrário das redundâncias de codificação e interpixel, a redundância psicovisual é associada a informações visuais quantificáveis ou reais. Sua eliminação é possível apenas pelo fato de a informação propriamente dita não ser essencial para o processamento visual normal.

Considerando que a eliminação de dados psicovisualmente redundantes resulta em uma perda de informação quantitativa, a mesma é comumente chamada de quantização. Esta terminologia é consistente com o uso normal da palavra, a qual geralmente significa o mapeamento de uma ampla faixa de valores de entrada para um número limitado de valores de saída. Como esta é uma operação irreversível (a informação visual é perdida), a quantização resulta em uma compressão de dados com perdas.

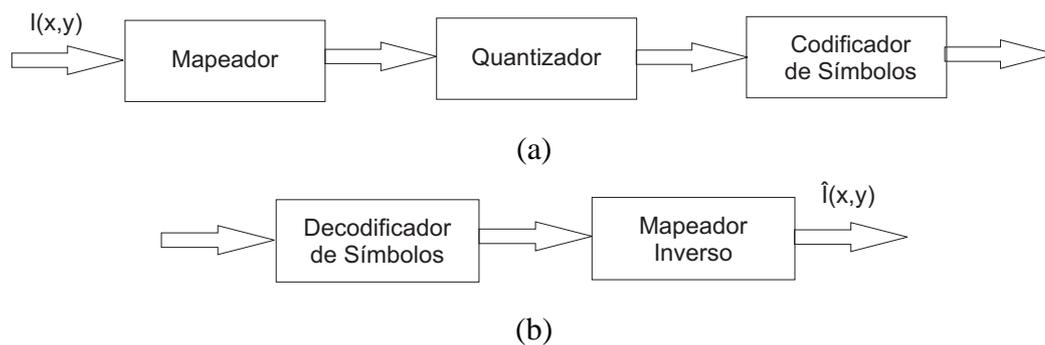


Figura 4.1: Diagramas em blocos: (a) codificador e (b) decodificador.

4.2.2 Codificador e Decodificador

Um sistema de codificação/compressão genérico consiste de dois blocos estruturais distintos: um codificador e um decodificador. O codificador é responsável pela redução ou eliminação de qualquer redundância (de codificação, interpixel, ou psicovisual) presente na imagem de entrada. Sua função normalmente pode ser modelada por uma série de três operações independentes. Como mostra a Figura 4.1(a), cada operação está projetada para reduzir uma das três redundâncias descritas na subseção anterior. A Figura 4.1(b) representa o decodificador de fonte correspondente.

No primeiro estágio do processo de codificação, o mapeador transforma os dados de entrada reduzindo as redundâncias interpixel da imagem de entrada. Esta operação geralmente é reversível e pode, ou não, reduzir diretamente a quantidade de dados exigidos para representar a imagem.

O segundo estágio, ou bloco quantizador na Figura 4.1(a), reduz a precisão de saída do mapeador de acordo com alguns critérios de fidelidade preestabelecidos. Este estágio reduz as redundâncias psicovisuais da imagem de entrada e as alterações que promove no sinal são irreversíveis. Portanto, este bloco deve ser omitido, quando se desejar uma compressão sem perdas.

No terceiro e último estágio de codificação, aparece o codificador de símbolos, responsável por produzir uma palavra-código de comprimento fixo ou variável para representar cada saída do quantizador. Ao final desta etapa, a imagem não deve apresentar qualquer tipo de redundância.

O decodificador apresentado na Figura 4.1(b) contém apenas dois blocos componentes: um decodificador de símbolos e um mapeador inverso. Estes blocos realizam, em sequência oposta, as operações inversas dos blocos mapeadores e codificadores de símbolos do codificador.

4.2.3 Critérios de Fidelidade

A necessidade de obtenção de maiores taxas de compressão, aliada à exploração adequada de limitações e peculiaridades do sistema visual humano, permite a elaboração de técnicas de compressão de imagens nas quais ocorre uma perda de informação visual quantitativa ou real. Considerando que informações de interesse podem ser perdidas, torna-se desejável quantificar a natureza e a extensão da perda de informação. Dois gru-

pos gerais de critérios são utilizados como base para tal análise: (1) critérios de fidelidade objetiva e (2) critérios de fidelidade subjetiva.

Quando o nível de perda de informação puder ser expresso como uma função da imagem original, ou imagem de entrada, e da imagem de saída, comprimida e descomprimida, subsequentemente, diz-se que este baseia-se em um critério de fidelidade objetiva. Um bom exemplo é o erro médio quadrático (MSE) entre a imagem original e a processada. Seja $f(x, y)$ a imagem de entrada e seja $f'(x, y)$ a estimativa ou aproximação de $f(x, y)$ resultante de sua compressão e subsequente descompressão. Para qualquer valor de x e y , o erro $e(x, y)$ entre $f(x, y)$ e $f'(x, y)$ pode ser definido como

$$e(x, y) = f'(x, y) - f(x, y) \quad (4.5)$$

de forma que o erro total entre as duas imagens é

$$\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f'(x, y) - f(x, y)] \quad (4.6)$$

sendo as imagens de tamanho $M * N$. O erro médio quadrático, e_{mse} , entre $f(x, y)$ e $f'(x, y)$ pode ser obtido por:

$$e_{mse} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f'(x, y) - f(x, y)]^2 \quad (4.7)$$

Outro critério de fidelidade objetiva existente é a relação sinal-ruído (SNR) entre a imagem comprimida e a descomprimida, definida em decibéis (dB) e dada por (JAIN, 1989):

$$SNR = 10 \log_{10} \frac{\sigma^2}{e_{mse}} \quad (4.8)$$

onde σ^2 é a variância da imagem original.

Outra definição relação sinal-ruído, usada comumente em aplicações de codificação de imagens é o PSNR (Peak Signal to Noise Ratio), dada por:

$$PSNR = 10 \log_{10} \frac{(d_p)^2}{e_{mse}} \quad (4.9)$$

onde d_p é a distância de pico a pico da imagem original (do máximo ao mínimo).

Embora os critérios de fidelidade objetiva ofereçam mecanismos simples e convenientes para se avaliar a perda de informação decorrente da compressão, na maioria das vezes as imagens comprimidas e descomprimidas são observadas por seres humanos ao final do processo. O sistema visual humano possui características peculiares, a ponto de duas imagens contendo a mesma quantidade de erro MSE poderem ser avaliadas como se possuíssem qualidade visual completamente distinta. Uma destas características é a maior sensibilidade a erros em áreas escuras da imagem e nas regiões de bordas. Para aplicações em que a qualidade visual de uma imagem é mais relevante, duas formas de avaliação podem ser utilizadas: avaliação subjetiva por observadores humanos ou avaliações objetivas que modelem a resposta do sistema visual humano.

4.3 Compressão sem perdas

Nesta seção são apresentadas algumas das principais técnicas de compressão de dados sem perdas, juntamente com métodos que constituem o estado da arte em compressão de imagens sem perdas, também chamada de 'compactação'.

No caso específico de compressão de imagens, convém notar que, em diversas circunstâncias práticas, a compressão deve obrigatoriamente ocorrer sem perdas. Este normalmente é o caso da compressão aplicada a imagens médicas, documentos legais, sensoriamento remoto, etc.

As chamadas técnicas de compressão de dados com palavras-código de comprimento variável são comumente utilizadas no processo de codificação de símbolos, conforme mostrado na Figura 4.1. Entre os métodos de compressão de dados utilizando palavras-código de comprimento variável, estão a Codificação de Huffman (HUFFMAN, 1952), a Codificação Aritmética (ABRAMSON, 1963) e a Codificação LZW (*Lempel-Ziv-Welch*) (WELCH, 1984). Embora possam ser utilizados a partir da aplicação direta na imagem de entrada, estes algoritmos são amplamente utilizados em diversos métodos de compressão de imagens, com ou sem perdas, como etapa final do processo de compressão, para eliminar a redundância de codificação de símbolos.

Uma forma muito utilizada de se obter compressão sem perdas em imagens de tons de cinza ou a cores é a partir da decomposição da imagem em planos de bits (*bit planes*). Dessa forma, uma imagem com 8 bits de cor, é decomposta em 8 planos, cada plano sendo representado por uma imagem binária. A compressão é atingida ao se aplicar, em cada plano, um dos inúmeros métodos de compressão de imagens binárias, tais como: codificação de área constante, codificação de contornos (GONZALEZ; WOODS, 1992) e códigos de corridas (*Run-Length Encodings*) (GOLOMB, 1966). O padrão JBIG (*Joint Bi-level Image Experts Group*) (ISO/IEC, 1993), para compressão de imagens binárias, foi utilizado por Abdal e Bellanger (ABDAT; BELLANGER, 1994) através da aplicação em planos de bits. Ao invés de representar os planos segundo o código natural binário, os autores utilizaram uma representação baseada no Código de Gray (*Gray Code*), também chamado de Código Refletido, para aumentar a eficiência do método. A importante propriedade do Código de Gray é que a transição de um valor V para um valor $V - 1$ ou $V + 1$ necessita apenas da mudança de um único bit. Um exemplo do Código de Gray de 3 bits é mostrado na Tabela 4.1. A propriedade de continuidade do Código de Gray é importante para a compressão de imagens, pois tende a aumentar a correlação de bits na representação digital. Geralmente, em imagens naturais, muitas regiões da imagem apresentam mudanças de pixels lentas e suaves. Em regiões aproximadamente uniformes, representadas pelo Código de Gray, muitos bits de um plano de bits são idênticos e as técnicas de compressão de dados podem ser aplicadas de forma mais eficiente, visto que há maior redundância interpixel neste caso.

As chamadas técnicas preditivas de codificação, amplamente utilizadas em compressão de imagens, partem da idéia básica de remoção de redundância entre pixels próximos, extraíndo e codificando apenas a informação nova trazida por cada pixel. Esta informação nova é normalmente definida como a diferença entre o valor real do pixel e o valor predito para aquele pixel.

Tabela 4.1: Código de Gray de 3 bits

Código Binário Natural	Código de Gray
000	000
001	001
010	011
011	010
100	110
101	111
110	101
111	100

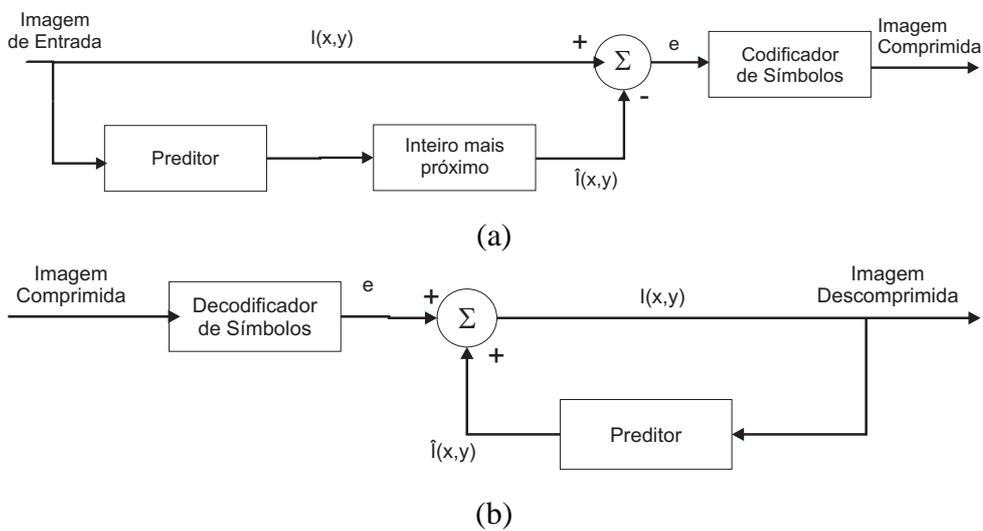


Figura 4.2: Modelo de codificação preditiva: (a) codificador e (b) decodificador.

A Figura 4.2 mostra os principais componentes de um codificador preditivo sem perdas e seu respectivo decodificador. Em ambos os blocos, aparece um elemento-chave, que é o estágio preditor. À medida que cada pixel da imagem de entrada, indicado pela notação f_n , é introduzido no codificador, o preditor gera um número, que é o valor previsto para aquele pixel, com base em entradas anteriores. A saída do preditor é então arredondada para um valor inteiro, a que denominaremos \hat{f}_n , utilizado para calcular o erro de predição, dado por

$$e_n = f_n - \hat{f}_n \quad (4.10)$$

que é então codificado por um código de comprimento variável (pelo codificador de símbolos do estágio codificador), gerando o próximo elemento do conjunto de dados comprimidos. O decodificador da Figura 4.2(b) reconstrói e_n a partir das palavras-código de comprimento variável recebidas e executa a operação inversa

$$f_n = e_n + \hat{f}_n \quad (4.11)$$

Diversos métodos locais, globais e adaptativos podem ser usados para gerar \hat{f}_n .

Um conhecido exemplo de método de decorrelação preditivo sem perdas é o preditor causal encontrado no padrão JPEG (*Joint Photographic Experts Group*) para compressão

sem perdas (PENNEBAKER; MITCHELL, 1993). Outro exemplo é o método CALIC (WU; MEMON, 1997), o qual utiliza um preditor não-linear que se adapta às estatísticas da imagem original, e ajusta automaticamente a ordem do preditor. O mesmo método, com algumas melhorias, é apresentado em (DENG et al., 2003). Preditores adaptativos também são o princípio básico no método descrito por Baligar et al. (BALIGAR; PATNAIK; NAGABHUSHANA, 2003), onde a imagem de entrada é dividida em blocos de tamanhos iguais, e o melhor preditor é estimado para cada bloco. A imagem erro é então comprimida utilizando uma representação por *quad-trees*.

Os métodos preditivos citados acima tendem a codificar os dados originais pixel a pixel, ou em pequenas sequências de pixel. Métodos preditores mais flexíveis, que tratam de blocos de pixels com diferentes tamanhos e formas, selecionados de acordo com características da imagem tendem a ser mais eficientes. Os métodos de compressão sem perdas baseados em segmentação de imagens partem deste pressuposto, e utilizam o resultado da segmentação como preditor dos pixels de uma determinada região.

Recentemente, os métodos baseados em segmentação tem apresentado os melhores resultados, especialmente se tratando de compressão sem perdas de imagens médicas (RATAKONDA; AHUJA, 2002; SHEN; RANGAYYAN, 1997). A idéia por trás destes métodos é segmentar uma imagem em regiões homogêneas, gerar um modelo para as regiões e fronteiras entre regiões, de forma a funcionarem como um preditor. Shen e Rangayyan (SHEN; RANGAYYAN, 1997) propuseram uma abordagem baseada em um método de segmentação por crescimento de regiões onde os modelos são dados pelos bits mais significativos dos pixels sementes. Um mapa de índices de descontinuidade, representando as fronteiras, e uma erro residual de aproximação são obtidos do processo de segmentação. A representação final da imagem é formada pelos bits mais significativos dos pixels semente, pelo erro residual, e pelo mapa de índices de descontinuidade, ambos codificados pelo método JBIG. Ratakonda e Ahuja (RATAKONDA; AHUJA, 2002) propuseram um método de compressão sem perdas baseado em um método de segmentação hierárquico. Este trabalho discute como identificar a escala ideal de segmentação para a codificação, e propõe um modelo de regiões no qual pixels de bordas e pixels internos de regiões são descritos separadamente. Embora métodos de compressão baseados de segmentação possam aproximar bem a imagem original, produzindo representações compactas das regiões, eles tendem a gerar um considerável *overhead* na codificação de fronteiras.

Ultimamente, uma nova classe de técnicas tem gerado grande interesse entre os pesquisadores. São os métodos de compressão de imagens aproximadamente sem perdas. Com estes métodos, não é possível garantir a reconstrução exata da imagem original. No entanto, um limite rígido na magnitude do erro garante que o máximo erro absoluto não ultrapasse um valor pré-definido. A abordagem apresentada em (PINHO, 2001) baseia-se em codificar regiões homogêneas da imagem, obtidas através da quantização da imagem original em um número reduzido de cores. O resíduo gerado (erro de quantização) é então codificado pelo método DPCM.

4.4 Compressão com perdas

Diferentemente das técnicas livres de erro, apresentadas na seção anterior, a codificação com perdas baseia-se no conceito de comprometimento da precisão da imagem reconstruída em troca de uma maior compressão. Se a distorção resultante for tolerável, o aumento na compressão poderá ser bastante significativo.

Na seção anterior falamos sobre técnicas preditivas de codificação para a compressão de imagens sem perdas. Métodos preditivos também podem ser utilizados na compressão com perdas, ao se acrescentar uma etapa de quantização após o cálculo do erro de predição. Assim, ao invés de o preditor gerar apenas valores discretos, este passa a gerar valores reais, e o erro de predição passa a ser quantizado. Este tipo de técnica é conhecida como Modulação por Codificação Diferencial de Pulsos (*Differential Pulse Code Modulation* - DPCM) (SONKA; HLAVAC; BOYLE, 1999).

As técnicas preditivas operam diretamente nos pixels de uma imagem e, por isso, são chamadas de métodos de domínio espacial. Existem técnicas, por outro lado, onde a eliminação da redundância de uma imagem é feito em outros domínios. Estas são as chamadas técnicas de codificação por transformadas. Na codificação por transformadas, utiliza-se uma operação matemática linear reversível para mapear a imagem dentro de um conjunto de coeficientes, os quais em seguida são quantizados e codificados. Para a maioria das imagens naturais, um número significativo de coeficientes têm pequena magnitude e podem, portanto, ser quantizados (ou mesmo descartados), causando pouca distorção na imagem decodificada. Diversas transformadas matemáticas conhecidas podem ser utilizadas para transformar os dados da imagem, tais como (VIERGEVER; ROOS, 1993): Transformada de Karhunen-Loève (KL), Transformada Discreta de Fourier (*Discrete Fourier Transform* - DFT), Transformada Discreta de Hartley (*Discrete Hartley Transform* - DHT), Transformada Discreta do Cosseno (*Discrete Cosine Transform* - DCT), Transformada de Walsh-Hadamard (*Walsh-Hadamard Transform* - WHT) e Transformada Wavelet Discreta (*Discrete Wavelet Transform* - DWT). Quanto maior a capacidade da transformada de compactar informação em poucos coeficientes, melhor ela será para fins de compressão.

5 PROPOSTAS PARA A REPRESENTAÇÃO GEOMÉTRICA DE IMAGENS

5.1 Modelagem de Imagens por Malhas

Para o presente trabalho, foram utilizadas representações de imagens através de malhas triangulares irregulares, dadas pela Triangulação de Delaunay (BERG et al., 1997) e pela Triangulação de Delaunay com Restrições (CHEW, 1987). Assim, a modelagem da imagem através de uma malha é vista como um problema de aproximação de superfícies, em que alguns vértices são selecionados segundo um critério qualquer e os pontos restantes são aproximados por interpolação. Quando a Triangulação de Delaunay com Restrições é considerada, é utilizado um detector de bordas para extrair as fronteiras entre regiões da imagem. As bordas detectadas são então aproximadas por segmentos de reta que são, por sua vez, utilizados como restrições da triangulação (ou seja, os segmentos de reta serão arestas da triangulação resultante). Dois métodos de aproximação da função imagem nas regiões triangulares resultantes são utilizados, bem como dois métodos de seleção das posições dos vértices da triangulação, conforme a aplicação.

5.1.1 Aproximação de Bordas

Uma das formas de modelagem da imagem através de malhas propostas no presente trabalho utiliza uma Triangulação de Delaunay com Restrições em que as bordas das imagens são mantidas na triangulação resultante. Para isso, é necessário detectar as bordas e aproximá-las por segmentos de retas (arestas da triangulação). Esta sub-seção demonstra como é feito este processo.

Dada uma imagem I , devemos obter um mapa de bordas E , do mesmo tamanho de I , que tenha valor 1 para pixels de borda e valor 0 para os demais pixels. Para detectar as bordas da imagem, é utilizado o método de Canny (CANNY, 1986). Caso a imagem I seja colorida, formada pelas componentes I_R, I_G e I_B (vermelho, verde e azul), é utilizada a luminância I_L da imagem, dada por (JAIN, 1989):

$$I_L = 0.299I_R + 0.587I_G + 0.114I_B$$

Bordas com um pixel de largura são obtidas ao se procurar pelos máximos locais do gradiente G . O gradiente é calculado usando a derivada de um filtro gaussiano aplicado na imagem I (ou componente I_L , para imagens a cores). Após o cálculo dos máximos locais do gradiente, dois limiares de gradientes, thr_{high} e thr_{low} , são utilizados para detectar

pixels de borda fortes e pixels de borda fracos, sendo que pixels de borda fortes são sempre incluídos no mapa de bordas de saída, enquanto pixels de borda fracos são incluídos apenas se estiverem conectados a pixels de borda fortes. Um pixel de borda p_w é dito fraco quando possui um valor de gradiente no intervalo $thr_{low} \leq G(p_w) < thr_{high}$, enquanto um pixel de borda p_s é dito forte caso o valor correspondente do gradiente seja $G(p_s) \geq thr_{high}$. Isto faz com que o detector de bordas seja menos suscetível ao ruído.

Quando os elementos de malha (triângulos) são aproximados por interpolação linear, a utilização de vértices nos pontos de borda de máximo gradiente pode resultar em uma má aproximação da imagem nas regiões vizinhas às bordas. Isto porque pontos de máximo gradiente são regiões de transição, instáveis por natureza, fazendo com que um erro de aproximação seja propagado nas regiões vizinhas às bordas. Uma alternativa para contornar este problema, é utilizar um detector de bordas duplas. Isto pode ser feito utilizando-se um detector de bordas de segunda ordem, e selecionando como borda os pontos de máximo e mínimo locais. Para este propósito, utilizamos o Laplaciano da imagem $I(x, y)$, filtrado por uma gaussiana $g(x, y, \sigma)$, sendo σ o desvio padrão da gaussiana. Este operador é conhecido como LoG (*Laplacian of Gaussian*) e possui a forma (SONKA; HLAVAC; BOYLE, 1999)

$$\nabla^2 [g(x, y, \sigma) * I(x, y)] \quad (5.1)$$

onde $*$ denota a operação de convolução. A Figura 5.1 mostra um exemplo em uma dimensão de como o erro de aproximação se propaga para a vizinhança da borda quando são detectadas bordas simples, e como a aproximação pode ser melhorada ao se detectar bordas duplas. Deve-se reparar que, desta forma, as bordas são modeladas como sendo rampas (*ramp edges*). A mesma abordagem de dois limiares descrita para a detecção de bordas simples é utilizada também para a seleção dos pontos das bordas duplas.

O mapa de bordas E deve ser aproximado por um conjunto de segmentos de retas, que constituirão arestas da Triangulação de Delaunay com Restrições. Para esta tarefa, é empregado o método de aproximação de bordas apresentado em (KASHIMURA; SATO; OZAWA, 1992).

Sejam $\{e_i : i = 1, \dots, n\}$ as n bordas conexas em E , e $\{p_i(j) : j = 1, \dots, m_i\}$ os pixels na i -ésima borda conexa. Cada pixel $p_i(j)$ em e_i deve ter no máximo dois pixels vizinhos em e_i . Caso um pixel $p_i(j)$ possua mais de dois vizinhos, este é um pixel de bifurcação, e a borda conexa e_i deve ser quebrada em duas, no pixel $p_i(j)$, como mostrado na Figura 5.2. Este processo é repetido até que todas os pixels da borda $p_i(j)$ tenham no máximo dois vizinhos em e_i . Após, para cada borda e_i , serão criados conjuntos de segmentos de retas $S_i = \{s_i(1), s_i(2), \dots, s_i(q_i)\}$. Inicialmente, um segmento de reta $S_i(1)$ unindo os dois pontos extremos de e_i (se e_i é uma borda circular, dois pixels vizinhos quaisquer são usados) é computado. O ponto $p_i(j)$, em que a distância em relação a $S_i(1)$ é a maior, é usado para dividir $S_i(1)$ em dois segmentos, e o mesmo procedimento é recursivamente aplicado aos novos segmentos, até que a maior distância seja menor que um limiar thr_{dist} , conforme mostrado na Figura 5.3.

Ao final do processo de aproximação de bordas, temos um conjunto de segmentos de retas $S = \{S_1, S_2, \dots, S_n\}$ que irão formar arestas da triangulação. O conjunto N de vértices da triangulação é inicializado com os pontos extremos dos segmentos de S . A Triangulação de Delaunay com Restrições é então computada, gerando um conjunto

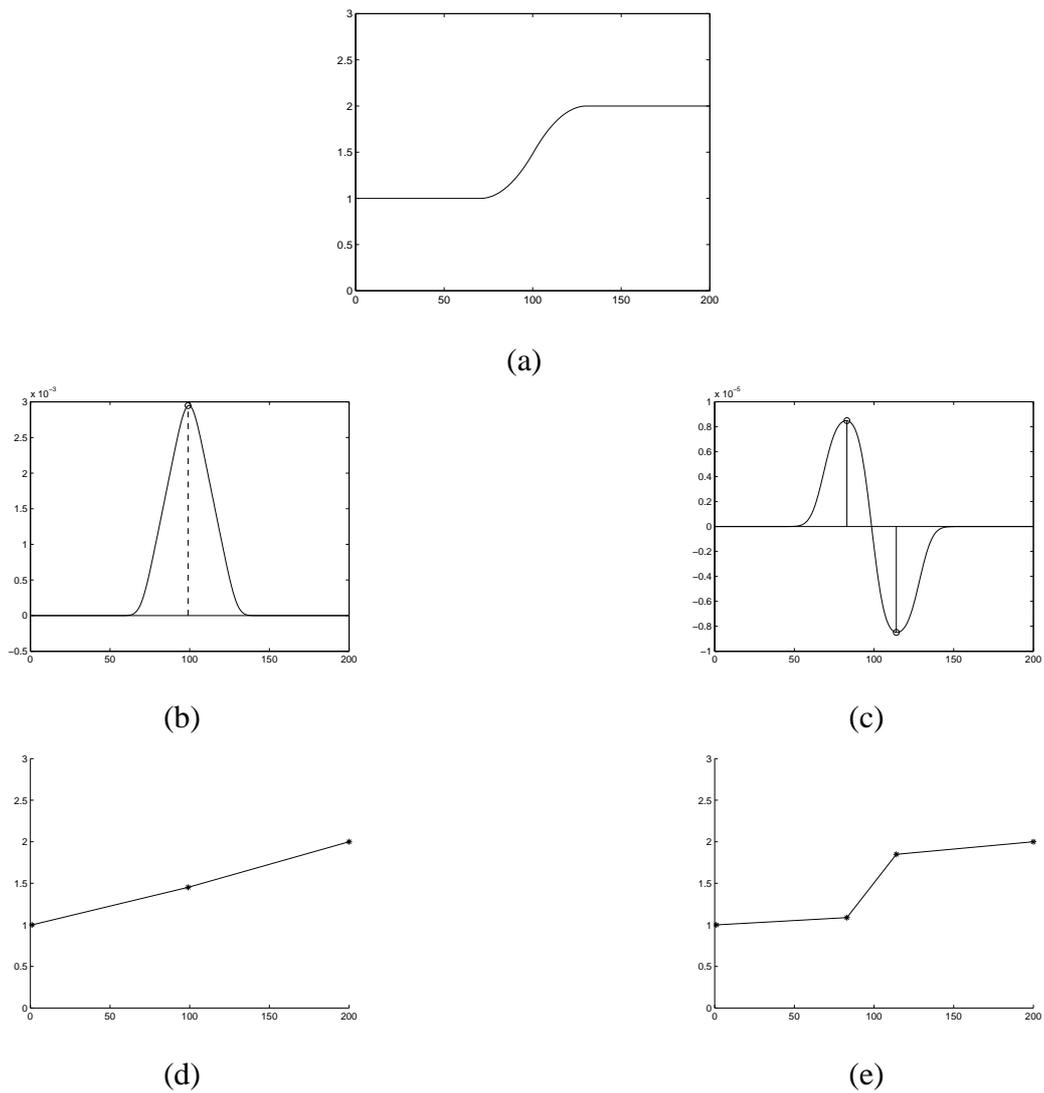


Figura 5.1: Aproximação de regiões por detecção de bordas (a) original; (b) borda simples (detector de primeira ordem); (c) borda dupla (detector de segunda ordem); (d) aproximação com base no máximo do gradiente; (e) aproximação com base nos máximos absolutos do laplaciano.



Figura 5.2: Exemplo de uma borda conexa com uma bifurcação (a) o pixel com mais de dois vizinhos é localizado; (b) a borda é quebrada em duas.

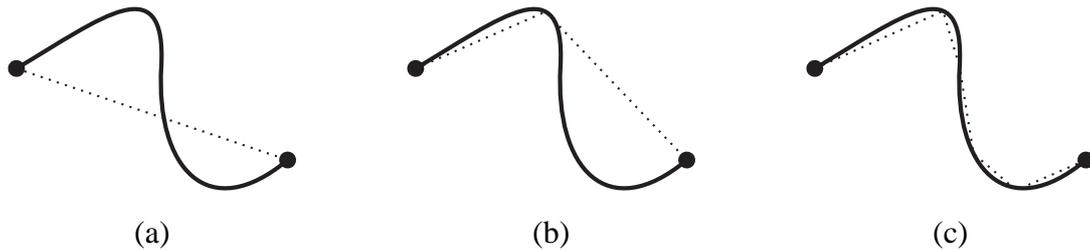


Figura 5.3: Aproximação de uma borda por segmentos de retas (a) inicialização; (b) o segmento é dividido em dois no ponto mais distante; (c) final do processo.

$D = \{T_1, T_2, \dots, T_t\}$ de triângulos. Um exemplo desta triangulação é mostrado na Figura 5.4.

5.1.2 Aproximação de Regiões

Conforme dito anteriormente, dada uma divisão triangular da imagem, as regiões triangulares são aproximadas por interpolação. Muitas formas de interpolação podem ser obtidas, como interpolação linear, *splines*, entre outras. Duas formas de aproximação são utilizadas no presente trabalho: interpolação linear e aproximação constante de regiões. A primeira é utilizada na modelagem de malhas visando a compressão. Assim, a imagem é vista como uma superfície 2 1/2 D (ou campo de alturas) a ser modelada, onde a altura dos vértices é dada pelas cores dos pixels correspondentes, e os demais pontos são obtidos por interpolação com base no valor dos três vértices do triângulo onde o ponto está contido.

A aproximação constante de regiões é utilizada para a segmentação de imagens. Neste caso, as regiões triangulares são vistas como platôs, ou conjuntos de pixels com nível constante, onde a altura de cada região triangular é dada pela média das alturas individuais (cor) de cada pixel contido no triângulo.

Para que possamos computar corretamente as malhas em todo o domínio da imagem, devemos saber em quais triângulos cada ponto (pixel) da imagem está contido. Existem diversas formas de calcular a localização de pontos em uma determinada triangulação. Neste trabalho, foi utilizada a técnica proposta em (SUNDARESWARA; SCHRATER, 2003), em que a localização dos pontos é obtida por caminhamento dentro da malha. Seja $T = \{t_1, t_2, \dots, t_n\}$ o conjunto de triângulos da malha, e seja p o ponto a ser localizado. Na solução proposta por Sundareswara e Schrater, é selecionado um triângulo qualquer t_i aleatoriamente, e então caminha-se pela malha na direção do triângulo t_p que contém o ponto p . A cada passo, é escolhido o triângulo vizinho ao atual que está mais próximo ao ponto p , terminando quando o triângulo atual contém o ponto p . Para isso, os vértices do triângulo atual são tratados como uma base afim para a malha, e a localização do ponto p em relação a esta base é calculada. Ao computarmos as coordenadas baricêntricas (afins) dos vértices de triângulos adjacentes, é possível achar rapidamente o triângulo vizinho mais próximo ao ponto p . Dados três vértices A_1, A_2 e A_3 do triângulo atual, o ponto p pode ser representado por uma combinação ponderada destes pontos:

$$p = \sum_{i=1}^3 \alpha_i A_i = A_j + \sum_{i=1, i \neq j}^3 \alpha_i (A_i - A_j)$$



Figura 5.4: Processo de criação de uma Triangulação de Delaunay com Restrições (a) imagem original; (b) mapa de bordas; (c) aproximação por segmentos de retas; (d) Triangulação de Delaunay com Restrições correspondente.

onde o vetor de pesos $\vec{\alpha} = \alpha_{1,2,3}$ representa as coordenadas baricêntricas do ponto p , e $\sum_{i=1}^3 \alpha_i = 1$. Se e somente se todas as coordenadas $0 < \alpha_i < 1$ então o ponto está contido no triângulo. Caso nem todas as coordenadas α_i sejam positivas, selecionamos os dois vértices cujas coordenadas baricêntricas são as maiores, e movemos para o triângulo adjacente que compartilha estes dois vértices.

Coordenadas baricêntricas podem ser computadas como uma divisão de matrizes, para resolver a equação:

$$\begin{bmatrix} A_1 & A_2 & A_3 \\ 1 & 1 & 1 \end{bmatrix} \vec{\alpha} = \begin{bmatrix} p \\ 1 \end{bmatrix}$$

que tem a forma:

$$Q\vec{\alpha} = P$$

e a solução:

$$\vec{\alpha} = Q^{-1}P$$

Deve-se salientar, ainda, que as coordenadas baricêntricas podem ser utilizadas como pesos para o cálculo da interploação linear, aproveitando a computação feita para a localização dos vértices. Então, seja I a imagem original contendo a informação de cor dos pontos. Uma imagem aproximada \hat{I} é criada a partir dos valores dos vértices da malha interpolando-se nos pontos internos aos triângulos:

$$\hat{I}(p) = \alpha_1 I(A_1) + \alpha_2 I(A_2) + \alpha_3 I(A_3)$$

onde A_i são os vértices do triângulo em que o ponto p está contido, e $\vec{\alpha}$ é o vetor de coordenadas baricêntricas do ponto p em relação aos vértices A_i .

5.1.3 Seleção dos Vértices

Embora já tenha sido descrito como se dá a criação da malha e aproximação da imagem original, dado um conjunto de vértices, ainda não foi especificado como é feita a seleção dos pontos da imagem que formam o conjunto de vértices da malha. A seleção de vértices pode ser vista como um problema de amostragem irregular, em que selecionamos pontos nas áreas de maior interesse da imagem. No presente trabalho, foram utilizadas duas estratégias para a seleção dos vértices da triangulação. Uma está relacionada ao método de aproximação de regiões por interpolação linear, enquanto outra é utilizada quando as regiões triangulares são aproximadas por platôs.

Aproximação por Interpolação Linear

Quando é utilizado o método de interpolação linear para a aproximação das regiões triangulares, a seleção dos vértices é feita de forma iterativa, sendo escolhido um vértice a cada iteração. Em uma dada iteração k , temos uma aproximação da imagem I , dada por \hat{I}_k , um conjunto de vértices da triangulação atual S_k e o conjunto de triângulos atual T_k . Um novo ponto de amostragem p_{n_k} é selecionado na imagem original I na região onde a aproximação \hat{I}_k é pior, gerando um novo conjunto de vértices $S_{k+1} = S_k \cup \{p_{n_k}\}$. Baseado no novo conjunto S_{k+1} , a triangulação de Delaunay D_k é atualizada, gerando uma nova triangulação D_{k+1} . Então, os novos triângulos criados são interpolados de forma a gerar uma nova aproximação \hat{I}_{k+1} .

Duas estratégias são utilizadas aqui na seleção do ponto de amostragem p_{n_k} . Uma das estratégias visa minimizar o máximo erro de aproximação enquanto a outra é utilizada para minimizar a soma dos erros absolutos de aproximação através da imagem. Ambas estratégias são utilizadas em alternância a cada iteração. Seja $R_k = I - \hat{I}_k$ o erro de aproximação obtido no passo k . A primeira estratégia escolhe o ponto p_{n_k} onde o máximo erro de aproximação ocorre:

$$p_{n_k} = \max_{p \in I} \{|R_k(p)|\}$$

enquanto a segunda estratégia seleciona o triângulo com a maior soma de erros absolutos:

$$T_{max_k} = \max_{T_m} \left\{ \sum_{p \in T_m} |R_k(p)| \right\}$$

e escolhe o ponto p_{n_k} contido em T_{max_k} onde:

$$p_{n_k} = \max_{p \in T_{max_k}} \{|R_k(p)|\}$$

Estas duas estratégias são aplicadas alternadamente para obtermos os pontos de amostragem; ou seja, quando a primeira estratégia é utilizada na iteração k , a segunda estratégia é utilizada na iteração $k + 1$. Esta abordagem é utilizada para evitar erros muito grandes de aproximação, e para que se tenha erros de aproximação similares em todas as regiões da imagem. O critério de parada pode ser definido a partir de um erro máximo aceitável, como por exemplo o máximo erro absoluto da imagem:

$$\max_{p \in I} \{|R_k(p)|\} \quad (5.2)$$

a máxima variação do erro na imagem:

$$\max_{p \in I} \{R_k(p)\} - \min_{p \in I} \{R_k(p)\} \quad (5.3)$$

ou ainda o SNR (*Signal to Noise Ratio*) entre a imagem original e a imagem aproximada (JAIN, 1989):

$$SNR = 10 \log_{10} \frac{\sigma^2}{\sigma_e^2} \quad (5.4)$$

onde σ^2 é a variância da imagem original I e σ_e^2 é o erro médio quadrático da imagem aproximada \hat{I} em relação à imagem I .

O algoritmo de seleção dos vértices para construção de uma malha através de interpolação linear é sumarizado abaixo:

1. Fazer $k = 1$ e inicializar o conjunto de vértices S_k com os quatro cantos da imagem I , e o conjunto D_k com dois triângulos, conforme mostrado na Figura 5.5;
2. Atualizar a imagem aproximada \hat{I}_k com o(s) novo(s) vértice(s) e interpolar linearmente os pontos contidos nos novos triângulos;
3. Calcular a imagem erro de aproximação $R_k = I - \hat{I}_k$;

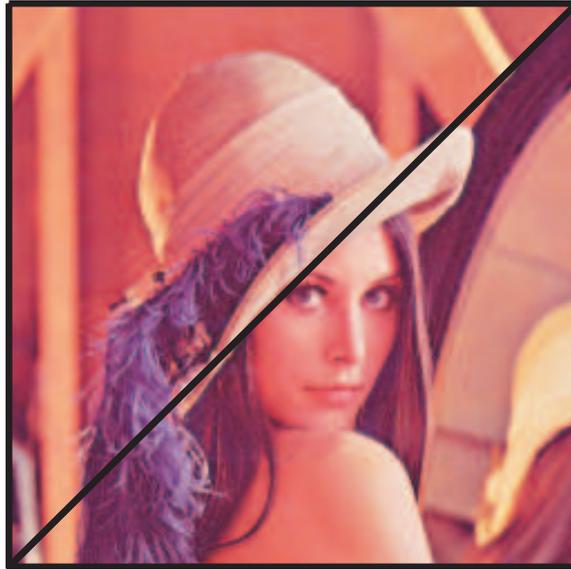


Figura 5.5: Inicialização do processo de inserção de vértices.

4. Calcular o erro de aproximação para cada novo triângulo T_{m_k} :

$$E_{m_k} = \sum_{p \in T_{m_k}} |R_k(p)|$$

5. Se k é ímpar, selecionar o triângulo $T_{max_k} = \max_{T_k} \{E_{m_k}\}$ onde a aproximação é a pior, e selecionar o ponto p_{max_k} contido em T_{max_k} onde o máximo erro absoluto ocorre:

$$p_{max_k} = \max_{p \in T_{max_k}} |R_k(p)|$$

6. Se k é par, selecionar o ponto da imagem p_{max_k} onde o máximo erro absoluto de aproximação ocorre:

$$p_{max_k} = \max_{p \in I} |R_k(p)|$$

7. Se uma ou mais medidas de erro apresentadas nas Equações 5.2, 5.3 e 5.4 estão abaixo do desejado, o processo pára. Caso contrário, fazer $k = k + 1$, atualizar o conjunto S_k com p_{max_k} e ir para o passo 2.

Quando a Triangulação de Delaunay com Restrições é utilizada para manter as bordas aproximadas por segmentos de retas na triangulação resultante, ao invés de inicializarmos o conjunto de vértices S com os quatro cantos da imagem, conforme indicado no passo 1 do algoritmo acima, utilizamos os pontos extremos dos segmentos de retas aproximados, e inicializamos o conjunto inicial de triângulos com o resultado da Triangulação de Delaunay com Restrições equivalente.

Aproximação Constante

A seleção dos vértices da triangulação visando uma aproximação por regiões triangulares constantes é efetuada por um método guloso de seleção de pontos, onde vários

vértices são incluídos na triangulação a cada iteração. A seleção dos vértices é orientada à homogeneidade de feições, no sentido de que triângulos heterogêneos são subdivididos até que o critério de homogeneidade seja satisfeito para todos os triângulos em cada uma das feições utilizadas.

1. Fazer $k = 1$ e inicializar o conjunto de vértices S_k com os quatro cantos da imagem I , e o conjunto D_k com dois triângulos, conforme mostrado na Figura 5.5;
2. Atualizar a imagem aproximada \hat{I}_k nos novos triângulos. Sejam $\{p_j(i) : i = 1, \dots, n\}$ os pontos contidos em um novo triângulo T_{k_j} . Então:

$$\hat{I}_k(p_j) = \frac{1}{n} \sum_{i=1}^n I(p_j(i))$$

3. Achar os novos triângulos T_{k_j} que não satisfazem o critério de homogeneidade:

$$\sqrt{\frac{1}{n} \sum_{p_k(i) \in T_{k_j}} [I(p_k(i)) - \hat{I}(p_k(i))]^2} < thr_{std}$$

onde n é o número de pontos $p_k(i)$ contidos em T_{k_j}

4. Se todos os triângulos satisfazem o critério de homogeneidade, o processo pára. Caso contrário, para cada triângulo T_{k_j} que não satisfaz o critério de homogeneidade, um novo vértice é criado no ponto mais próximo ao seu centróide. O conjunto de vértices S_{k+1} é atualizado, bem como a triangulação D_{k+1} . Fazer $k = k + 1$ e voltar ao passo 2.

Quando a Triangulação de Delaunay com Restrições é utilizada para manter as bordas aproximadas por segmentos de retas na triangulação resultante, ao invés de inicializarmos o conjunto de vértices S com os quatro cantos da imagem, conforme indicado no passo 1 do algoritmo acima, utilizamos os pontos extremos dos segmentos de retas aproximados, e inicializamos o conjunto inicial de triângulos com o resultado da Triangulação de Delaunay com Restrições equivalente.

5.2 Um Método para a Segmentação de Imagens usando Representação Geométrica

Nesta seção será descrito o funcionamento do método de segmentação proposto. A abordagem utilizada para a segmentação de imagens utilizando malhas funciona como um algoritmo de divisão e fusão de regiões (*split and merge*) (GONZALEZ; WOODS, 1992), em que na primeira fase (etapa de divisão), a imagem é dividida em regiões triangulares homogêneas, e na segunda fase (etapa de fusão), é gerado um agrupamento de triângulos semelhantes em classes, através de um método de *clustering* por similaridade no espaço de feições.

Para os testes realizados neste trabalho, foram utilizadas como feições dos pontos da imagem, suas respectivas coordenadas no espaço de cores $L^*a^*b^*$. No entanto, o procedimento apresentado pode ser generalizado para o uso de diferentes feições para os pixels,

como posição espacial, textura, movimento, etc. O espaço de cores $L^*a^*b^*$ foi desenvolvido para representar diferenças perceptuais de maneira uniforme, e os pontos neste espaço podem ser obtidos a partir das componentes R , G e B pelas seguintes transformações (JAIN, 1989):

$$\begin{aligned} L^* &= 25 \left(\frac{100Y}{Y_0} \right)^{1/3} - 16 \\ a^* &= 500 \left[\left(\frac{X}{X_0} \right)^{1/3} - \left(\frac{Y}{Y_0} \right)^{1/3} \right] \\ b^* &= 200 \left[\left(\frac{Y}{Y_0} \right)^{1/3} - \left(\frac{Z}{Z_0} \right)^{1/3} \right] \end{aligned}$$

onde:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

e X_0 , Y_0 e Z_0 são os valores de X , Y e Z para o branco de referência.

A seguir são apresentadas as técnicas de subdivisão triangular (*split*) e de *clustering* de triângulos (*merge*) utilizadas para gerar a segmentação de imagens.

5.2.1 Subdivisão Triangular

No processo de subdivisão triangular visando a segmentação, é utilizado o método de detecção e aproximação de bordas descrito na sub-seção 5.1.1, página 35. As bordas da imagem I de entrada são aproximadas por um conjunto de segmentos de retas $S = \{s_1, s_2, \dots, s_a\}$ das quais os pontos extremos formam o conjunto de vértices $N = \{n_1, n_2, \dots, n_b\}$. A partir do conjunto de vértices N , e dadas as restrições impostas pelos segmentos de S , é computada a Triangulação de Delaunay com Restrições correspondente, gerando o conjunto de triângulos $T = \{t_1, t_2, \dots, t_c\}$. Uma aproximação \hat{I} da imagem original é calculada, gerando platôs (regiões com valores de feições constantes) nas regiões triangulares.

Em alguns casos, apenas a triangulação gerada a partir das bordas aproximadas é suficiente. Mas, em muitos casos, triângulos excessivamente heterogêneos ou com bordas suaves não detectadas em seu interior podem ser passados para a etapa de clustering, gerando artefatos na segmentação resultante. Para evitar isso, são selecionados novos vértices na imagem (ver sub-seção 5.1.3, página 40), refinando a triangulação até que o critério de homogeneidade seja satisfeito para todos os triângulos. Um triângulo é dito homogêneo quando o desvio padrão das feições dos pixels contidos no triângulo é menor que um limiar thr_{std} .

5.2.2 Clustering

Após a imagem I ser devidamente dividida em um conjunto $T = \{t_1, t_2, \dots, t_n\}$ de triângulos homogêneos, um método de *clustering* é utilizado para encontrar classes de triângulos similares entre si. Para cada triângulo t_i , é extraído um vetor representativo das feições dos pixels, que chamaremos de x_i . Em nossos testes, calculamos o vetor x_i pela média aritmética das feições (coordenadas no espaço $L^*a^*b^*$) dos pontos contidos

no triângulo. Deve-se notar que a triangulação também pode ser considerada como um método de amostragem, em que seleciona para o *clustering* apenas um conjunto reduzido de pontos representativos da imagem.

Para o presente trabalho, foi utilizado o método de *clustering* proposto por Yang e Wu (YANG; WU, 2004), com pequenas modificações. Este método recupera *clusters* com base na densidade de pontos no espaço de feições. Os autores propuseram uma expressão para o cálculo da densidade do espaço de feições em um ponto x_i , dada pela similaridade do ponto em relação a todos os outros pontos:

$$J(x_i) = \sum_{j=1}^n \left[\exp \left(-\frac{\|x_j - x_i\|^2}{\beta} \right) \right]^\gamma \quad (5.5)$$

onde γ é o fator de escala da função $J()$, e β é a variância amostral, definida por:

$$\beta = \frac{\sum_{j=1}^n \|x_j - \bar{x}\|^2}{n} \quad \text{onde } \bar{x} = \frac{\sum_{j=1}^n x_j}{n} \quad (5.6)$$

Os clusters são definidos como os picos da função $J()$. O parâmetro γ define a forma da função $J()$ e, conseqüentemente, seus picos. Para a seleção do parâmetro γ , os autores propuseram um método baseado na estabilidade da função $J()$ no espaço-escala. Assim, γ é inicializado com um valor pequeno e a cada passo é incrementado. É medida a correlação entre o valor da função $J()$ para duas iterações consecutivas (dois valores de γ consecutivos) e quando o coeficiente de correlação entre duas iterações consecutivas for maior que um limiar thr_{corr} , o valor de γ atual é escolhido. O algoritmo para seleção do parâmetro γ é sumarizado a seguir (assumir $\gamma_m = 5m$):

1. Fazer $m = 1$ e $thr_{corr} = 0.97$
2. Calcular a correlação dos valores de $J(x_k)_{\gamma_m}$ e $J(x_k)_{\gamma_{m+1}}$
3. Se a correlação for maior ou igual a thr_{corr} , escolher γ_m como estimativa de γ . Caso contrário, fazer $m = m + 1$ e vá para o passo 2.

Definido o valor de γ , devemos procurar pelos valores $Z = \{z_1, z_2, \dots, z_c\}$ que maximizam a similaridade total (picos):

$$J_s(Z) = \sum_{i=1}^c \sum_{j=1}^n \left(\exp -\frac{\|x_j - z_i\|^2}{\beta} \right)^\gamma \quad (5.7)$$

Para encontrar os picos, primeiramente fazemos com que o conjunto Z seja inicializado com o conjunto de pontos $X = x_1, x_2, \dots, x_n$. A seguir, os pontos do conjunto Z são iterativamente aproximados aos picos da função $J()$:

$$z_i = \frac{\sum_{j=1}^n S_{ij}^\gamma x_j}{\sum_{j=1}^n S_{ij}^\gamma} \quad (5.8)$$

onde S_{ij} é a relação de similaridade:

$$S_{ij} = \exp \left(-\frac{\|x_j - z_i\|^2}{\beta} \right) \quad (5.9)$$

Cada ponto z_i é então aproximado, a cada iteração, ao máximo local predominante na vizinhança de x_i . Após a estabilidade, podemos ter vários pontos de Z convergindo para o mesmo pico da função $J()$. Para determinarmos o número de picos para os quais os pontos z_i convergiram, geramos uma árvore de classificação hierárquica (HARTIGAN, 1975) com as posições finais de z_i , utilizando o critério de Ward (soma de quadrados incremental). A árvore então é percorrida da raiz em direção às folhas, e para cada nível k , são calculadas as distâncias intra-classes e inter-classes correspondentes:

$$d_{intra}(k) = \sum_{i=1}^k \sum_{j=1}^{n_k^i} \|z_j^i - \bar{z}^i\|^2 \quad (5.10)$$

$$d_{inter}(k) = \sum_{i=1}^{k-1} \sum_{j=1+1}^k \left(\sum_{a=1}^{n_k^i} \|z_a^i - \bar{z}^{ij}\|^2 + \sum_{b=1}^{n_k^j} \|z_b^j - \bar{z}^{ij}\|^2 \right) \quad (5.11)$$

onde n_k^i é o número de pontos da i -ésima classe do nível k , z_j^i é o j -ésimo ponto da i -ésima classe, \bar{z}^i é a média dos pontos da i -ésima classe, \bar{z}^{ij} é a média da união dos pontos da i -ésima e j -ésima classes.

A decisão de qual o nível k da árvore será podado, ou seja, quantas classes serão selecionadas, é feita com base no critério de maximização da distância inter-classes e minimização da distância intra-classes:

$$n_{clusters} = \max_k \left\{ \frac{d_{inter}(k)}{d_{intra}(k)}, k = 2, \dots, n \right\} \quad (5.12)$$

Após ser selecionado o número de classes $n_{clusters}$, a partir da árvore de classificação hierárquica obtemos os pontos z_i^j pertencentes a cada classe j . Os pontos x_i correspondentes, por sua vez, recebem os mesmos rótulos de classe, bem como os triângulos t_i que originaram estes pontos. Ao final, temos um conjunto de triângulos divididos em classes. Ao unirmos triângulos adjacentes da mesma classe, geramos uma representação das regiões da imagem através de regiões poligonais. Esta representação é especialmente útil em aplicações de imagens e vídeos orientadas a objetos, tais como detecção, rastreamento e descrição de objetos, codificação baseada em objetos, recuperação de informações, entre outras. O método de classificação utilizado é robusto contra ruído, e gera uma segmentação compacta, com um número reduzido de cores e regiões conexas, fazendo com que a saída da segmentação seja facilmente tratável por aplicações de alto nível.

Resultados do método de segmentação proposto, bem como comparações com outros métodos da literatura são apresentados e discutidos no Capítulo 6.

5.3 Um Método para a Compressão de Imagens Usando Representação Geométrica

O método de compressão de imagens proposto baseia-se na modelagem da imagem de entrada em malhas triangulares, tal como a aproximação de uma superfície. Esta aproximação é utilizada como um preditor, de forma semelhante aos métodos apresentados em (SHEN; RANGAYYAN, 1997) e (RATAKONDA; AHUJA, 2002) e discutidos no capítulo

4. A diferença é que estes dois últimos utilizam como preditor modelos gerados por técnicas de segmentação de imagens, enquanto no método proposto é utilizada a modelagem de imagens baseada em malhas. A modelagem através de malhas como forma de predição oferece vantagens sobre os métodos baseados em segmentação visando compressão. Isto porque os métodos baseados em segmentação, embora possam gerar uma boa aproximação da imagem original e, conseqüentemente, um resíduo pequeno a ser codificado (quando a codificação sem perdas é considerada), eles tendem a gerar um *overhead* considerável na codificação das fronteiras entre as regiões segmentadas. Na predição baseada em malhas triangulares, apenas os vértices da triangulação devem ser codificados, sendo que a imagem aproximada pode ser reconstruída pela interpolação dos pontos restantes e gerando, ainda assim, um pequeno resíduo.

Quando o objetivo for a compressão com perdas, apenas os dados necessários para a reconstrução da malha são codificados. Para obtermos uma codificação sem perdas, o erro de aproximação deve ser calculado e codificado junto à malha. No presente trabalho, uma atenção especial foi dada para a compressão de imagens digitalizadas de mamografias. A descrição do método, voltado para esta aplicação, pode ser vista em (SILVA; SCHARCANSKI, 2005).

A seguir serão apresentadas em detalhes as abordagens utilizadas na subdivisão triangular e o esquema de codificação.

5.3.1 Subdivisão Triangular

Dada uma imagem de entrada I , devemos gerar uma imagem aproximada \hat{I} , modelada a partir de uma malha triangular, que funcionará como um preditor. Para uma codificação reversível (sem perdas), o erro residual de aproximação dado por

$$R = \hat{I} - I$$

deve ser calculado, codificado, e armazenado junto aos parâmetros de modelagem da malha.

A modelagem da malha se dá conforme descrito na seção 5.1. Dependendo do tipo de imagem que desejamos comprimir, a técnica de aproximação de bordas duplas por segmentos pode ser utilizada ou não. No caso de imagens que apresentam muitas bordas, em especial bordas do tipo rampa (*ramp edges*), é conveniente utilizarmos a detecção de bordas duplas. Isto ocasiona um *overhead* na codificação da malha, pela necessidade de se armazenar as informações das arestas de bordas; porém, isto reduz consideravelmente o erro de aproximação, melhorando a eficiência da codificação como um todo.

Algumas classes de imagens não apresentam bordas do tipo degrau ou rampa. Para imagens com estas características, a aproximação de bordas e, conseqüentemente, a codificação das arestas, se torna um *overhead* desnecessário. Em nossos teste, o método foi aplicado a imagens digitalizadas de mamografias. As mamografias são imagens de alta resolução, com alta correlação espacial dos pixels, não possuem bordas abruptas e, por isso, o método de compressão proposto apresenta bons resultados apenas representando a malha pela Triangulação de Delaunay simples. Desta forma, o conjunto de vértices é inicializado com os quatro cantos da imagem original, e novos vértices são selecionado

iterativamente nos pontos onde a aproximação por interpolação linear é pior, conforme mostrado na subseção 5.1.3.

5.3.2 Esquema de Codificação

Após finalizado o processo de inserção de pontos na Triangulação de Delaunay ou Triangulação de Delaunay com Restrições, já podemos codificar os dados. Para que o decodificador possa reconstruir a malha adequadamente, é necessário codificar os vértices da triangulação (posição e informação de cor), as arestas correspondentes às bordas (pares de vértices da triangulação) e, se for o caso de compressão sem perdas, o erro de aproximação.

Tendo as posições dos vértices, e as arestas correspondentes às bordas da imagem, o decodificador pode reconstruir a Triangulação de Delaunay ou Triangulação de Delaunay com Restrições, sem qualquer perda com relação à adjacência original dos vértices. A informação de cor dos vértices é utilizada então para gerar uma aproximação da imagem original a partir da interpolação linear dentro dos elementos triangulares da malha.

Seja B uma imagem binária, do mesmo tamanho da imagem original I , onde o valor 1 é atribuído aos pontos onde os vértices da triangulação estão localizados, e o valor 0 aos pontos restantes. A imagem B é percorrida de cima para baixo, esquerda para a direita, sendo considerada como um vetor uni-dimensional. Então, as corridas de zeros são codificadas usando um código de corridas (*Run-Length Coding* = RLC) (GOLOMB, 1966). Tanto as corridas de zeros, que representam as posições dos vértices, como as cores dos vértices, são então codificados pelo código de Huffman (HUFFMAN, 1952). Caso seja usada a Triangulação de Delaunay com Restrições - para fixar as bordas da imagem na triangulação - as arestas, representadas por pares de índices apontando para os vértices correspondentes, também são codificados pelo código de Huffman.

Aqui, já temos informação suficiente codificada para reconstruir a imagem original com perdas. Para que uma reconstrução sem perdas seja possível, devemos codificar o resíduo da aproximação junto à imagem codificada. O método JPEG-LS (WEINBERGER; SEROUSSI; SAPIRO, 1996) se mostrou eficaz na codificação do resíduo e, portanto, foi utilizado neste trabalho.

Nos testes realizados utilizando-se mamografias digitalizadas de 16 bits, verificou-se que os bits menos significativos destas imagens apresentavam basicamente ruído branco, com muito pouca correlação espacial. Isto faz com que a aproximação da malha à imagem não apresente um bom desempenho, dificultando a convergência da aproximação. Para aumentar a eficiência da aproximação por malhas, melhorando a convergência, a triangulação é construída utilizando-se os bits mais significativos, tal como descrito anteriormente, enquanto os bits menos significativos são codificados separadamente, utilizando-se o método PNG (RANDERS-PEHRSON, 1999), que se mostrou eficiente para esta aplicação.

6 RESULTADOS EXPERIMENTAIS

Neste capítulo são apresentados resultados experimentais dos métodos propostos no capítulo 5. Será estudada a variação de parâmetros e metodologias e sua relação com os resultados obtidos, bem como uma comparação com outros métodos existentes. A seção 6.1 apresenta resultados experimentais do método de segmentação de imagens proposto, enquanto as seções 6.2 e 6.3 apresentam resultados dos métodos propostos para a compressão sem perdas e com perdas, respectivamente.

6.1 Segmentação de Imagens

Conforme dito na subseção 5.2.1, o método de segmentação proposto parte da detecção de bordas pelo método de Canny, as bordas são aproximadas por segmentos de reta, e com base nestes segmentos, é gerada uma Triangulação de Delaunay com Restrições. Foi mostrado ainda um método de refinamento da triangulação, opcional, para evitar que triângulos heterogêneos sejam utilizados na etapa de *clustering*. A seguir, mostraremos a influência da detecção de bordas e do refinamento no resultado final da segmentação. Na Figura 6.1, a primeira linha mostra as três imagens originais utilizadas neste teste. A segunda linha mostra os resultados da detecção de bordas. A terceira linha mostra as triangulações geradas apenas pela aproximação das bordas por arestas. A quarta linha mostra os resultados da segmentação para as triangulações da terceira linha. A quinta linha mostra as triangulações geradas após o refinamento de regiões heterogêneas. A sexta linha mostra os resultados da segmentação para as triangulações da quinta linha.

Na detecção de borda das duas últimas colunas da Figura 6.1 os limiares foram alterados intencionalmente para gerarem maus resultados. Como pode ser visto, o refinamento da triangulação é necessário apenas quando o resultado da detecção de bordas apresenta falhas na detecção. Caso a detecção de bordas seja robusta o suficiente para a aplicação, a etapa de refinamento pode ser ignorada em prol de maior eficiência. Caso contrário, o refinamento pode ser útil para dar mais robustez ao método.

Para demonstrar a capacidade de extrair regiões com alguma semântica das imagens, mantendo um resultado compacto, a Figura 6.2 mostra resultados comparativos entre o método de segmentação proposto e o método baseado em *mean-shift* apresentado em (COMANICIU; MEER, 2002), e considerado estado-da-arte em segmentação de imagens.

A seguir são apresentados resultados do método de segmentação de imagens visando a quantização de paletas de cores. Para isso, os resultados obtidos são comparados com



Figura 6.1: Teste de segmentação com e sem refinamento de regiões. As linhas do topo para baixo mostram: Imagens Originais, mapas de bordas, triangulações sem refinamento, aproximações sem refinamento, triangulações com refinamento e aproximações com refinamento.

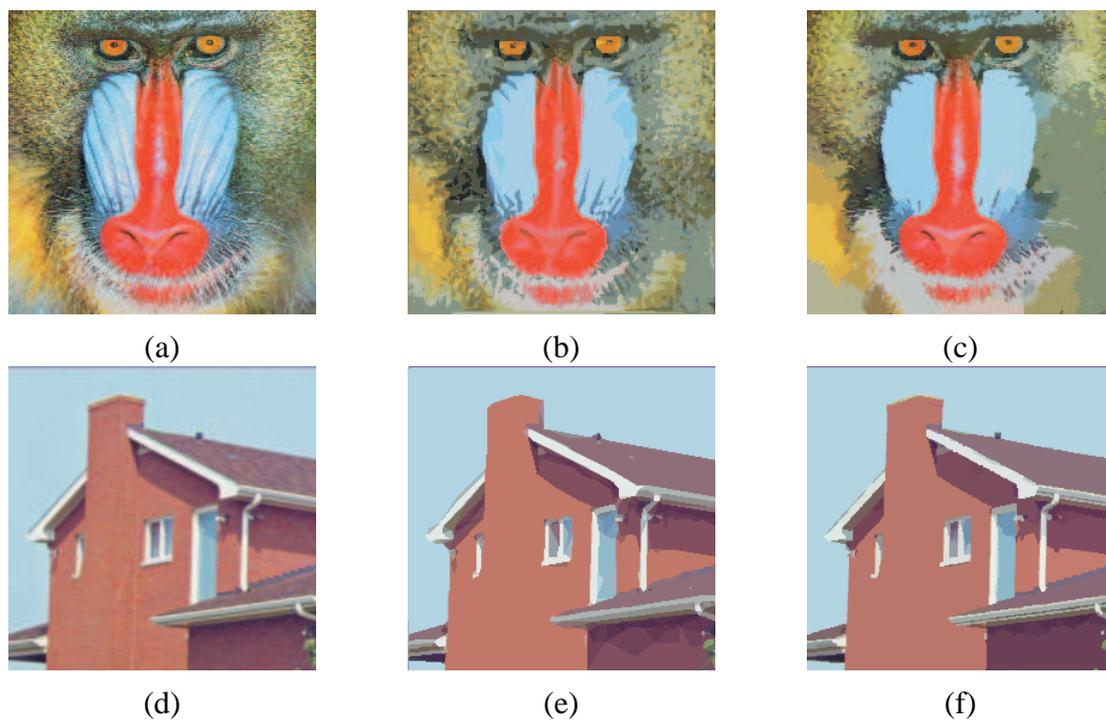


Figura 6.2: Comparação do método proposto com o *mean-shift*. (a,d) imagens originais; (b) imagem segmentada pelo método proposto com 26 cores e 2070 regiões conexas; (c) imagem segmentada pelo método *mean-shift* com 2399 cores e 2407 regiões conexas; (e) imagem segmentada pelo método proposto com 53 cores e 347 regiões conexas; (f) imagem segmentada pelo método *mean-shift* com 318 cores e 319 regiões conexas.

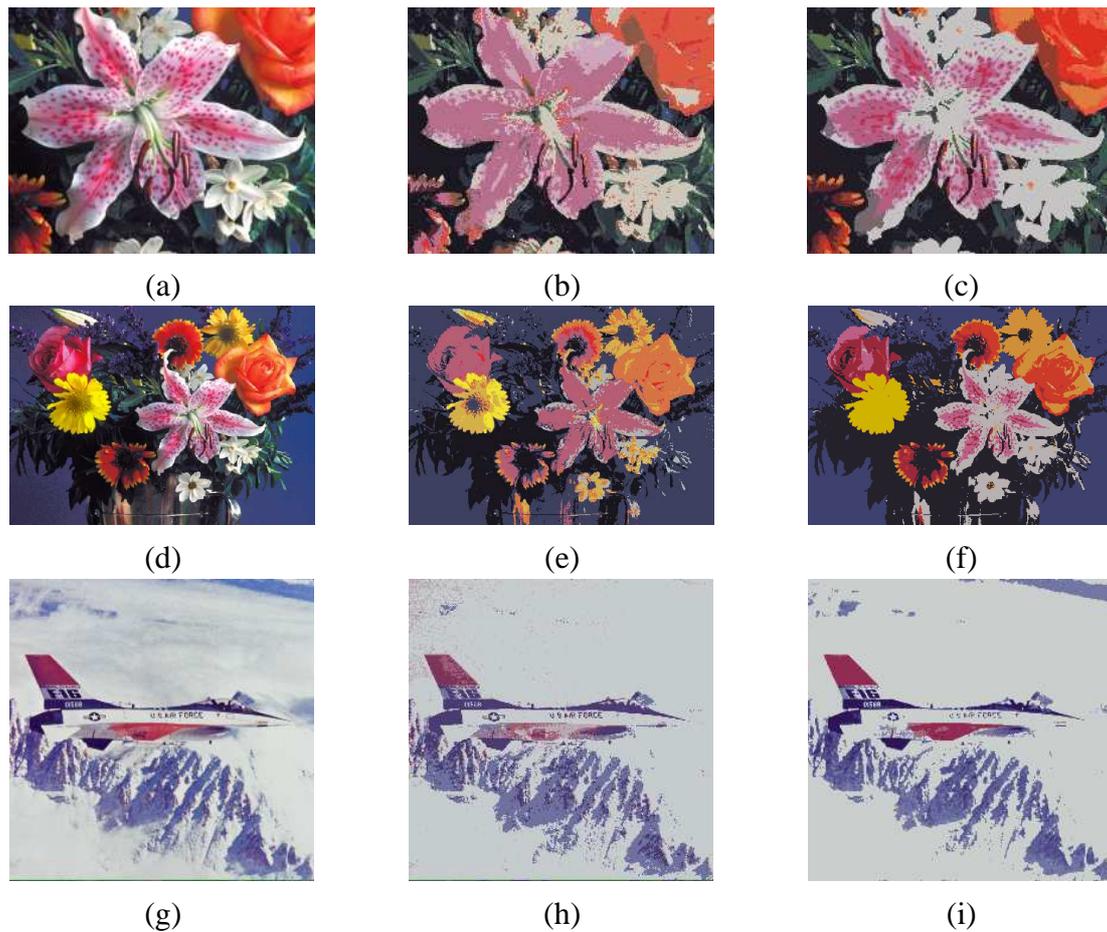


Figura 6.3: Método de segmentação proposto aplicado à quantização de paletas de cores. (a,d,g) imagens originais; (b) resultado do método proposto por Cheng e Sun, com 14 cores e PSNR = 20,98dB; (c) resultado obtido a partir do método proposto, com 14 cores e PSNR = 23,97dB; (e) resultado do método proposto por Cheng e Sun, com 8 cores e PSNR = 20,72dB; (f) resultado obtido a partir do método proposto, com 8 cores e PSNR = 22,75dB; (h) resultado do método proposto por Cheng e Sun, com 9 cores e PSNR = 22,99dB; (i) resultado obtido a partir do método proposto, com 9 cores e PSNR = 23,49dB.

o método proposto por Cheng e Sun (CHENG; SUN, 2000), cuja principal característica é gerar resultados extremamente compactos em relação ao número de cores resultantes. Para efeito de comparação, ao invés de utilizarmos nosso método adaptativo de seleção do número de *clusters*, fixamos o número de *clusters* como sendo igual ao obtido pelo método proposto por Cheng e Sun. Ainda, a imagem é classificada pixel a pixel, ao invés de triângulo a triângulo, como fora considerado nos exemplos anteriores. A intenção aqui é apenas verificar a capacidade do algoritmo de segmentação desenvolvido de manter o quanto possível a informação contida na imagem com um pequeno número de cores. A Figura 6.3 apresenta os resultados obtidos para algumas imagens.

Tabela 6.1: Taxas de compressão em bits/pixel para mamografias de 16 bits.

	Método Proposto	JPEG 2000	JPEG-LS	JPEG-lossless	PNG
mamografia1	7,755	10,994	10,690	11,292	9,328
mamografia2	7,164	10,554	10,207	10,519	8,869
mamografia3	7,315	10,726	10,377	10,793	8,949
mamografia4	8,413	11,388	11,192	11,574	9,788
mamografia5	7,705	10,968	10,672	11,207	9,284
mamografia6	7,442	10,843	10,521	10,985	9,117
mamografia7	8,326	11,368	11,111	11,590	9,744
mamografia8	6,838	10,433	10,001	10,388	8,590
mamografia9	7,667	10,984	10,705	11,145	9,310
mamografia10	7,528	10,831	10,528	11,008	9,174
média	7,615	10,910	10,600	11,050	9,215

6.2 Compressão de Imagens sem Perdas

Para a compressão sem perdas, duas modalidades de criação da malha são utilizadas: a Triangulação de Delaunay e a Triangulação de Delaunay com Restrições. A última utiliza o método de aproximação de bordas duplas e é apropriada para classes de imagens que apresentam bordas do tipo degrau ou do tipo rampa. Para outros tipos de imagens, a detecção e codificação de arestas de bordas é um custo desnecessário para o processo.

Durante o desenvolvimento deste trabalho, o método de compressão desenvolvido foi utilizado na compressão sem perdas de mamografias. Os resultados para este tipo de imagem foram bastante satisfatórios, sem a utilização de detecção de bordas. As mamografias utilizadas neste trabalho são imagens digitalizadas de 8 bits. Algumas das imagens de mamografias testadas são mostradas na Figura 6.4. Os resultados da compressão sem perdas para mamografias a partir do método proposto, bem como os resultados obtidos a partir dos métodos JPEG lossless (PENNEBAKER; MITCHELL, 1993), JPEG-LS (WEINBERGER; SEROUSSI; SAPIRO, 1996), JPEG 2000 (no modo sem perdas) (ADAMS, 2001) e PNG (RANDERS-PEHRSON, 1999), são mostrados na Tabela 6.1.

6.3 Compressão de Imagens com Perdas

A seguir são apresentados resultados da compressão com perdas segundo o método proposto. Para os testes apresentados aqui, foi utilizada a abordagem de detecção de bordas duplas e aproximação da malha através de uma Triangulação de Delaunay com Restrições. Diferentemente da compressão sem perdas, o erro de aproximação não é codificado, sendo apenas armazenados os dados necessários para a reconstrução da malha.

A Figura 6.5 apresenta as imagens testadas. Na primeira coluna aparecem as imagens originais, na segunda coluna são mostradas as imagens reconstruídas após a compressão pelo método proposto enquanto na terceira e última coluna são apresentados resultados

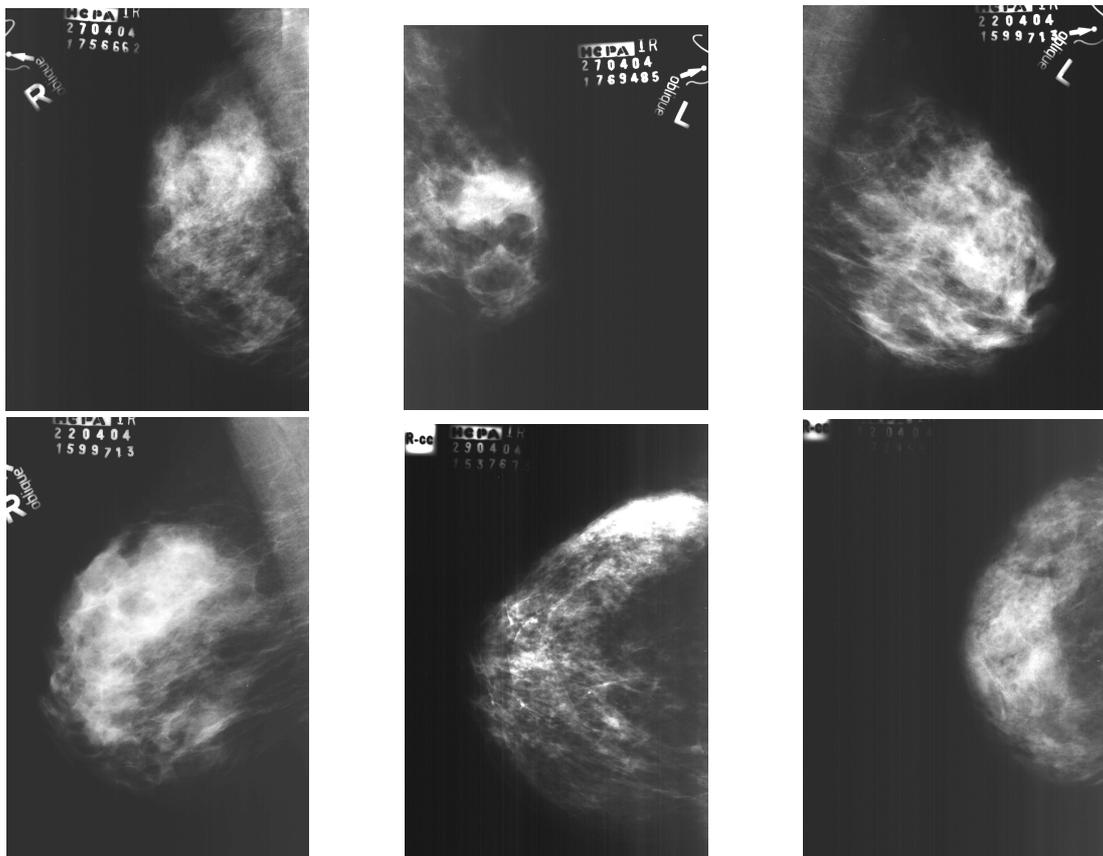


Figura 6.4: Mamografias digitalizadas de 16 bits.

Tabela 6.2: Resultados da compressão com perdas

	<i>girl</i>	<i>flowers</i>	<i>splash</i>	<i>peppers</i>
Taxa de compressão JPEG em bits/pixel	0.819	4.213	3.173	4.256
Taxa de compressão método proposto em bits/pixel	2.130	7.145	4.869	7.673
Custo relativo da codificação das posições dos vértices (bits/pixel)	20% (0.426)	14% (1.005)	20% (0.959)	16% (1.214)
Custo relativo da codificação das arestas de bordas (bits/pixel)	4% (0.085)	9% (0.656)	8% (0.391)	5% (0.423)
Custo relativo da codificação das cores dos vértices (bits/pixel)	76% (1.619)	77% (5.484)	72% (3.518)	79% (6.025)

da compressão através do método JPEG. Algumas medidas quantitativas referentes aos resultados mostrados na Figura 6.5 são apresentados na Tabela 6.2. Os parâmetros de controle das taxas de compressão nos dois métodos foram selecionados de maneira que os dois apresentassem resultados com valores de PSNR aproximadamente iguais a 33dB em relação à imagem original.



Figura 6.5: Compressão com perdas. (a) imagens originais; (b) imagens comprimidas pelo método proposto e (c) imagens comprimidas pelo método JPEG.

7 CONCLUSÕES

Neste capítulo serão apresentadas as considerações finais do trabalho. Aspectos positivos e negativos da abordagem utilizada serão discutidos, bem como algumas possíveis melhorias nas técnicas apresentadas e as perspectivas de trabalhos futuros.

Este trabalho apresenta uma forma de representação de imagens através de malhas triangulares. O potencial da representação é demonstrado através da sua aplicação em duas áreas de suma importância no processamento de imagens: segmentação e compressão. Os resultados do método de segmentação de imagens proposto mostram que o método possui uma grande capacidade de extrair informação semântica de uma imagem, de uma forma bastante compacta, demonstrando superioridade em relação a outros métodos da literatura. No melhor do nosso conhecimento, este é o primeiro método de segmentação de imagens baseado na modelagem de malhas que utiliza um método de *clustering* de triângulos para gerar um conjunto de regiões poligonais como resultado.

A aplicação de compressão de imagens, embora não tenha apresentado um desempenho igualmente satisfatório em todas as classes de imagens, se comparada a outros métodos existentes, apresenta características interessantes a serem desenvolvidas, tal como a eficiência da malha em armazenar informação de conteúdo. Além do mais, foi o primeiro trabalho a utilizar a modelagem de malhas visando a compressão sem perdas, que por sua vez se mostrou bastante eficiente quando aplicada à compressão de mamografias. O desenvolvimento de técnicas de codificação adequadas para esta abordagem pode trazer resultados melhores futuramente. A técnica de detecção de bordas duplas introduzida na subseção 5.1.1, visando a aproximação de imagens por meio de malhas, é outra novidade apresentada neste trabalho.

Os resultados apresentados no capítulo 6 para a compressão com perdas mostram uma inferioridade do método proposto quando comparado ao método JPEG. No entanto, deve-se salientar que, de acordo com o método proposto, a informação de cor dos vértices é codificada diretamente pelo método de Huffman, sem perdas, o que é desnecessário. Justamente o *overhead* causado pela codificação da cor dos vértices é o grande gargalo na eficiência do método.

A compressão sem perdas se mostrou bastante eficiente na codificação de mamografias digitalizadas. Em média, o método proposto obteve taxas de compressão em média 43%, 40%, 45% e 25% maiores do que as obtidas com os métodos JPEG 2000, JPEG-LS, JPEG-lossless, e PNG, respectivamente.

Na seção 5.2.2 foi visto que o processo de *clustering* envolve a criação de uma árvore

de classificação hierárquica. Esta árvore pode ser utilizada para criar uma representação multi-escala correspondente de regiões da imagem, com uma hierarquia binária. Uma idéia semelhante a esta foi utilizada em (SALEMBIER; GARRIDO, 2000), onde a árvore era construída a partir de um método de segmentação baseado em crescimento de regiões, e as regiões tinham formas quaisquer e eram representadas por códigos de cadeia. A representação poligonal obtida no método de segmentação proposto, aliada a uma representação hierárquica, pode ser bastante útil em aplicações de recuperação de informações. Além do mais, a utilização de modelos de representação baseados em malhas visando recuperação de imagens ainda é algo pouco explorado.

7.1 Trabalhos Futuros

As vantagens apresentadas pelo método de segmentação desenvolvido lhe credenciam a ser utilizado futuramente em aplicações que envolvam a ligação entre as características de baixo nível e as de alto nível semântico de uma imagem, tal como na codificação de vídeos baseada em objetos ou na recuperação de informações.

A estratégia de orientar a criação da malha com base na detecção de bordas duplas traz melhorias para a aproximação baseada na Triangulação de Delaunay com Restrições, ao dar um tratamento especial para as regiões de bordas. Isto porque nas estratégias de aproximação de imagens por meio de malhas, os maiores erros de aproximação geralmente ocorrem nas regiões próximas às bordas. Ainda assim, mesmo com o método de detecção de bordas duplas, os maiores erros de aproximação continuam se dando nas regiões próximas às bordas. A aproximação em áreas texturadas também é problemática. Isto se deve ao fato de que a interpolação linear não é o método mais adequado para a aproximação nestas ocasiões. A utilização de novos modelos de aproximação, específicos para pixels de bordas ou pixels internos de regiões, são alternativas que podem trazer maior eficiência na codificação, e que serão testadas futuramente.

A utilização de um método mais eficiente de codificação da cor dos vértices pode ser utilizado para contornar o problema da falta eficiência na compressão com perdas. Métodos preditivos baseados na adjacência dos vértices da malha (CHOI et al., 2000), e que introduzam perdas aceitáveis são alternativas para aumentar a eficiência da compressão de imagens com perdas. Ainda, ao se utilizar diferentes métodos de representação de regiões, pode ser feita a conexão entre os resultados da segmentação de imagens proposta e a codificação de imagens e vídeos. A saída da segmentação pode ser utilizada junto a novas formas de modelagem de regiões para fornecer uma codificação bastante compacta, ao explorar a correlação existente entre elementos de malha de mesma classe.

O método de segmentação proposto ainda pode ser aperfeiçoado ao se adicionar informação de textura e movimento na etapa de *clustering*. Como foi exposto no capítulo 2, as malhas são estruturas adequadas para a codificação de objetos em movimento, e por isso a aplicação do método em vídeos será desenvolvido futuramente.

REFERÊNCIAS

ABDAT, M.; BELLANGER, M. G. Combining Gray coding and JBIG for lossless image compression. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 1994, Austin, TX, USA. **Proceedings...** [S.l.: s.n.], 1994. v.3, p.851–855.

ABRAMSON, N. **Information Theory and Coding**. [S.l.]: McGraw-Hill, 1963.

ADAMS, M. **The JPEG-2000 still image compression standard**. [S.l.]: ISO/IEC JTC1/SC29/WG1, 2001. (N2412).

BALIGAR, V. P.; PATNAIK, L. M.; NAGABHUSHANA, G. R. High compression and low order linear predictor for lossless coding of grayscale images. **Image and Vision Computing**, [S.l.], v.21, n.6, p.473–562, June 2003.

BEEK, P. van; TEKALP, A. M.; ZHUANG, N.; CELASUN, I.; XIA, M. Hierarchical 2-D Mesh Representation, Tracking, and Compression for Object-Based Video. **IEEE Trans. on Circuits and Systems for Video Technology**, [S.l.], v.9, n.2, p.353–369, March 1999.

BERG, M. de; KREVELD, M.; OVERMARS, M.; SHWARZKOPF, O. **Computational Geometry: algorithms and applications**. 2nd ed. New York: Springer-Verlag, 1997.

BRANKOV, J. G.; YANG, Y.; WERNICK, M. N. Accurate mesh representation of vector-valued (color) images. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 2003. **Proceedings...** [S.l.: s.n.], 2003. v.1, p.705–708.

CANNY, J. A Computational Approach to Edge Detection. **IEEE Trans. on Pattern Analysis and Machine Intelligence**, [S.l.], v.8, n.6, p.679–698, 1986.

CHENG, H. D.; JIANG, X. H.; SUN, Y.; WANG, J. Color image segmentation: advances and prospects. **Pattern Recognition**, [S.l.], v.34, n.12, p.2259–2281, 2001.

CHENG, H.-D.; SUN, Y. A hierarchical approach to color image segmentation using homogeneity. **IEEE Trans. on Image Processing**, [S.l.], v.9, n.12, p.2071–2082, December 2000.

CHEW, L. P. Constrained Delaunay triangulations. In: ANNUAL SYMPOSIUM ON COMPUTATIONAL GEOMETRY, SCG, 3., 1987. **Proceedings...** New York: ACM Press, 1987. p.215–222.

CHOI, J. S.; KIM, Y. H.; LEE, H.-J.; PARK, I.-S.; LEE, M. H.; AHN, C. Geometry Compression of 3-D Mesh Models Using Predictive Two-Stage Quantization. **IEEE Trans. on Circuits and Systems for Video Technology**, [S.l.], v.10, n.2, p.312–322, March 2000.

CHRISTOUDIAS, C. M.; GEORGESCU, B.; MEER, P. Synergism in low level vision. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 2002, Quebec City, Canada. **Proceedings...** [S.l.: s.n.], 2002.

COMANICIU, D.; MEER, P. Mean Shift: a robust approach toward feature space analysis. **IEEE Trans. Pattern Anal. Mach. Intell.**, [S.l.], v.24, n.5, p.603–619, 2002.

DEMARET, L.; ROBERT, G. Scalable Image Coder Mixing DCT and Triangular Meshes. In: IEEE INT. CONF. IMAGE PROCESSING, 2000. **Proceedings...** [S.l.: s.n.], 2000. v.3, p.849–852.

DENG, G.; YE, H.; MARUSIC, S.; TAY, D. A Method for Predictive Order Adaptation Based on Model Averaging. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 2003. **Proceedings...** [S.l.: s.n.], 2003. v.2, p.189–192.

DIMITROVA, N.; ZHANG, H.-J.; SHAHRARAY, B.; SEZAN, I.; HUANG, T.; ZAKHOR, A. Applications of Video-Content Analysis and Retrieval. **IEEE Multimedia**, [S.l.], v.9, n.3, p.42–55, July-September 2002.

DUDON, M.; AVARO, O.; ROUX, C. Triangular active mesh for motion estimation. **Signal Processing: Image Communication**, [S.l.], v.10, p.21–41, 1997.

FU, K. S.; MUI, J. K. A survey on image segmentation. **Pattern Recognition**, [S.l.], v.13, n.1, p.3–16, 1981.

FUKUNAGA, K.; HOSTETLER, L. D. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. **IEEE Trans. Information Theory**, [S.l.], v.21, p.32–40, 1975.

GEVERS, T. Image Segmentation and Similarity of Color-Texture Objects. **IEEE Trans. Multimedia**, [S.l.], v.4, n.4, p.509–516, December 2002.

GOKÇETEKIN, M.; HARMANCI, M. D.; CELASUN, I.; TEKALP, A. M. Mesh-Based Segmentation and Update for Object-Based Video. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 2000. **Proceedings...** [S.l.: s.n.], 2000.

GOLOMB, S. W. Run-length Encodings. **IEEE Trans. Information Theory**, [S.l.], v.12, n.3, p.399–401, 1966.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. [S.l.]: Addison-Wesley Longman Publishing Co., 1992.

GUO, G. D.; YU, S.; MA, S. D. Unsupervised segmentation of color images. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 1998. ICIP 98, 1998. **Proceedings...** [S.l.: s.n.], 1998. v.3, p.299–302.

- HARALICK, R. M.; SHAPIRO, L. G. Image Segmentation Techniques. **Computer Vision, Graphics and Image Processing**, [S.l.], v.29, p.100–132, 1985.
- HARTIGAN, J. A. **Clustering Algorithms**. [S.l.]: John Wiley & Sons, 1975.
- HERMES, L.; BUHMANN, J. M. A Minimum Entropy Approach to Adaptive Image Polygonization. **IEEE Trans. on Image Processing**, [S.l.], v.12, n.10, p.1243–1258, October 2003.
- HSIEH, J.-W.; GRIMSON, W. E. L. Spatial Template Extraction for Image Retrieval Matching. **IEEE Trans. Image Processing**, [S.l.], v.12, n.11, p.1404–1415, November 2003.
- HSU, P.; LIU, K. J. R.; CHEN, T. A Low Bit-Rate Video Codec Based on Two-Dimensional Mesh Motion Compensation with Adaptive-Interpolation. **IEEE Trans. on Circuits and Systems for Video Technology**, [S.l.], v.11, n.1, p.111–117, January 2001.
- HUFFMAN, D. A. A Method For Construction of Minimum Redundancy Codes. In: IRE, 1952. **Proceedings...** [S.l.: s.n.], 1952. p.1098–1101.
- IP, H. H. S.; LAM, S. W. C. Alternative strategies for irregular pyramid construction. **Image and Vision Computing**, [S.l.], v.14, n.4, p.297–304, May 1996.
- ISO/IEC. **ISO 11544**: progressive bi-level image compression. [S.l.], 1993.
- JAIN, A. K. **Fundamentals of digital image processing**. [S.l.]: Prentice-Hall, Inc., 1989.
- KASHIMURA, M.; SATO, Y.; OZAWA, S. Image description for coding using triangular patch structure. In: ICCS/ISITA, 1992. **Proceedings...** [S.l.: s.n.], 1992.
- KOHOUT, J.; KOLINGEROVÁ, I. Parallel Delaunay triangulation based on circum-circle criterion. In: SPRING CONFERENCE ON COMPUTER GRAPHICS, SCCG, 19., 2003. **Proceedings...** New York: ACM Press, 2003. p.73–81.
- KONYHA, L.; ENYEDI, B.; TRAN, S. M.; FAZEKAS, K. Content-based mesh generation algorithm. In: EURASIP CONFERENCE ON VIDEO, IMAGE PROCESSING AND MULTIMEDIA COMMUNICATIONS, 4., 2003. **Proceedings...** [S.l.: s.n.], 2003. v.1, p.175–180.
- KURUGOLLU, F.; SANKUR, B.; HARMANCI, A. E. Color image segmentation using histogram multithresholding and fusion. **Image and Vision Computing**, [S.l.], v.19, n.13, p.915–928, 2001.
- LUO, J.; GUO, C. Perceptual Grouping of Segmented Regions in Color Images. **Pattern Recognition**, [S.l.], v.36, n.12, p.2781–2792, December 2003.
- MARTÍNEZ, J. M. MPEG-7: overview of mpeg-7 description tools, part 2. **IEEE Multimedia**, [S.l.], v.9, n.3, p.83–93, July-September 2002.
- MARTÍNEZ, J. M.; KOENEN, R.; PEREIRA, F. MPEG-7: the generic multimedia content description standard, part 1. **IEEE Multimedia**, [S.l.], v.9, n.2, p.78–87, April-June 2002.

NACK, F.; LINDSEY, A. T. Everything You Wanted to Know About MPEG-7: part 1. **IEEE Multimedia**, [S.l.], v.6, n.3, p.65–77, July-September 2002.

NACK, F.; LINDSEY, A. T. Everything You Wanted to Know About MPEG-7: part 2. **IEEE Multimedia**, [S.l.], v.6, n.4, p.64–73, October-December 2002.

PAL, N. R.; PAL, S. K. A review on image segmentation techniques. **Pattern Recognition**, [S.l.], v.26, n.9, p.1277–1294, September 1993.

PAVLIDIS, T. **Structural Pattern Recognition**. New York: Springer, 1980.

PENNEBAKER, W. B.; MITCHELL, J. L. **JPEG Still Image Data Compression Standard**. 3rd ed. New York: Van Nostrand Reinhold, 1993.

PINHO, A. J. Region-Based Near-Lossless Image Compression. In: IEEE INT. CONF. ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2001, Salt Lake City, Utah, USA. **Proceedings...** [S.l.: s.n.], 2001. p.1761–1764.

RANDERS-PEHRSON, G. **PNG (Portable Network Graphics) specification version 1.1**. [S.l.]: PNG Development Group, 1999.

RATAKONDA, K.; AHUJA, N. Lossless Image Compression With Multiscale Segmentation. **IEEE Trans. Image Processing**, [S.l.], v.11, n.11, p.1228–1237, November 2002.

RAZDAN, A.; BAE, M. S. A hybrid approach to feature segmentation of triangle meshes. **Computer-Aided Design**, [S.l.], v.35, p.783–789, 2003.

SALEMBIER, P.; GARRIDO, L. Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval. **IEEE Trans. on Image Processing**, [S.l.], v.9, n.4, p.561–576, April 2000.

SALEMBIER, P.; MARTÍ, E.; PARDÀS, M. Active Mesh Coding and Rate-Distortion Theory. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 1996, Lausanne, Switzerland. **Proceedings...** [S.l.: s.n.], 1996. v.2, p.113–116.

SHEN, L.; RANGAYYAN, R. M. A Segmentation-Based Lossless Image Coding Method for High-Resolution Medical Image Compression. **IEEE Trans. Med. Imaging**, [S.l.], v.16, n.3, p.301–307, June 1997.

SHEWCHUK, J. R. Delaunay Refinement Algorithms for Triangular Mesh Generation. **Computational Geometry: Theory and Applications**, [S.l.], v.22, n.1-3, p.21–74, May 2002.

SHEWCHUK, J. R. Updating and constructing constrained delaunay and constrained regular triangulations by flips. In: ANNUAL SYMPOSIUM ON COMPUTATIONAL GEOMETRY, SCG, 19., 2003. **Proceedings...** New York: ACM Press, 2003. p.181–190.

SILVA, L.; SCHARCANSKI, J. A Lossless Compression Approach for Mammographic Digital Images Based on the Delaunay Triangulation. In: **IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 2005**, Genova, Italy. **Proceedings...** [S.l.: s.n.], 2005.

SKARBEK, W.; KOSCHAN, A. **Segmentation of Colour Images - A survey**. Berlin: Department of Computer Science, Technical University of Berlin, 1994. (TR 94-32).

SONKA, M.; HLAVAC, V.; BOYLE, R. **Image processing, analysis and machine vision**. 2nd ed. Pacific Grove, CA, USA: Brooks/Cole, 1999.

SUNDARESWARA, R.; SCHRATER, P. Extensible Point Location Algorithm. In: **INTERNATIONAL CONFERENCE ON GEOMETRIC MODELING AND GRAPHICS, GMAG, 2003**. **Proceedings...** [S.l.: s.n.], 2003.

TRÉMEAU, A.; COLANTONI, P. Regions Adjacency Graph Applied to Color Image Segmentation. **IEEE Trans. on Image Processing**, [S.l.], v.9, n.4, p.735–744, April 2000.

TSENG, B. L.; LIN, C.-Y.; SMITH, J. R. Using MPEG-7 and MPEG-21 for Personalizing Video. **IEEE Multimedia**, [S.l.], v.11, n.1, p.42–52, January-March 2004.

VALETTE, S.; MAGNIN, I.; PROST, R. Active mesh for video segmentation and objects tracking. In: **IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 2001**, Thessaloniki, Greece. **Proceedings...** [S.l.: s.n.], 2001. v.2, p.77–80.

VIERGEVER, M.; ROOS, P. Hierarchical Interpolation. **IEEE Engineering in Medicine and Biology**, [S.l.], v.12, n.1, p.48–55, March 1993.

WEINBERGER, M.; SEROUSSI, G.; SAPIRO, G. LOCO-I: a low complexity, context-based lossless image compression algorithm. In: **DATA COMPRESSION CONFERENCE, 1996**. **Proceedings...** [S.l.: s.n.], 1996. p.140–149.

WELCH, T. A. A Technique for High Performance Data Compression. **IEEE Computer**, [S.l.], v.17, n.6, June 1984.

WU, X.; MEMON, N. Context-Based, Adaptive, Lossless Image Coding. **IEEE Trans. on Communications**, [S.l.], v.45, n.4, p.437–444, April 1997.

YANG, M.-S.; WU, K.-L. A Similarity-Based Robust Clustering Method. **IEEE Trans. Pattern Anal. Mach. Intell.**, [S.l.], v.26, n.4, p.434–448, 2004.

YANG, Y.; WERNICK, M. N.; BRANKOV, J. G. A Fast Approach for Accurate Content-Adaptive Mesh Generation. **IEEE Trans. on Image Processing**, [S.l.], v.12, n.8, p.866–881, August 2003.

ZHAOYANG, L.; XINGNI, Z.; XIQUAN, G.; CHANGXIN, F. Multiresolution Delaunay triangulation model-based image representation scheme. In: **IEEE INT. CONF. ELECTRICAL AND ELECTRONIC TECHNOLOGY, 2001**. **Proceedings...** [S.l.: s.n.], 2001. v.1, p.202–205.