

QUANTIFICAÇÃO DAS CARACTERÍSTICAS MORFOMÉTRICAS E DE
TEXTURA DE TUMORES DE MAMA EM IMAGENS POR ULTRA-SOM E A
AVALIAÇÃO DA SUA CONTRIBUIÇÃO DIAGNÓSTICA

André Victor Alvarenga

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS
EM ENGENHARIA BIOMÉDICA.

Aprovada por:

Prof. Antonio Fernando Catelli Infantosi, Ph.D.

Prof. Wagner Coelho de Albuquerque Pereira, D.Sc.

Prof. Marco Antonio von Krüger, Ph.D.

Prof. Eduardo Antonio Barros da Silva, Ph.D.

Profª. Annie France Frère Slaets, D.Sc.

Prof. Edson dos Santos Marchiori, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2005

ALVARENGA, ANDRÉ VICTOR

Quantificação das Características
Morfométricas e de Textura de Tumores
de Mama em Imagens por Ultra-som e a
Avaliação da sua Contribuição Diagnóstica
[Rio de Janeiro] 2005

XXI, 100 p. 29,7 cm (COPPE/UFRJ,
D.Sc., Engenharia Biomédica, 2005)

Tese – Universidade Federal do Rio de
Janeiro, COPPE

1. Câncer de mama
2. Ultra-som da mama
3. Parâmetros morfométricos e de textura

I. COPPE/UFRJ II. Título (série)

DEDICATÓRIA

Dedico este trabalho aos meus pais Gilson (*in memorian*) e Terezinha e meus padrinhos Octacílio e Sueli por todos os sacrifícios feitos em prol da minha educação, pela paciência e por todo o amor e carinho doados sem restrição. Vocês são tudo para mim. Com Amor!

Ao meu amor Aline. Que estará para sempre na caminhada de nossas vidas.

AGRADECIMENTOS

Aos meus Orientadores Wagner Coelho de Albuquerque Pereira e Antonio Fernando Catelli Infantosi por todo o tempo e cuidado dedicado ao trabalho. Sem a dedicação de vocês nosso trabalho não teria sido tão reconhecido.

A minha “Mãe de Mama” Dra. Carolina Maria de Azevedo por me contaminar com o “vírus” da luta contra o câncer de mama, por todo o conhecimento transferido e pela doação total para os avanços da pesquisa.

Aos “Amigos do Peito” Luciana Leite, Leonardo Muller, Fabiana Valadares, Isabela Miller e Christiano Machado por acreditarem que podemos contribuir para as pesquisas do câncer de mama.

Aos amigos Igor da Guia Rosa e Danilo Barbosa Melges pelo apoio durante todo o desenvolvimento da Tese e, principalmente, pelas sugestões para a apresentação do trabalho na Sessão Especial do Prêmio Cândido Pinto de Melo no CLAEB 2004. Vocês fazem parte da história desse prêmio.



*Uma imagem é a concretização de um sonho
Nem sempre a imagem o representa por completo
Nem sempre o completo se representa pela imagem
Um conjunto de pontos pretos é suficiente?
Um plano. Alguns pontos. Uma imagem. Duas almas. O amor.
Os padrões são identificados. A morfologia vem à tona.
Onde está o significado? Onde estarás?
A segmentação te apresenta. A imagem transcende seu significado.*



Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

QUANTIFICAÇÃO DAS CARACTERÍSTICAS MORFOMÉTRICAS E DE
TEXTURA DE TUMORES DE MAMA EM IMAGENS POR ULTRA-SOM E A
AVALIAÇÃO DA SUA CONTRIBUIÇÃO DIAGNÓSTICA

André Victor Alvarenga

Junho/2005

Orientadores: Wagner Coelho de Albuquerque Pereira
Antonio Fernando Catelli Infantsi

Programa: Engenharia Biomédica

Este trabalho apresenta o desempenho de sete parâmetros morfométricos e 20 de textura em quantificar as características morfológicas de 272 tumores de mama em imagens por ultra-som (US), bem como o desempenho destes na distinção dos tumores em maligno ou benigno. As imagens foram submetidas, previamente, a um método de segmentação baseado em Morfologia Matemática, tendo sido capaz de segmentar cerca de 90% destas. Aplicando Análise Discriminante Linear e utilizando a curva ROC como figura de mérito, o melhor desempenho, entre os parâmetros morfométricos, foi obtido com o valor médio quadrático residual normalizado (*nrv*), proposto neste trabalho, que distinguiu corretamente quase 88% dos tumores. Entre os parâmetros de textura destacaram-se o contraste (*con*) da matriz de co-ocorrência e o valor máximo (*mv*) da curva de complexidade. Estes últimos, quando combinados, foram capazes de classificar corretamente aproximadamente 80,0% dos tumores de mama. Os melhores parâmetros morfométricos e de textura, em conjunto, resultaram em área sob a curva ROC de 0,93, acurácia de 92,8%, sensibilidade de 92,0% e especificidade de 94,2%. Tais achados sugerem o emprego destes parâmetros no auxílio ao diagnóstico de tumores de mama em imagens por US, contribuindo, portanto, para a redução no número de biópsias.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

QUANTIFICATION OF MORPHOMETRIC AND TEXTURE CHARACTERISTICS
OF BREAST TUMOURS FROM ULTRASOUND IMAGES AND EVALUATION OF
THEIR DIAGNOSTIC CONTRIBUTION

André Victor Alvarenga

June/2005

Advisors: Wagner Coelho de Albuquerque Pereira
Antonio Fernando Catelli Infantosi

Department: Biomedical Engineering

The performance of seven morphometric and 20 texture parameters for quantifying the morphologic characteristics of 272 breast tumours from ultrasonic images was investigated. Besides, their contributions in distinguishing tumours into malign or benign were assessed. First, a segmentation procedure based on Mathematical Morphology was developed allowing the segmentation of about 90% of the tumours. By applying the Linear Discriminant Analysis and using the ROC curve, the normalised residual mean square (*nrv*) morphometric parameter, introduced in this work, was found to be the most effective one (accuracy of 88%). Concerning the texture parameters, using both contrast (*con*), calculated using the co-occurrence matrix, and the maximum value (*mv*) of the complexity curve resulted in nearly 80% of the tumours correctly classified. Joining the best morphometric and texture parameters led to higher performance in the classification, that is, an area under the ROC curve of 0.93, accuracy of 92.8%, sensitivity of 92.0% and specificity of 94.2%. This finding suggests these parameters to be helpful in breast tumours diagnosis, based on ultrasonic images.

ÍNDICE

LISTA DE FIGURAS	X
LISTA DE TABELAS.....	XVII
LISTA DE SÍMBOLOS E SIGLAS.....	XIX
I INTRODUÇÃO.....	1
I.1 OBJETIVO.....	3
I.2 ESTRUTURA DO TEXTO	3
II DIAGNÓSTICO DO CÂNCER DE MAMA.....	4
II.1 A LUTA CONTRA O CÂNCER DE MAMA	4
II.2 O ULTRA-SOM NO DIAGNÓSTICO DO CÂNCER DE MAMA	6
II.3 METODOLOGIAS DE AUXÍLIO AO DIAGNÓSTICO.....	8
III OPERADORES MORFOLÓGICOS.....	10
III.1 OPERADORES BÁSICOS.....	10
III.2 TRANSFORMAÇÕES GEODÉSICAS E OPERADORES BASEADOS EM RECONSTRUÇÃO.....	12
III.3 ABERTURA E FECHAMENTO POR RECONSTRUÇÃO	13
III.4 IMPOSIÇÃO DE MÍNIMOS	14
III.5 WATERSHED	15
IV SEGMENTAÇÃO DE TUMORES DE MAMA EM IMAGENS POR US USANDO MORFOLOGIA MATEMÁTICA.....	17
IV.1 BANCO DE IMAGENS US DE MAMA.....	18
IV.2 MÉTODO DE SEGMENTAÇÃO: SAC (<i>SEMI-AUTOMATIC CONTOUR</i>).....	20
IV.2.1 <i>Detecção do Contorno</i>	23
IV.3 SEGMENTAÇÃO DE IMAGENS SIMULADAS	24
IV.4 SEGMENTAÇÃO DE IMAGENS DE US DE MAMA USANDO SAC.....	25
V PARÂMETROS MORFOMÉTRICOS NA DISTINÇÃO DE TUMORES: CONTORNO E DIAGNÓSTICO.....	29
V.1 TÉCNICAS PARA QUANTIFICAR AS CARACTERÍSTICAS MORFOMÉTRICAS DE IMAGENS	29
V.1.1 <i>Comprimento Radial Normalizado (Normalized Radial Length)</i>	30
V.1.2 <i>Razão entre largura e profundidade</i>	31
V.1.3 <i>Circularidade</i>	31
V.1.4 <i>Número Substancial de Protuberâncias e Depressões</i>	32
V.1.5 <i>Índice de Lobulação</i>	32
V.1.6 <i>Perímetro Normalizado por Elipse Equivalente (Elliptic-Normalized circumference)</i>	33
V.1.7 <i>Esqueleto Normalizado por Elipse Equivalente (Elliptic-Normalized Skeleton)</i>	34
V.1.8 <i>Razão de Superposição</i>	35
V.1.9 <i>Valor Médio Quadrático Residual Normalizado</i>	35
V.1.10 <i>Mshape</i>	36
V.2 METODOLOGIA DE AVALIAÇÃO DE DESEMPENHO.....	36
V.2.1 <i>Análise Discriminante Linear</i>	36
V.2.2 <i>A Curva ROC</i>	37
V.3 APLICAÇÃO DE PARÂMETROS MORFOMÉTRICOS PARA CLASSIFICAÇÃO DE TUMORES DE MAMA EM IMAGENS POR US.....	40
V.4 CÁLCULO DOS PARÂMETROS MORFOMÉTRICOS PARA IMAGENS DE US DE MAMA.....	41
V.5 AVALIANDO OS PARÂMETROS MORFOMÉTRICOS NA DISTINÇÃO DOS TUMORES: CONTORNO E DIAGNÓSTICO.....	42
V.5.1 <i>Irregular x regular</i>	42
V.5.2 <i>Maligno x benigno</i>	46
V.6 DISCUSSÃO	50
VI PARÂMETROS DE TEXTURA NA DISTINÇÃO DOS TUMORES: ECOTEXTURA E DIAGNÓSTICO.....	52

VI.1	TÉCNICAS PARA QUANTIFICAR AS CARACTERÍSTICAS DE TEXTURA DE IMAGENS	52
VI.1.1	<i>Parâmetros estatísticos do histograma da imagem</i>	53
VI.1.2	<i>Matriz de Co-ocorrência</i>	53
VI.1.3	<i>Run-Length Analysis</i>	55
VI.1.4	<i>Markoviano</i>	57
VI.1.5	<i>Coeficientes de Autocovariância Bi-dimensional Normalizados</i>	58
VI.1.6	<i>Curva de Complexidade</i>	58
VI.1.7	<i>Diferença Mínima Lateral</i>	61
VI.2	APLICAÇÃO DE PARÂMETROS DE TEXTURA NA CLASSIFICAÇÃO DE TUMORES DE MAMA EM IMAGENS POR US	62
VI.3	CÁLCULO DOS PARÂMETROS DE TEXTURA PARA IMAGENS DE US DE MAMA	63
VI.4	AVALIANDO OS PARÂMETROS DE TEXTURA NA DISTINÇÃO DOS TUMORES: ECOTEXTURA E DIAGNÓSTICO.....	67
VI.4.1	<i>Tecido adiposo x glandular</i>	67
VI.4.2	<i>Homogêneo x heterogêneo</i>	70
VI.4.3	<i>Maligno x benigno</i>	75
VI.5	DISCUSSÃO	80
VII	PARÂMETROS MORFOMÉTRICOS E DE TEXTURA: DISTINÇÃO DE TUMORES QUANTO AO DIAGNÓSTICO.....	82
VII.1	APLICAÇÃO DOS PARÂMETROS MORFOMÉTRICOS E DE TEXTURA TOMADOS EM CONJUNTO NA CLASSIFICAÇÃO DE TUMORES DE MAMA EM IMAGENS POR US	82
VII.2	AVALIANDO PARÂMETROS MORFOMÉTRICOS E DE TEXTURA TOMADOS EM CONJUNTO	83
VII.3	DISCUSSÃO	87
VIII	CONCLUSÃO	89
IX	REFERÊNCIAS BIBLIOGRÁFICAS.....	91
X	APÊNDICE I.....	98

Lista de Figuras

- Figura III.1. Exemplo de **(a)** elemento estruturante B de tamanho 3×3 *pixels* em forma de “L”, com centro definido pelo ponto preto, e **(b)** seu transposto B^T 10
- Figura III.2. **(a)** Imagem a ser processada f . **(b) Erosão** e **(c) dilatação** usando um ES quadrado com 5 *pixels* de lado.....11
- Figura III.3. **(a)** Imagem a ser processada f . **(b) Abertura** e **(c) fechamento** usando um ES circular de raio 19 *pixels*.....12
- Figura III.4. **(a)** Imagem original g usada como máscara. **(b)** Imagem f usada como marca. **(c)** Resultado da **reconstrução por dilatação** da máscara g a partir da marca f13
- Figura III.5. **(a)** Imagem original após **(b)** uma **erosão** usando um ES quadrado (3×3 *pixels*). **(c) Dilatação por reconstrução** até que coincida com a original (mesmo ES).....14
- Figura III.6. Exemplo da aplicação da **imposição de mínimos**. **(a)** Curva f (sólida) e a marca f_m (pontilhada). **(b)** Mínimo entre as curvas f e f_m . **(c)** Resultado da **imposição de mínimos** (sólida) sobre a curva f (pontilhada).....14
- Figura III.7. **(a)** Imagem original contendo três mínimos locais (em cinza) respectivamente **(b)** rotulados com os valores 1, 2 e 3.15
- Figura III.8. **(a)** Análise dos *pixels* da vizinhança definida pelo elemento estruturante (em cinza escuro) e **(b)** rotulados como pertencendo ao vale definido pelo respectivo mínimo local.....16
- Figura III.9. Resultado do processo de análise da vizinhança obtendo as linhas divisórias (em cinza escuro) das regiões de mínimos locais.16
- Figura IV.1. Exemplo de imagem original de um tumor de mama adquirida com o equipamento de US de 7,5 MHz (*Sonoline – Sienna* ®*Siemens*) e gravada no formato TIF.....19
- Figura IV.2. Exemplos de ROI contendo tumores de contorno **(a)** irregular e **(b)** regular, determinados pelo radiologista.19

Figura IV.3. Exemplos de ROI contendo de tumores de ecotextura (a) heterogênea e (b) homogênea classificados pelo radiologista.....	20
Figura IV.4. Exemplo de regiões saudáveis com predominância de (a e b) tecido glandular e (c e d) tecido adiposo, selecionadas pelo radiologista.....	20
Figura IV.5. (a) Imagem original de um tumor de mama. (b) Após a aplicação da abertura por reconstrução para evidenciar as diferenças. (c) Imagem gerada a partir dos <i>pixels</i> rotulados e ordenados. Esta contém um número menor de níveis de cinza que a imagem original (a)	21
Figura IV.6. (a) Histograma da imagem (VII.2b). (b) Histograma (a) filtrado por uma abertura morfológica (linha tracejada) e rótulos numéricos (linha contínua) gerados pelo <i>watershed</i> . (c) Rótulos numéricos originais (curva tracejada) e ordenados (curva contínua).....	22
Figura IV.7. (a) Imagem (IV.1c) após aplicação da imposição de mínimos e (b) respectivo histograma, onde cada pico representa uma região rotulada. (c) Curva da área cumulativa do histograma (IV.3b) e seu (d) respectivo gradiente, onde o pico representa a transição entre a área da ROI e o restante a imagem.....	23
Figura IV.8. (a) Imagem original. (b) Após a aplicação da segmentação. (c) Após aplicação do operador de fechamento por reconstrução na imagem (b) . (d) Contorno determinado após aplicação do <i>Laplaciano</i>	24
Figura IV.9. (a) Imagem original. (b) Após a aplicação da segmentação. (c) Após aplicação do operador de fechamento por reconstrução na imagem (b)	25
Figura IV.10. (a) ROI de imagem de US de mama. (b) Após a realização da segmentação. (c) Após aplicação do operador de fechamento por reconstrução na imagem (b) . (d) Contorno determinado após aplicação do <i>Laplaciano</i>	25
Figura IV.11. (a) ROI de imagem de US de mama. (b) Após a realização da segmentação. (c) Após aplicação do operador de fechamento por reconstrução . (d) Contorno determinado após aplicação do <i>Laplaciano</i>	26
Figura IV.12. (a) ROI contendo tumor de ecotextura mista e elevada heterogeneidade e (b) respectivo contorno resultante da aplicação do SAC. Nota-se que o SAC não foi capaz de	

determinar completamente os limites entre o tumor e o tecido adjacente, principalmente na sua porção superior (seta)	26
Figura IV.13. (a) ROI contendo tumor de ecotextura da interna similar ao tecido adjacente. Nota-se que (b) seu respectivo contorno, resultante da aplicação do SAC, excedeu os limites do tumor (setas)	27
Figura IV.14. (a) ROI de imagem de tumor de mama de contorno regular. Contornos estabelecidos pelo (b) 1º radiologista, (c) 2º radiologista e (d) SAC.....	28
Figura IV.15. (a) ROI de imagem de tumor de mama de contorno irregular. Contornos estabelecidos pelo (b) 1º radiologista, (c) 2º radiologista e (d) SAC.....	28
Figura V.1. Exemplo da determinação da (b) distância radial normalizada num (a) tumor previamente segmentado.....	30
Figura V.2. Representação da largura (linha contínua) e profundidade (linha pontilhada) de um objeto em relação a posição do transdutor de US.....	31
Figura V.3. Exemplo de objeto (contorno com protuberâncias e depressões) e seu respectivo polígono convexo (menor contorno convexo)	32
Figura V.4. Desenho representando os lóbulos em um objeto, onde $[w_1, w_2, w_3, w_4]$ são quatro pontos de mínimos das concavidades presentes no contorno e $[A_1, A_2, A_3, A_4]$ são as áreas dos lóbulos.....	33
Figura V.5. Exemplo de (a) um objeto quadrado medindo 6 x 8 (em preto os pontos do contorno), (b) o objeto após a transformação de distância, onde cada valor indica a distância, em <i>pixels</i> , até o ponto mais próximo do contorno. Nota-se ainda a janela de define a vizinhança para a determinação do máximo local. (c) representação com os valores máximos locais. (d) esqueleto (em cinza) do objeto (a)	34
Figura V.6. Um (a) objeto e contorno irregular e seu respectivo (b) esqueleto. Note que devido à irregularidade do contorno, o esqueleto é mais complexo que o obtido na Figura V.5.....	35

Figura V.7. Exemplo de fechamento morfológico (em branco) de um objeto (a) regular e outro (b) irregular.....	36
Figura V.8. Exemplo das distribuições de probabilidade das amostras benignas e malignas para análise da curva ROC, supondo um possível limiar. Nota-se a superposição entre as distribuições das amostras benignas e malignas, região onde ocorrem as classificações Falso-Positivo e Falso Negativo.....	39
Figura V.9. Exemplo de curva ROC. Os (*) representam os pontos de operação sensibilidade x (1-especificidade) sobre a curva.....	39
Figura V.10. Polígono convexo (em branco) de dois tumores de mama (a) regular e (b) irregular, segmentados usando SAC.....	41
Figura V.11. Curva ROC dos três parâmetros de melhor desempenho analisados individualmente. As setas indicam a região (E entre 65,5% e 98,0%) onde se obtém o melhor desempenho pelo parâmetro nrv	43
Figura V.12. Diagrama de espalhamento nrv x C e a função linear que melhor separa os tumores irregulares dos regulares, reta $y = -2,26x + 0,38$	44
Figura V.13. Diagrama de espalhamento de nrv x C x R e o plano ($z = -2,48x - 3,45y - 0,62$) que melhor separa os tumores irregulares dos regulares.....	45
Figura V.14. Curvas ROC das combinações de dois e três parâmetros de melhor desempenho na separação entre tumores quanto ao contorno. A seta indica o ponto da curva onde se obtém o melhor desempenho ($S = 95,7\%$ e $E = 96,7\%$).....	46
Figura V.15. Diagrama de espalhamento de nrv x D_{NRL} x R e o plano ($z = -2,26x + 0,55y + 0,10$) que melhor separa os tumores malignos dos benignos.....	48
Figura V.16. Diagrama de espalhamento de nrv x C x R e o plano ($z = -1,43x - 0,99y - 0,05$) que melhor separa os tumores malignos dos benignos.....	49
Figura V.17. Curvas ROC das combinações de três parâmetros de melhor desempenho na separação dos tumores quanto ao diagnóstico. A seta indica o ponto da curva onde se obtém o melhor desempenho ($S = 86,0\%$ e $E = 94,2\%$).....	49

Figura VI.1. Determinação da GLCM (b) $m(i,j)$, a partir da imagem (a) $f(x,y)$ de dois níveis de cinza (0 e 1, sendo o último mais escuro) para distância k de dois <i>pixels</i> e direção θ nula. A matriz $m(i,j)$ resulta em 10 pares $\{0,0\}$ e em $\{1,1\}$, e nenhum par em $\{0,1\}$ e em $\{1,0\}$	54
Figura VI.2. (a) Imagem $f(x,y)$ e (b) sua respectiva matriz <i>run-length</i> , para a direção horizontal. Notam-se grifadas as corridas de comprimento $l = 2$, para o nível de cinza $i = 1$ (Figura II.2a) e sua respectiva quantidade corridas $r(1,2) = 4$	56
Figura VI.3. (a) Imagem $f(x,y)$ com 16 níveis de cinza e (b) sua respectiva imagem binária $f^\alpha(x,y)$ para um valor de limiar $\alpha = 12$	59
Figura VI.4. Curva de Complexidade $C(\alpha)$ para a imagem $f(x,y)$ da Figura II.3a.....	60
Figura VI.5. Representação das regiões utilizada para definir o parâmetro <i>MSD</i>	62
Figura VI.6. Diagrama esquemático de uma ROI, com o tumor em cinza, apontando as regiões contempladas pelos parâmetros de GLCM e CC.	64
Figura VI.7. Exemplos (c) GLCM calculada a partir de ROI de (a) tumor de ecotextura homogênea e (d) GLCM da (b) região interna do mesmo, além das respectivas (e) CCs.....	65
Figura VI.8. Exemplos (c) GLCM calculada a partir de ROI de (a) tumor de ecotextura heterogênea e (d) GLCM da (b) região interna do mesmo, além das respectivas (e) CCs.....	66
Figura VI.9. Diagrama de espalhamento dos parâmetros calculados a partir de CC: (a) mv x av . (b) sm x ssd . (c) mv x ent	68
Figura VI.10. Diagrama de espalhamento dos parâmetros calculados a partir de GLCM: (a) std x coo . (b) con x asm . (c) cor x con	69
Figura VI.11. Diagrama de espalhamento dos parâmetros mv_i e con_i	72
Figura VI.12. Diagrama de espalhamento dos parâmetros av e ent_i	73

Figura VI.13. Curvas ROC do desempenho obtido considerando-se dois, cinco e onze parâmetros de textura na separação dos tumores quanto a ecotextura. As setas indicam a região (especificidade entre 86,7% e 91,8%) onde os parâmetros apresentam desempenho similar.....74

Figura VI.14. Diagrama de espalhamento entre os parâmetros mv_i e con . Nota-se que os tumores malignos tendem a apresentar valores elevados de contraste (con), enquanto que os valores de mv_i tendem a ser elevados em tumores benignos.....77

Figura VI.15. Curvas ROC do desempenho obtido considerando-se dois, nove e doze parâmetros de textura na separação dos tumores quanto ao diagnóstico. Nota-se que na região onde a sensibilidade varia de 82,0 a 87,0% (setas), o conjunto de nove parâmetros apresenta uma especificidade maior (88,5%) que o de 12 parâmetros (86,5%).....78

Figura VI.16. Diagrama de espalhamento entre os parâmetros mv_i e ent . Nota-se que os tumores malignos tendem a apresentar valores de ent mais elevados que os tumores benignos.....79

Figura VI.17. Diagrama de espalhamento entre os parâmetros con e std . Nota-se que os tumores malignos tendem a apresentar valores de std menores, além de estarem mais agrupados que os benignos.....80

Figura VII.1. Diagrama de espalhamento de nrv x std_i e o plano ($y = -3,3x + 0,83$) que melhor separa os tumores malignos dos benignos.....84

Figura VII.2. Diagrama de espalhamento de nrv x R x std e o plano ($z = -3,93x - 0,92y + 0,91$) que melhor separa os tumores malignos dos benignos.....85

Figura VII.3. Curvas ROC do desempenho obtido considerando-se dois, oito e onze parâmetros de textura e morfométricos combinados na separação dos tumores quanto ao diagnóstico. Nota-se que na região onde a especificidade varia de 75,0% a 92,3% (setas) os conjuntos de 8 e 11 parâmetros mostram um comportamento idêntico.....85

Figura VII.4. Curvas ROC do desempenho obtido na separação dos tumores em malignos ou benignos, entre das 246 amostras, quando aplicadas à LDA obtida com as 152 amostras considerando-se a melhor combinação entre parâmetros morfométricos e de textura (onze parâmetros).....86

Figura A1. Exemplo da projeção das amostras de duas classes (quadrado e círculo) projetadas sobre **(a)** uma reta que não as separa e **(b)** outra que as separa.....98

Figura A2. Projeção $v'x_i$ da amostra x_i , na direção do vetor unitário v98

Lista de Tabelas

Tabela IV.1. Total de tumores segmentados com o SAC.....	28
Tabela V.1. Desempenho individual dos parâmetros na classificação dos tumores, quanto ao contorno, em irregular ou regular.....	43
Tabela V.2. Desempenho das melhores combinações entre os parâmetros na classificação dos tumores, quanto ao contorno, em irregular ou regular.....	44
Tabela V.3. Desempenho individual dos parâmetros na distinção dos tumores entre benigno ou maligno.....	48
Tabela V.4. Desempenho das melhores combinações entre os parâmetros na distinção dos tumores entre benigno ou maligno.....	48
Tabela VI.1. Desempenho individual de cada um dos 10 parâmetros de textura, ordenados por A_z , na distinção entre regiões da mama com predominância de tecido glandular ou adiposo.....	70
Tabela VI.2. Desempenho individual de cada um dos 20 parâmetros de textura parâmetros de textura na separação entre tumores heterogêneos ou homogêneos, ordenados por A_z	71
Tabela VI.3. Desempenho das melhores combinações de parâmetros na distinção entre tumores de mama de textura heterogênea ou homogênea.....	74
Tabela VI.4. Desempenho individual de cada um dos vinte parâmetros de parâmetros de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z	75
Tabela VI.5. Desempenho das melhores combinações de parâmetros de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z	76
Tabela VI.6. Desempenho, na distinção quanto ao diagnóstico, das melhores combinações de parâmetros de textura obtido usando-se, separadamente, parâmetros calculados a partir de CC e GLCM.....	79

Tabela VI.7. Desempenho, na distinção quanto ao diagnóstico, das melhores combinações de parâmetros de textura obtido usando-se, separadamente, parâmetros calculados a partir da ROI e da região interna do tumor.....79

Tabela VII.1. Desempenho das melhores combinações de parâmetros morfométricos e de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z84

Tabela VII.2. Desempenho da separação dos tumores em malignos ou benignos, entre das 246 amostras, quando aplicadas a LDA obtida com as 152 amostras considerando-se a melhor combinação entre parâmetros morfométricos e de textura (onze parâmetros).....86

Lista de Símbolos e Siglas

A	área total em <i>pixels</i> de um objeto em uma imagem binária
A_{max}	área do maior lóbulo de um objeto em uma imagem binária
A_{min}	área do menor lóbulo de um objeto em uma imagem binária
A_R	Razão de área do comprimento radial normalizado
asm	Segundo momento angular
av	Valor médio das transições na Curva de Complexidade
A_z	Área sob a curva ROC
B	Elemento estruturante
C	Circularidade de um objeto em uma imagem binária
$c(\alpha)$	Função da curva de complexidade (função)
CC	Curva de Complexidade (sigla)
con	Contraste da GLCM
coo	Entropia da GLCM
cor	Correlação da GLCM
d	Distância separando dois <i>pixels</i>
$d(i)$	Comprimento radial normalizado
\bar{d}	Média do comprimento radial normalizado
DV	<i>variância da diferença</i>
DWR	Razão entre largura e profundidade
ENC	Perímetro Normalizado por Elipse Equivalente
ENS	Esqueleto Normalizado por Elipse Equivalente
ent	Entropia da Curva de Complexidade
$f(x,y)$	Imagem digital, onde (x,y) representa as coordenadas de um dado <i>pixel</i>
\bar{f}	Média dos níveis de cinza de $f(x,y)$
$f^\alpha(x,y)$	Imagem $f(x,y)$ binarizada para dado nível de cinza α
LDA	Análise Discriminante Linear
$\bar{g}_{posterior}$	Valor médio dos níveis de cinza na região posterior da lesão
$\bar{g}_{esquerda}$	Valor médio dos níveis de cinza na região posterior da lesão à esquerda da lesão
$\bar{g}_{direita}$	Valor médio dos níveis de cinza na região posterior da lesão à direita da lesão
G	Nível de cinza máximo em uma imagem
gl	<i>Gray-Level Nonuniformity</i> do <i>run-length</i>
GLCM	Matriz de co-ocorrência (sigla)

ic	Contraste inverso de M
IMC_2	Medida de informação da correlação
j	
$J(\mathbf{v})$	Função objetivo do discriminante de linear
k	Distância entre dois <i>pixels</i>
l	Comprimento em <i>pixels</i> de uma “corrida” da <i>run-length</i>
LI	Índice de lobulação
lre	<i>Long Run Emphasis</i> do <i>run-length</i>
m	Média
$m(i,j)$	Matriz de co-ocorrência (função), onde i e j são valores de nível de cinza de dois <i>pixels</i> quaisquer
Mk	Matriz da estatística Markoviana
MSD	Diferença mínima lateral
$Mshape$	Razão entre a área do objeto sob avaliação e o seu fechamento morfológico
Mv	Valor máximo de transições na Curva de Complexidade
$M \times N$	Dimensões de uma imagem $f(x,y)$
N_l	Número de lóbulos de um objeto em uma imagem binária
n_p	Número total de <i>pixels</i> na imagem
n_r	Número total de <i>runs</i>
NRG	Gradiente radial normalizado
NRL	Comprimento radial normalizado
nrv	Valor Médio Quadrático Residual Normalizado
$NSPD$	Número Substancial de Protuberâncias e Depressões
P	perímetro em <i>pixels</i> do contorno de um objeto em uma imagem binária
$p_k(i,j)$	Probabilidade da matriz Markoviana de um nível de cinza i ocorrer k <i>pixels</i> de um nível de cinza j
$r(i,l)$	Matriz <i>run-length</i> (função)
R	Rugosidade do comprimento radial normalizado
$R_g^*(f)$	Reconstrução por erosão da imagem f com respeito à máscara g
$R_g(f)$	Reconstrução por dilatação da imagem f com respeito à máscara g
RA	Razão de área do comprimento radial normalizado
rl	<i>Run-Length Nonuniformity</i>
ROC	Característica de resposta do observador
ROI	Região de interesse
rp	<i>Run Percentage</i> do <i>run-length</i>
r_r	Resíduo

RS	Razão de superposição
S	Objeto em uma imagem
s^2	Variância
SAC	Algoritmo de segmentação semi-automático
sm	Valor médio dos níveis de cinza promediados pelo número de transições da Curva de Complexidade
sre	<i>Short Run Emphasis</i> do <i>run-length</i>
ssd	Desvio padrão dos níveis de cinza promediado pelo número de transições da Curva de Complexidade
std	Desvio padrão da GLCM
$T_{i,j}^\alpha$	Número total de transições em $f^\alpha(x,y)$ com um passo (i,j)
US	Ultra-som
$x(i),y(i)$	Coordenadas dos <i>pixels</i> que pertencem ao contorno de um objeto em uma imagem binária
X_0, Y_0	Coordenadas do centróide de um objeto em uma imagem binária
α	Um dado nível de cinza na Curva de Complexidade
$\gamma_B(f)$	Abertura morfológica da imagem f com o elemento estruturante B
$\gamma_R^{(n)}(f)$	Abertura por reconstrução da imagem f
$\delta_B(f)$	Dilatação morfológica da imagem f com o elemento estruturante B
$\delta_g^{(1)}(f)$	Dilatação geodésica de tamanho 1 da imagem f com respeito à máscara g
$\varepsilon_B(f)$	Erosão morfológica da imagem f com o elemento estruturante B
$\varepsilon_g^{(1)}(f)$	Erosão geodésica de tamanho 1 da imagem f com respeito à máscara g
σ	Desvio padrão
$\phi_R^{(n)}$	Fechamento por reconstrução da imagem f
$\nabla^2 f(x,y)$	Gradiente bidimensional da imagem $f(x,y)$
θ	Ângulo
$\rho(p,q)$	Autocovariância normalizada
ψ^2	Valor médio quadrático
$\varphi_B(f)$	Fechamento morfológico da imagem f com o elemento estruturante B

I INTRODUÇÃO

No Brasil, o câncer de mama é um dos que mais causa morte entre as mulheres, sendo o diagnóstico precoce imprescindível para aumentar as chances de cura. Segundo o Ministério da Saúde (MS-Brasil), “... o investimento nas áreas de prevenção e diagnóstico precoce traz, assim, maior benefício social e econômico do que o custeio do tratamento da doença na fase avançada” (INCa, 1997).

A mamografia tem sido considerada a única técnica diagnóstica que comprovadamente contribui, por meio de um programa de acompanhamento periódico, para a redução da mortalidade por câncer de mama (SKAANE, 1999). Porém, sua acurácia depende da composição do parênquima mamário e das características tissulares dos tumores, pois um parênquima denso dificulta a localização e até mesmo o diagnóstico radiológico de um tumor (AZEVEDO, 1994). Devido a esta limitação, o exame da mama por meio de imagens por ultra-som (US) tem emergido como o mais importante coadjuvante da mamografia, em particular para pacientes com massas palpáveis e mamografias não-conclusivas (SKAANE, 1999).

Tumores malignos geralmente se infiltram no tecido adjacente resultando em irregularidades no contorno (AZEVEDO, 1994). Características morfológicas do contorno, tais como espículas, margens anguladas e irregulares, em imagens de mama por US, sugerem malignidade (CHOU *et al.*, 2001). Além do contorno, os ecos provenientes das regiões interna e adjacente contêm informação sobre a organização do tecido na região do tumor (DONOHUE *et al.*, 2001); uma ecotextura heterogênea tende a ser uma característica de malignidade (SKANNE, 1999). Tais aspectos apontam para a análise do contorno e da ecotextura de um tumor como abordagem para o estabelecimento de uma hipótese diagnóstica.

Com o intuito de reduzir o número de biópsias realizadas em tumores sólidos benignos e aumentar a consistência da interpretação diagnóstica do US, diversos trabalhos têm proposto metodologias para auxiliar no diagnóstico do câncer de mama em imagens por US (HORSCH *et al.*, 2001; DRUKKER *et al.*, 2002; HUBER *et al.*, 2000; CHANG *et al.*, 2003; CHEN *et al.*, 2003b; CHEN D. *et al.*, 2000; LEFEBVRE *et al.*, 2000). Estas metodologias têm como base comum o cálculo de parâmetros que visam quantificar

características morfológicas dos tumores de mama, tais como forma e textura, sendo, então, avaliados quanto à capacidade em distinguir tumores benignos de malignos. O estudo de técnicas para o cálculo de parâmetros, bem como a determinação daqueles que melhor distinguem entre tumores malignos e benignos, são etapas complexas e ainda objeto de investigação (TOURASSI, 1999).

Outro aspecto fundamental para o uso destas metodologias é a segmentação da imagem do tumor, pois é a partir desta que se determinam os parâmetros requeridos para a quantificação das características morfométricas. Em imagens de mama por US, a segmentação tem sido geralmente realizada por radiologista experiente que delinea as margens do tumor com auxílio do *mouse*, a partir da imagem exibida em monitor de computador (LEFEBVRE *et al.*, 2000; HUBER *et al.*, 2000; CHOU *et al.*, 2001; CHEN *et al.*, 2003b). Devido ao componente subjetivo deste procedimento, existe uma elevada variabilidade entre os observadores na interpretação das imagens (BARR, 2001; HUBER *et al.*, 2000; RAHBAR *et al.*, 1999; SKAANE, 1999). Portanto, a utilização de um método de segmentação automático adequado, precedendo o cálculo de parâmetros que quantifiquem as características morfológicas dos tumores, pode contribuir para a redução desta variabilidade.

Entretanto, a segmentação de uma região de interesse em aplicações médicas é de extrema dificuldade devido às restrições impostas pelo processo de aquisição da imagem, à patologia e a variações biológicas (OLABARRIAGA & SMEULDERS, 2001). Particularmente em imagens por US, a segmentação é dificultada pela natureza textural deste tipo de imagem. Esta característica resulta de fatores como o processamento do sinal para a formação da imagem, as propriedades físicas do tecido, o *speckle* e artefatos (CHEN *et al.*, 2001). Todavia, visando contornar esta dificuldade, métodos de segmentação de imagens por US têm sido investigados por diversos autores (CHEN, C. *et al.*, 2000; CHEN *et al.*, 2001; ZIMMER *et al.*, 1999, ZIMMER *et al.*, 1996), alguns dos quais aplicados diretamente a imagens da mama (HORSCH *et al.*, 2001, BOUKERROUI *et al.*, 2003, HUANG *et al.*, 2004).

I.1 Objetivo

Este trabalho tem como objetivo caracterizar quantitativamente, por meio de parâmetros morfométricos e de textura, tumores de mama em imagens por ultra-som e avaliar seu potencial diagnóstico. Ainda, com vistas a enfatizar detalhes do contorno de tumores, um método de segmentação de imagens de mama baseado em Morfologia Matemática é proposto.

I.2 Estrutura do Texto

No Capítulo II, apresenta-se revisão bibliográfica sobre o câncer de mama, o emprego do ultra-som em seu diagnóstico e as metodologias empregadas para a quantificação de características morfológicas dos tumores de mama em imagens por US. Uma breve introdução aos Operadores Morfológicos é apresentada no Capítulo III. Estes são aplicados no método de segmentação de tumores de mama em imagens por ultra-som detalhado no Capítulo IV, que apresenta ainda as imagens de US de mama utilizadas neste trabalho e os resultados da aplicação do método de segmentação nestas e em imagens simuladas.

O Capítulo V, no item V.1, contém uma descrição sucinta das técnicas que têm sido empregadas para calcular parâmetros morfométricos, algumas das quais são utilizadas neste trabalho. No item V.2, a Análise Discriminante Linear, classificador utilizado para avaliar o desempenho dos parâmetros estudados, é apresentada, e sua dedução matemática detalhada no Apêndice I. Ainda neste item, apresenta-se a utilização da curva ROC como uma medida de desempenho para o classificador. A revisão da aplicação dos parâmetros morfométricos na classificação dos tumores de mama é apresentada no item V.3, seguida do cálculo (V.4), avaliação (V.5) e discussão (V.6) do desempenho destes parâmetros, na distinção dos tumores quanto ao contorno e ao diagnóstico. O Capítulo VI apresenta descrição semelhante para os parâmetros de textura.

O desempenho combinado dos parâmetros morfométricos e de textura é descrito e discutido no Capítulo VII, e, por fim, as Conclusões e as Referências Bibliográficas são apresentadas nos Capítulos VIII e IX, respectivamente.

II DIAGNÓSTICO DO CÂNCER DE MAMA

II.1 A Luta Contra O Câncer de Mama

“Cada tarde, em uma minúscula sala de trabalho no 12º andar da clínica Mayo, a batalha contra o câncer começa com uma discussão. À uma hora da tarde, o staff de apoio da mundialmente famosa clínica em Rochester, Minnesota, enche as paredes da sala com radiografias e tomografias do caso de estudo atual” (ROBERTS, 1999).

O câncer de mama representa, nos países ocidentais, uma das principais causas de morte entre as mulheres. As estatísticas apontam o seu crescimento tanto em países desenvolvidos quanto em países em desenvolvimento (INCa, 2005). Dos 467.440 novos casos de câncer, com previsão de serem diagnosticados em 2005, o de mama será o principal a atingir a população feminina, sendo o responsável por 49.470 novos casos (INCa, 2005).

Apesar do câncer de mama ser considerado de relativamente bom prognóstico, se diagnosticado e tratado oportunamente, as taxas de mortalidade por este câncer continuam elevadas no Brasil, muito provavelmente porque a doença ainda seja diagnosticada em estádios avançados (INCa, 2005).

O controle do câncer de mama no Brasil é considerado um problema de saúde pública. Assim, programas que estimulam o diagnóstico precoce da doença, tais como o *“Viva Mulher – Programa Nacional de Controle do Câncer do Colo do Útero e de Mama”* promovido pelo Ministério da Saúde, por meio do Instituto Nacional de Câncer – INCa têm o objetivo de *“reduzir a mortalidade e as repercussões físicas, psíquicas e sociais desses cânceres na mulher brasileira, por meio da oferta de serviços para prevenção e detecção em estágios iniciais da doença e do tratamento e reabilitação das mulheres”* (INCa, 2002).

Quanto mais precoce se detecta o câncer de mama, maior é a possibilidade de cura e, como consequência, permitindo oferecer, na maioria dos casos, um tratamento não mutilador para a paciente. Tumores não invasivos, conhecidos como *in situ*, apresentam

índice de cura próximo de 100%, enquanto que tumores invasivos, com diâmetro de até 2 centímetros, obtêm-se até 95% de cura (INCa, 2002).

Existem, também, muitos tipos de nódulos de mama que são de natureza benigna, sendo mais comuns os fibroadenomas e os cistos. O fibroadenoma surge quase sempre na mulher jovem, entre 15 e 30 anos, sendo um tumor sem potencial de tornar-se maligno. Certos tipos de câncer, chamados tumores circunscritos, podem ser interpretados como um fibroadenoma, sendo, portanto, prudente que os nódulos suspeitos sejam submetidos à confirmação histopatológica (INCa, 2002).

Os cistos são tumores de conteúdo líquido e pouca consistência, podendo atingir grandes volumes. Alguns tipos de cistos de grande volume podem exibir crescimento tumoral em seu interior, lembrando uma vegetação em desenvolvimento. Tal vegetação intracística merece investigação, pois pode representar tumor pré-maligno ou mesmo maligno (INCa, 2002).

O câncer de mama pode ser detectado precocemente por meio de três medidas básicas (INCa, 2002): auto-exame realizado mensalmente; exame clínico anual das mamas, realizado por médicos e enfermeiros, em todas as mulheres, especialmente naquelas com 40 anos ou mais de idade; e realização da mamografia que, idealmente, toda mulher com idade entre 50 e 69 anos deveria se submeter anualmente.

A mamografia é a radiografia da mama realizada com um aparelho de raios-X específico denominado mamógrafo. Durante o exame, adota-se como procedimento a compressão da mama com vistas a se obter imagem de melhor qualidade. Em imagens, assim obtidas, quando existir, é possível identificar tumores em fase inicial, de dimensão de milímetros, além de microcalcificações, que são consideradas como o sinal mais precoce de malignidade.

II.2 O Ultra-Som no Diagnóstico do Câncer de Mama

“O mamograma é denso, explica a Dra. Harttman. Em outras palavras, as radiografias são difíceis de serem analisadas. O que torna difícil para o cirurgião da equipe dizer até onde o tumor está se expandindo” (ROBERTS, 1999).

Dependendo de vários fatores, tais como o tamanho e a localização do tumor, a densidade do tecido mamário (mulheres jovens possuem mamas mais densas), qualidade dos recursos técnicos e a habilidade de interpretação do radiologista, a sensibilidade da mamografia pode variar de 46% a 88% e a especificidade, de 82% a 99% (INCa, 2005). Além disso, ensaios clínicos apontam que o uso da mamografia em programas de rastreamento propicia uma redução de 30% na mortalidade por câncer de mama em mulheres entre 50 e 69 anos (INCa, 2005).

Entretanto, existem situações onde a mamografia é incapaz de fornecer indícios suficientes para que o radiologista tome sua decisão diagnóstica. Nestes casos, em particular, o ultra-som (US) tem emergido como o mais importante coadjuvante da mamografia para pacientes com massas palpáveis e mamografias não-conclusivas (SKAANE, 1999). Segundo CHEN D. *et al.* (2000), o US tem apresentado uma efetividade similar à da mamografia, e até melhor quando o alvo são mulheres com idade inferior aos 35 anos.

Em geral, a acurácia, sensibilidade e especificidade do exame de US variam, respectivamente, de 81,3% a 96,6%, 75,3% a 85,1% e 78,5% a 96,8% (KOLB *et al.*, 2002; SKAANE, 1999). Além disso, diversos estudos clínicos indicam que o uso do US combinado a mamografia fornece uma melhora no desempenho diagnóstico, aumentando os valores de acurácia e sensibilidade (HOUSSAMI *et al.*, 2002, KOLB *et al.*, 2002, KAPLAN, 2001, SKAANE, 1999, ZONDERLAND *et al.*, 1999, RAHBAR *et al.*, 1999, DUIJM *et al.*, 1997, STAVROS *et al.*, 1995).

Os primeiros estudos do emprego do US, como método de imagem, para exame da mama, surgiram na década de 60. Na década seguinte, surgiram os primeiros artigos sugerindo que o US deveria ser usado para examinar pacientes sintomáticas com idade

inferior a 30 anos. Esta indicação baseava-se no fato de mulheres desta faixa etária apresentarem mamas densas, na necessidade de se evitar o uso de radiação em mulheres jovens, na capacidade do US permitir estabelecer se o tumor é cístico ou sólido e, no caso do tumor ser benigno, o US poderia ser empregado para monitorar a evolução (HAGEN-ANSERT, 1997). Inicialmente, esta técnica foi usada para diferenciar tumores císticos de sólidos em massas palpáveis não visualizadas nos mamogramas (BARR, 2001). Com o desenvolvimento de transdutores lineares na faixa de 7,5 – 10 MHz e uma melhor resolução no campo acústico próximo, a importância do US de mama expandiu-se para a análise das características morfológicas dos tumores (EVANS III *et al.*, 2002).

Do ponto de vista morfológico, segundo HAGEN-ANSERT (1997), os tumores benignos da mama em imagens por US geralmente apresentam contorno regular e bem definido, forma redonda ou oval, parede posterior bem definida e reforço acústico. Por outro lado, os tumores malignos se apresentam hipocóicos, com contorno irregular, ecos internos heterogêneos e sombra acústica.

Segundo RAHBAR *et al.* (1999), a análise de características de forma e contorno dos tumores sólidos de mama, em imagens por US, tem o potencial de auxiliar na redução do número de biópsias realizadas em tumores benignos. Esta evidência pode ser explicada pelo fato do contorno regular ser uma característica de elevada preditividade para benignidade (SKAANE, 1999; RAHBAR *et al.*, 1999; HUBER *et al.*, 2000). No estudo clínico realizado por RAHBAR *et al.* (1999), com 162 tumores de mama (38 malignos e 124 benignos) analisados independentemente por 3 radiologistas, 94% dos tumores de contorno regular eram benignos, sendo esta a característica mais preditiva de benignidade. De forma similar, mais de 85% dos 142 fibroadenomas (tumor benigno mais comum) avaliados por quatro radiologistas, no estudo de SKAANE (1999), possuíam contorno regular. Ainda, HUBER *et al.* (2000) mostraram que, em dois centros de radiológicos distintos, radiologistas diagnosticaram corretamente 76% e 88% dos tumores benignos levando em consideração apenas seus contornos regulares. Estes, associados à análise da textura do tecido adjacente ao tumor, resultaram numa maior contribuição diagnóstica tanto no centro 1 ($S = 83,0\%$ e $E = 92,0\%$), quanto no centro 2 ($S = 80\%$ e $E = 92\%$).

A literatura reporta diferenças no desempenho diagnóstico das características de textura e contorno. Como citado em diversos trabalhos (HUBER *et al.*, 2000; RAHBAR *et*

al., 1999; SKAANE, 1999), apesar da análise clínica da ecotextura interna dos tumores de mama ser, em alguns casos, útil no diagnóstico diferencial, esta não alcança valores de acurácia da ordem de grandeza dos obtidos com a análise das características morfológicas. Esta diferença pode ser explicada pela variabilidade existente na análise qualitativa da textura. RAHBAR *et al.* (1999) cita, por exemplo, que apesar da homogeneidade da ecotextura interna de tumores de mama ser uma característica com elevado potencial preditivo de benignidade, existe uma elevada variabilidade quanto ao diagnóstico fornecido por radiologistas com base nesta característica. Esta variabilidade diagnóstica, também foi observada por HUBER *et al.* (2000) em seu estudo clínico.

Assim, apesar dos vários aspectos relevantes do exame por US, como, por exemplo, não utilizar radiação ionizante, a portabilidade do equipamento, a facilidade de repetição do exame e a capacidade de gerar imagens em tempo real, este depende da qualidade da imagem gerada, podendo assim, resultar em elevada variabilidade na interpretação diagnóstica pelos radiologistas (BARR, 2001; HUBER *et al.*, 2000; RAHBAR *et al.*, 1999 e SKAANE, 1999).

II.3 Metodologias de Auxílio ao Diagnóstico

“Há quem afirme que os computadores nunca poderão exibir uma inteligência verdadeira – seja isso o que for. Mas parece-me que, se moléculas químicas muito complexas podem agir em seres humanos e torná-los inteligentes, circuitos eletrônicos complexos podem também fazer os computadores atuarem de forma inteligente” (HAWKING, 2001).

Com o intuito de aumentar a consistência da interpretação diagnóstica do US e reduzir o número de biópsias realizadas em tumores sólidos benignos, vários pesquisadores têm proposto metodologias distintas para auxiliar no diagnóstico (GOLDBERG *et al.*, 1992; GARRA *et al.*, 1993; ZHENG *et al.*, 1997; LEFEBVRE *et al.*, 2000; SIVARAMAKRISHNA *et al.*, 2002; HUBER *et al.*, 2000; CHEN, D., *et al.*, 2000; CHANG *et al.*, 2003, KUO *et al.*, 2002). Tais metodologias têm, em comum, basearem-se na segmentação da imagem e no cálculo de parâmetros que quantificam características

referentes ao contorno e a textura dos tumores, sendo, então, estes parâmetros utilizados na classificação e distinção entre tumores malignos de benignos.

A determinação das características de interesse da imagem e a extração de parâmetros que resultam em melhor desempenho na distinção entre tumores malignos e benignos, são etapas complexas que ainda continuam sob investigação. Duas abordagens têm sido usadas para quantificar as características morfológicas dos tumores de mama; uma delas baseia-se em parâmetros de textura, e a outra, em parâmetros morfométricos.

Os parâmetros de textura têm como base diferentes técnicas. Assim, GARRA *et al.* (1993), LEFEBVRE *et al.* (2000), HUBER *et al.* (2000) e SIVARAMAKRISHNA *et al.* (2002) utilizaram a matriz de co-ocorrência, enquanto que GOLDBERG *et al.* (1992) e ZHENG *et al.* (1997) basearam-se na determinação do *run-length* e da estatística do *Markoviano*. Por outro lado, CHEN, D. *et al.* (2000), KUO *et al.* (2002) e CHANG *et al.*, (2003) usaram os coeficientes de autocovariância, similarmente a HORSCH *et al.* (2002) e DRUKKER *et al.* (2002) que empregaram a autocorrelação. Mais recentemente a curva de complexidade (BAHEERATHAN *et al.*, 1999) foi usada por ALVARENGA *et al.* (2003b), e faz parte da contribuição deste trabalho.

Dentre os parâmetros morfométricos utilizados para quantificar a forma e o contorno dos tumores, deve-se destacar a razão entre a largura e a altura do tumor (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002; LEFEBVRE *et al.*, 2000; ZHENG *et al.* 1997), o gradiente radial normalizado (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002), o comprimento radial normalizado (CHOU *et al.*, 2001) e a razão de superposição (HORSCH *et al.*, 2001). Novamente, como contribuição deste trabalho, tem-se o valor médio quadrático residual normalizado, proposto por ALVARENGA *et al.* (2003a).

Detalhes sobre o cálculo e a aplicação dos parâmetros morfométricos e de textura são apresentados nos Capítulos V e VI, respectivamente.

III OPERADORES MORFOLÓGICOS

Este capítulo apresenta conceitos básicos sobre os Operadores Morfológicos para a compreensão do método de segmentação apresentado no Capítulo IV. As palavras em **negrito** indicam os nomes dos Operadores.

III.1 Operadores Básicos

Os Operadores Morfológicos (OM) são utilizados para identificar, remover ou realçar estruturas de interesse em uma imagem. Nas aplicações mais simples, utiliza-se apenas da imagem em estudo e de um elemento estruturante (ES). O ES é formado por um determinado conjunto de *pixels* e é usado para investigar a imagem durante a aplicação do Operador Morfológico. Caso o ES coincida com alguma estrutura da imagem, dependendo do operador aplicado, uma determinada transformação pode ser realizada. Logo, a forma e o tamanho do ES devem ser adequados às propriedades geométricas das estruturas em estudo (SOILLE, 1999). Assim, por exemplo, para o ES de tamanho 3 x 3 *pixels*, em forma de “L”, como ilustrado na Figura III.1a, tem-se os quadrados cinzas que representam *pixels* ativos (ou seja, que interagem com a imagem durante a aplicação do operador morfológico) e os quadrados brancos representam *pixels* inativos (o ponto preto indica o pixel central do ES, considerado referência para a operação, e o qual assume o valor resultante da aplicação do operador). Para este exemplo, o ES transposto é mostrado na Figura III 1b.

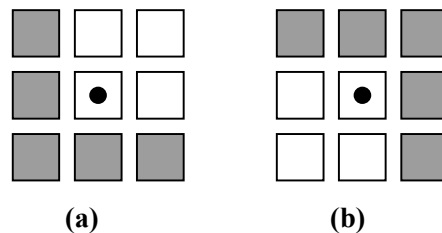


Figura III.1. Exemplo de (a) elemento estruturante B de tamanho 3 x 3 *pixels* em forma de “L”, com centro definido pelo ponto preto, e (b) seu transposto B^T .

Os operadores utilizados no processamento morfológico de imagens formam uma ampla classe de operadores não-lineares, sendo, entretanto, todos construídos pela interação de duas operações básicas: **erosão** e **dilatação**. A **erosão** de uma imagem f para um dado *pixel* x é definida como o valor mínimo da imagem em uma janela formada pelo elemento

estruturante B , estando a origem deste ES na posição x (SOILLE, 1999), ou seja, a erosão pode ser expressa por:

$$[\varepsilon_B(f)](x) = \min_{b \in B} f(x + b), \quad (\text{III.1})$$

onde b é um *pixel* pertencente ao ES B . Por outro lado, a **dilatação** pode ser definida como:

$$[\delta_B(f)](x) = \max_{b \in B} f(x + b) \quad (\text{III.2})$$

A **erosão** de uma imagem binária resulta na remoção de estruturas com dimensão menor que ES e na redução do tamanho de estruturas maiores que ES (Figura III.2b), enquanto que a **dilatação** aumenta o tamanho de todas as estruturas da imagem (Figura III.2c), unindo estruturas que estejam separadas por uma distância menor do que ES.

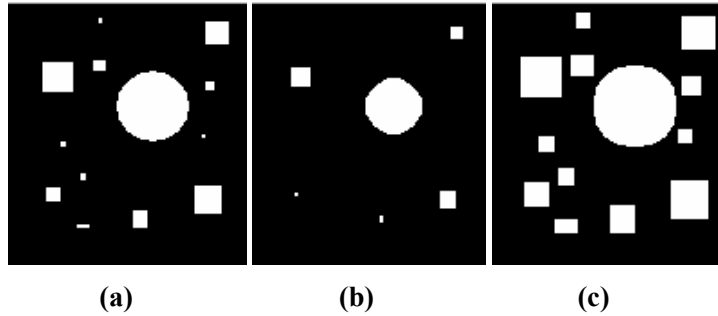


Figura III.2. (a) Imagem (160 x 160 *pixels*) a ser processada f . (b) **Erosão** e (c) **dilatação** usando um ES quadrado com 5 *pixels* de lado.

A aplicação de uma destas operações, **dilatação** e **erosão**, seguida da outra, forma operadores distintos: **abertura** e **fechamento**. A **abertura** morfológica da imagem f é definida como a **erosão** desta imagem com ES B , seguida da **dilatação** com o ES transposto (SOILLE, 1999), sendo expressa por:

$$\gamma_B(f) = \delta_{B^T}[\varepsilon_B(f)], \quad (\text{III.3})$$

A **abertura** tem o efeito de remover estruturas menores que o ES sem, entretanto, alterar as estruturas maiores (Figura III.3b).

Por outro lado, o **fechamento** morfológico é definido pela **dilatação** da imagem f , seguida da **erosão** com o ES transposto (SOILLE, 1999):

$$\phi_B(f) = \varepsilon_{B^T}[\delta_B(f)], \quad (\text{III.4})$$

pode preencher lacunas em estruturas da imagem f (Figura III.3c).

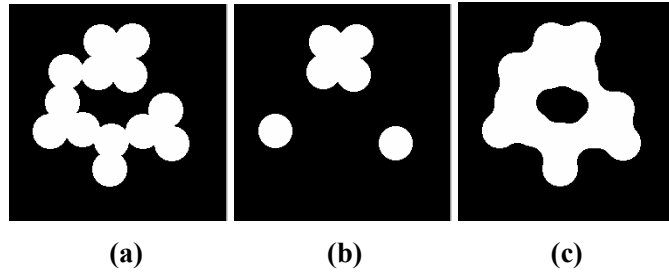


Figura III.3. (a) Imagem (256 x 256 *pixels*) a ser processada f . (b) **Abertura** e (c) **fechamento** usando um ES circular de raio 19 *pixels*.

III.2 Transformações Geodésicas e Operadores Baseados em Reconstrução

Diferentemente das operações mais simples, as transformações geodésicas necessitam do uso de uma outra imagem, além daquela que é operada (f) e do ES, para promover suas operações. Esta outra imagem, denominada de máscara (g), serve de domínio para o resultado das operações realizadas em f (marca). A marca f pode ser uma imagem criada para atender uma determinada condição quando se aplica à operação desejada, ou pode ser obtida por meio de uma operação morfológica na própria máscara g .

Nesta classe de operadores, as duas operações básicas são a **erosão geodésica** e a **dilatação geodésica**. A **dilatação geodésica** de tamanho 1 da marca f com respeito à máscara g é pode ser expressa por (SOILLE, 1999):

$$\delta_g^{(1)}(f) = \min[g; \delta_B(f)]. \quad (\text{III.5})$$

Por outro lado, a **erosão geodésica** de tamanho 1 da marca f com respeito à máscara g pode ser definida como (SOILLE, 1999):

$$\varepsilon_g^{(1)}(f) = \max[g; \varepsilon_B(f)]. \quad (\text{III.6})$$

A **erosão geodésica** e a **dilatação geodésica** sempre convergem após um número finito de aplicações sucessivas. A convergência ocorre quando a operação não causa mais nenhuma alteração na imagem resultante, sendo os operadores baseados em reconstrução definidos com base nesta propriedade. Assim, a **reconstrução por dilatação** (Figura III.4) de uma imagem máscara g a partir de uma imagem marca f é definida como a **dilatação geodésica** de f com respeito à g até a convergência, ou seja (SOILLE, 1999):

$$R_g(f) = \delta_g^{(i)}(f), \quad (\text{III.7})$$

onde i é tal que $\delta_g^{(i)}(f) = \delta_g^{(i+1)}(f)$. A Figura III.4 ilustra um exemplo da aplicação da **reconstrução por dilatação** para separar um par de letras do restante da imagem.

Similarmente, a **reconstrução por erosão** de uma imagem máscara g a partir de uma imagem marca f é definida como a **erosão geodésica** de f com respeito à g até a convergência e pode ser representada por (SOILLE, 1999):

$$R_g^*(f) = \varepsilon_g^{(i)}(f), \quad (\text{III.8})$$

onde i é tal que $\varepsilon_g^{(i)}(f) = \varepsilon_g^{(i+1)}(f)$.

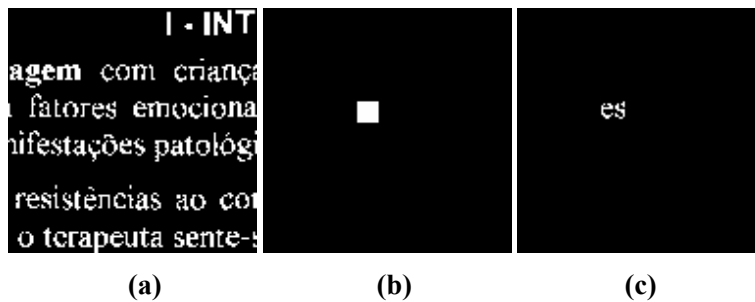


Figura III.4. (a) Imagem original g usada como máscara. (b) Imagem f usada como marca. (c) Resultado da **reconstrução por dilatação** da máscara g a partir da marca f .

III.3 Abertura e fechamento por reconstrução

A **abertura por reconstrução** é definida como a **reconstrução por dilatação** de f a partir da **erosão** de f (SOILLE, 1999):

$$\gamma_R(f) = R_f[\varepsilon_B(f)], \quad (\text{III.9})$$

enquanto que o **fechamento por reconstrução** é a **reconstrução por erosão** de f a partir da **dilatação** de f (SOILLE, 1999).

$$\phi_R(f) = R_f^*[\delta_B(f)] \quad (\text{III.10})$$

Assim, os operadores **abertura/fechamento por reconstrução** preservam a forma dos componentes que não são removidos pela **erosão/dilatação**, ou seja, todas as estruturas da imagem que não contêm o ES, permanecendo as demais inalteradas (SOILLE, 1999). Na Figura III.5 tem-se um exemplo de **abertura por reconstrução** para remover as bolas brancas menores que o ES.

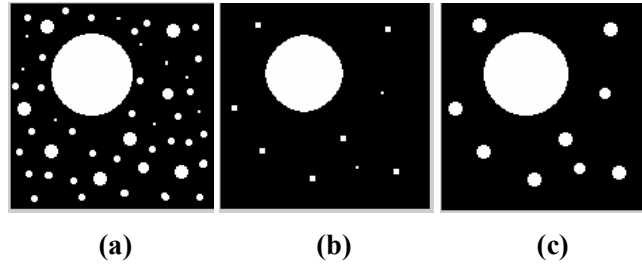


Figura III.5. (a) Imagem original após (b) uma **erosão** usando um ES quadrado (3 x 3 *pixels*). (c) **Dilatação por reconstrução** até que coincida com a original (mesmo ES).

III.4 Imposição de mínimos

Este operador força a existência de regiões de mínimo na imagem. Para tal, é necessária a definição de uma imagem marca determinando as regiões nas quais serão impostos os mínimos. A marca f_m deve ser definida como (SOILLE, 1999):

$$f_m(x) = \begin{cases} 0, & \text{se } x \text{ pertence a marca} \\ t_{\max}, & \text{nos demais casos} \end{cases} \quad (\text{III.11})$$

onde t_{\max} é o máximo valor da imagem f .

A **imposição de mínimos** (Eq. III.11) é obtida a partir da realização de duas operações: o mínimo entre a marca f_m e a imagem f , representado por $f \wedge f_m$ seguido da **reconstrução por erosão** de f_m com respeito à $f \wedge f_m$, conforme exemplificado na Figura III.6.

$$R_{[f \wedge f_m]}^*[f_m] \quad (\text{III.12})$$

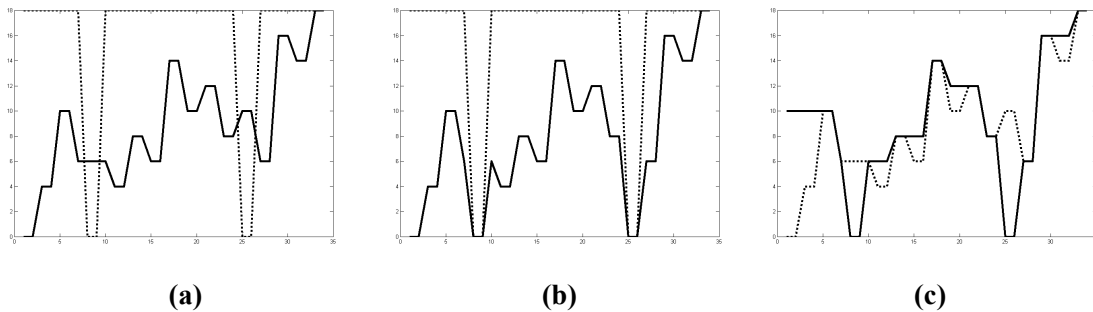


Figura III.6. Exemplo da aplicação da **imposição de mínimos**. (a) Curva f (sólida) e a marca f_m (pontilhada). (b) Mínimo entre as curvas f e f_m . (c) Resultado da **imposição de mínimos** (sólida) sobre a curva f (pontilhada).

III.5 Watershed

O operador *watershed* é utilizado com frequência no auxílio à segmentação de objetos em imagens, por meio da determinação das fronteiras entre os mesmos. Estas fronteiras são representadas por linhas que indicam os pontos de encontro entre regiões que contém mínimos locais. O procedimento que determina as linhas de *watershed* inicia-se com a localização dos mínimos locais da imagem (Figura III.7a), os quais são associados a um único rótulo (Figura III.7b). Em seguida, todos os *pixels* com valores de nível de cinza maiores que os mínimos rotulados são identificados e os *pixels* adjacentes ao mínimo local (Figura III.8a) são rotulados (Figura III.8b) como pertencendo ao vale correspondente ao respectivo mínimo. Aqueles *pixels* que não são adjacentes são colocados numa fila e recebem um rótulo apropriado no instante em que seus vizinhos são assinalados como pertencentes a um vale. Este procedimento é repetido sucessivamente (como se um nível d'água fosse progressivamente subindo), para diferentes níveis de cinza até alcançar o valor máximo da imagem. As linhas divisórias são construídas nos pontos onde as regiões de mínimos locais se encontram (Figura III.9), formando os *watersheds* da imagem (SOILLE, 1999).

160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	120	160	160	160	160	80	160	160	3	160	160	160	160	2	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	160	160	160	20	160	160	160	160	160	160	160	1	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160

(a)

(b)

Figura III.7. (a) Imagem original contendo três mínimos locais (em cinza) respectivamente (b) rotulados com os valores 1, 2 e 3.

160	160	160	160	160	160	160	160	160	3	160	160	160	160	2	160
160	3	160	160	160	160	2	160	3	3	3	160	160	2	2	2
160	160	160	160	160	160	160	160	160	3	160	160	160	160	2	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	1	160	160	160
160	160	160	160	1	160	160	160	160	160	160	1	1	1	160	160
160	160	160	160	160	160	160	160	160	160	160	160	1	160	160	160
160	160	160	160	160	160	160	160	160	160	160	160	160	160	160	160

(a)

(b)

Figura III.8. (a) Análise dos *pixels* da vizinhança definida pelo elemento estruturante (em cinza escuro) e (b) rotulados como pertencendo ao vale definido pelo respectivo mínimo local.

3	3	3	3	0	2	2	2
3	3	3	3	0	2	2	2
3	3	3	3	0	2	2	2
3	3	3	0	1	0	2	2
3	3	0	1	1	1	0	0
3	0	1	1	1	1	1	1
0	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1

Figura III.9. Resultado do processo de análise da vizinhança obtendo as linhas divisórias (em cinza escuro) das regiões de mínimos locais.

IV SEGMENTAÇÃO DE TUMORES DE MAMA EM IMAGENS POR US USANDO MORFOLOGIA MATEMÁTICA

“Acompanhados por uma enfermeira e um residente, três especialistas traçam meticulosamente formas imaginárias nas radiografias, tentando distinguir tecidos anormais dos tecidos saudáveis” (ROBERTS, 1999).

A segmentação é uma etapa indispensável na análise quantitativa de imagens, pois a partir da imagem segmentada pode-se quantificar as características dos objetos de interesse. Duas abordagens clássicas têm sido utilizadas na segmentação com base na detecção de regiões ou na detecção de fronteiras, denominadas respectivamente, *region-based* e *edge-based* (CASTLEMAN, 1996). Particularmente em imagens médicas, devido à natureza complexa dos tecidos biológicos, técnicas híbridas têm sido aplicadas, como, por exemplo, as de segmentação de imagens de US de mama (CHEN *et al.*, 2000; HORSCH *et al.*, 2001; MADABHUSHI *et al.*, 2003; HUANG *et al.*, 2004).

CHEN *et al.* (2000) adotaram abordagem baseada em textura e empregaram um Modelo de Contornos Ativos (MCA) para delinear as bordas dos tumores. Embora tenha sido utilizado com sucesso na segmentação de imagens médicas (PHAM *et al.*, 2000), esse método é muito sensível ao ruído e, portanto, requer a intervenção de especialista para traçar o contorno inicial do modelo que englobe a estrutura de interesse. A MCA também foi usada por MADABHUSHI *et al.* (2003), porém combinando informação de intensidade e de textura dos tumores e utilizando o método de crescimento de regiões para fornecer uma “inicialização” para a modelagem. Os autores reportam que essa técnica resultou em elevada rejeição a artefatos de sombra acústica e a falsos positivo, os quais provêm da gordura subcutânea e de tecidos glandulares.

Por outro lado, o método de segmentação proposto por HORSCH *et al.* (2001) baseia-se na maximização de uma função radial derivativa sobre contornos definidos *a priori* para diferentes níveis de cinza. Entretanto, esta técnica resultou em fraco desempenho em detectar detalhes nos contornos, tais como espiculações e pequenas irregularidades. Desempenho superior foi obtido por HUANG *et al.* (2004) ao utilizar

Redes Neurais Artificiais auto-organizáveis (aprendizado não-supervisionado) para separar os *pixels* imagens em classes distintas e o *watershed* para detectar os contornos.

No presente trabalho, método de segmentação distinto dos anteriormente citados é proposto para aplicação a imagens de US de mama. Esse método consiste em, inicialmente, aplicar o procedimento de *region-based* para agrupar os *pixels* do interior de um tumor ao redor de mínimos regionais, que foram determinados usando-se o operador de **imposição de mínimos**. A seguir, utiliza-se o *edge-based* para identificar as fronteiras entre grupos de *pixels* adjacentes, fazendo uso do operador morfológico *watershed* aplicado ao histograma da imagem obtida na etapa anterior. Este método foi aplicado a imagens reais e simuladas.

IV.1 Banco de Imagens US de Mama

As imagens de US de tumores de mama usadas neste estudo, cedidas pelo INCa, foram adquiridas por meio de um equipamento *Sonoline-Sienna* de 7,5 MHz e gravadas em formato TIF, com vistas a facilitar a constituição de um banco de dados de imagens. Este banco de dados tem 272 imagens de US de tumores de mama de 210 pacientes previamente submetidas a mamografia. Os resultados histológicos (203 malignas e 69 benignas) foram obtidos via *intranet* do INCa, a partir do número de matrícula de cada paciente. Cada imagem foi exibida no monitor de vídeo de um microcomputador (Figura IV.1), tendo então um radiologista experiente delimitado manualmente, com auxílio do *mouse*, as regiões de interesse (ROI) da imagem. Este radiologista estabeleceu ainda o contorno dos tumores e a classificação da ecotextura da imagem. Quanto ao contorno foram estabelecidas 191 irregulares e 81 regulares (Figura IV.2), enquanto que, no que concerne a ecotextura, 130 foram classificadas como heterogênea e 142 como homogêneas (Figura IV.3).

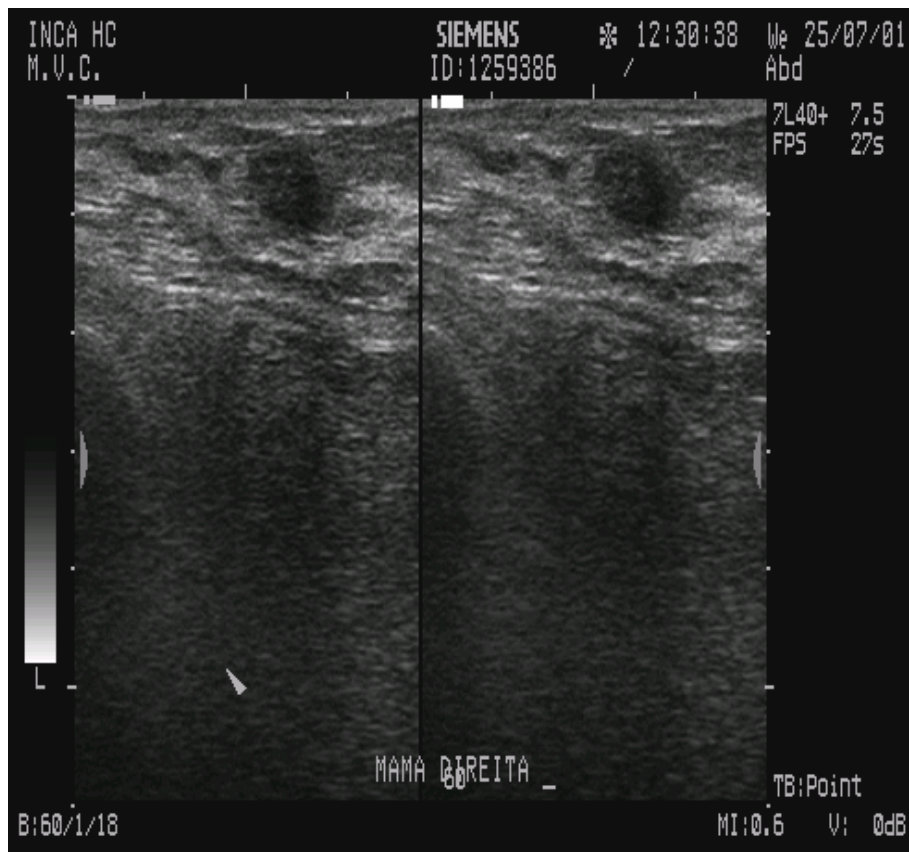
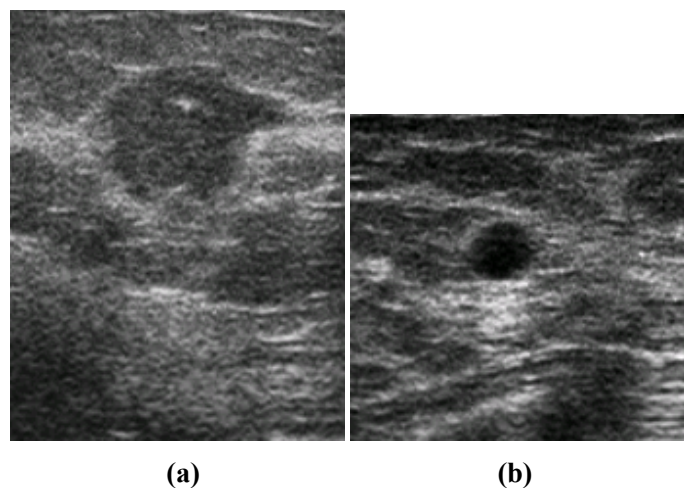


Figura IV.1. Exemplo de imagem original de um tumor de mama adquirida com o equipamento de US de 7,5 MHz (*Sonoline – Sienna ©Siemens*) e gravada no formato TIF.



(a)

(b)

Figura IV.2. Exemplos de ROI contendo tumores de contorno (a) irregular e (b) regular, determinados pelo radiologista.

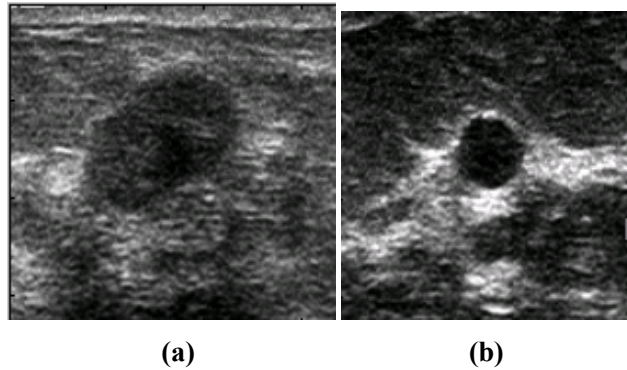


Figura IV.3. Exemplos de ROI contendo de tumores de ecotextura (a) heterogênea e (b) homogênea classificados pelo radiologista.

Com vistas a se ter um grupo controle, foram incluídos no banco de imagens, pacientes saudáveis, ou seja, 74 imagens de US de regiões contendo, predominantemente, tecido glandular (37) ou adiposo (37). Tais imagens foram utilizadas na fase inicial deste trabalho, mais especificamente na avaliação do desempenho dos parâmetros de textura em separar estes dois tipos de regiões (Figura IV.4). O parênquima mamário é sonograficamente variável e depende da quantidade de tecido adiposo, fibroso e glandular presente na mama. O tecido adiposo, na mama, tende a ser hipoecóico e homogêneo, enquanto que o tecido glandular apresenta uma ecogenicidade mista, tendendo à heterogeneidade (SKAANE, 1999). Estas características seriam adequadas a um primeiro teste para os parâmetros de textura.

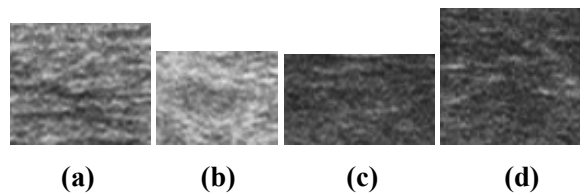


Figura IV.4. Exemplo de regiões saudáveis com predominância de (a) e (b) tecido glandular e (c) e (d) tecido adiposo, selecionadas pelo radiologista.

IV.2 Método de Segmentação: SAC (*Semi-Automatic Contour*)

Inicialmente, a imagem original (Figura IV.5a) é normalizada pelo seu máximo, e então se aplica o operador de **abertura por reconstrução** através de um ES de 3×3 pixels. Tal procedimento aumenta o contraste entre regiões distintas, evidenciando as diferenças (Figura IV.5b). O histograma da imagem reconstruída (Figura IV.6a) é filtrado, aplicando-

se o operador de **abertura morfológica** com um ES linear de 3 *pixels*, obtendo-se o histograma da Figura IV.6b. Em seguida, aplica-se o operador morfológico *watershed* (ES linha de 3 *pixels*) para rotular cada região modal, identificando assim seus *pixels* (Figura IV.6b). Os rótulos numéricos das regiões modais são ordenados de forma crescente (Figura IV.6c), obtendo-se uma nova imagem caracterizada por um número menor de níveis de cinza (Figura IV.5c).

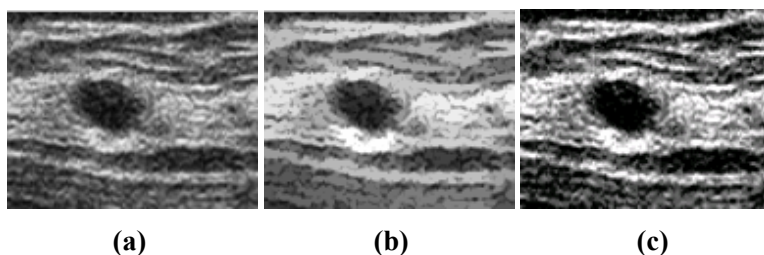
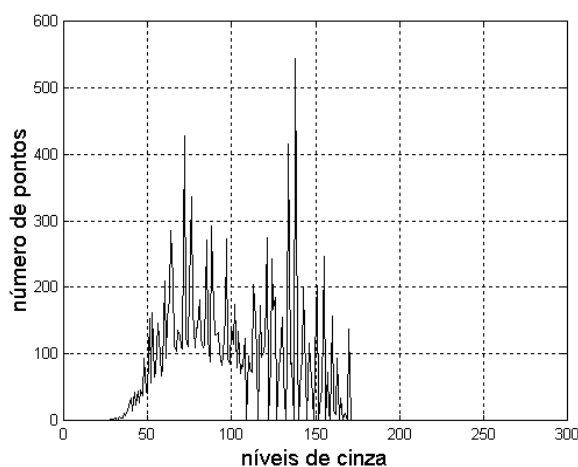
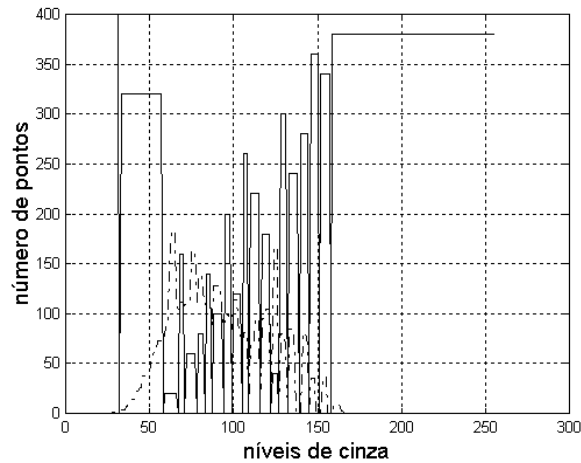


Figura IV.5. (a) Imagem original de um tumor de mama. (b) Após a aplicação da **abertura por reconstrução** para evidenciar as diferenças. (c) Imagem gerada a partir dos *pixels* rotulados e ordenados. Esta contém um número menor de níveis de cinza que a imagem original (a).

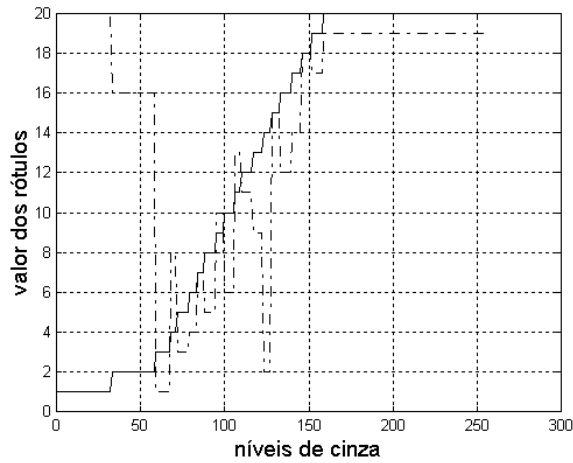
O operador de **imposição de mínimos** é aplicado para garantir que os *pixels* no interior do tumor possuam valores inferiores ao restante da imagem, sendo necessário à seleção manual de um deles, com auxílio do radiologista, para definição da marca. A imagem resultante (Figura IV.7a) mostra regiões bem definidas, com histograma (Figura IV.7b) caracterizado por picos facilmente identificados, sendo os primeiros picos (de menor amplitude) pertencentes a ROI.



(a)



(b)



(c)

Figura IV.6. (a) Histograma da imagem (IV.5b). (b) Histograma (a) filtrado por uma **abertura** morfológica (linha tracejada) e rótulos numéricos (linha contínua) gerados pelo *watershed*. (c) Rótulos numéricos originais (curva tracejada) e ordenados (curva contínua).

Para realizar a segmentação da ROI, escolhe-se o limiar em nível de cinza como o máximo gradiente da curva da área cumulativa (Figura IV.7d), determinada a partir do histograma da Figura IV.7b. Este critério baseia-se na suposição de que este máximo representa a transição entre a área interna do tumor e o restante da imagem. Esta característica pode ser observada no histograma da Figura IV.7b, onde os pequenos picos presentes entre 0 e 125, representam os *pixels* do interior do tumor. A segmentação resultante pode possuir algumas falhas provenientes da criação dos rótulos (Figura IV.8b).

Em sendo este o caso, as falhas são corrigidas realizando-se o **fechamento** morfológico com um ES circular 3 x 3, resultando, por exemplo, na imagem da Figura IV.8c. O método de segmentação foi implementado em MATLAB[®] (Mathworks Inc., Natick, MA, USA) usando a *toolbox* SDC Morphology Toolbox V0.14 (SDC Information Systems, Naperville, USA).

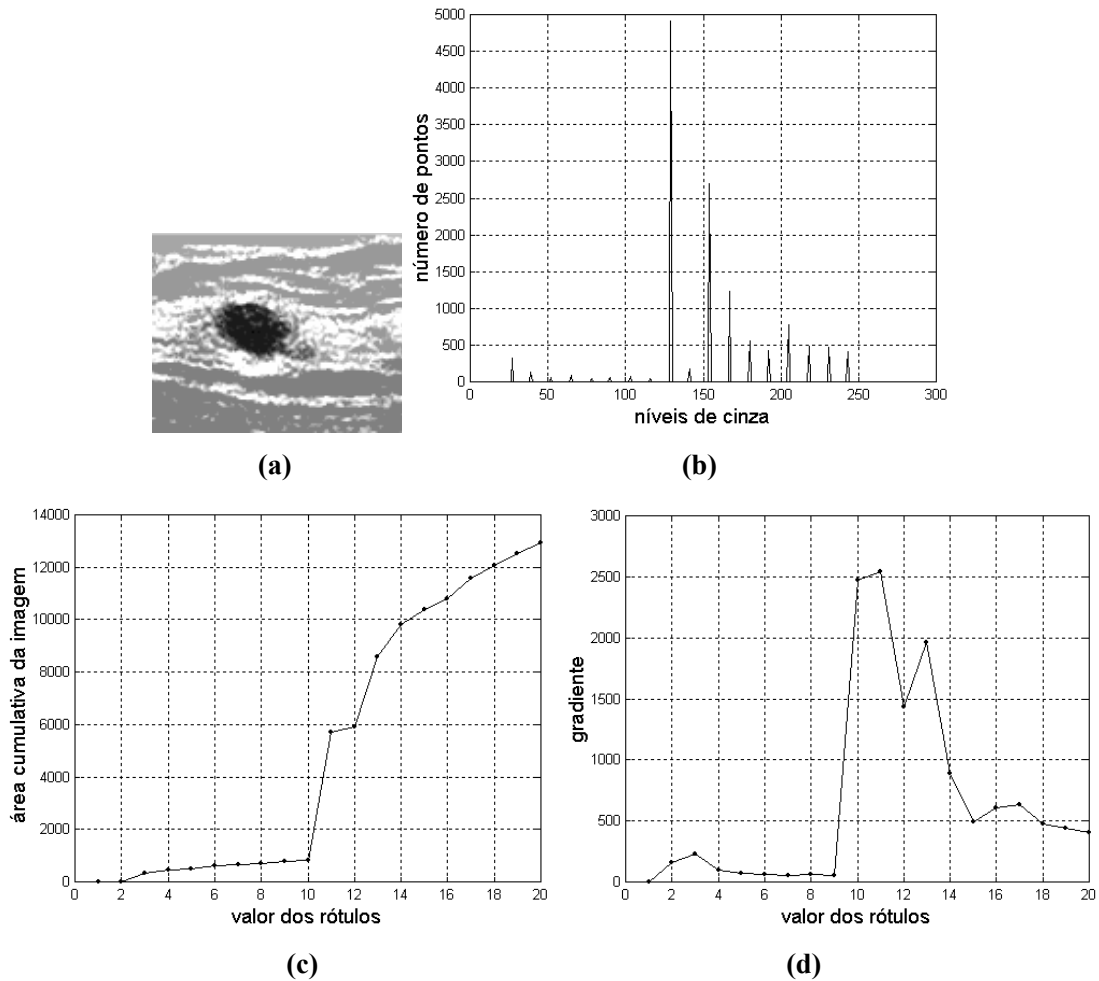


Figura IV.7. (a) Imagem (IV.1c) após aplicação da **imposição de mínimos** e (b) respectivo histograma, onde cada pico representa uma região rotulada. (c) Curva da área cumulativa do histograma (IV.3b) e seu (d) respectivo gradiente, onde o pico representa a transição entre a área da ROI e o restante a imagem.

IV.2.1 Detecção do Contorno

O contorno do tumor é determinado através do *Laplaciano* na imagem segmentada, o que permite determinar suas fronteiras abruptas, ou seja:

$$\nabla^2 f(x, y) = \frac{\partial^2}{\partial x^2} f(x, y) + \frac{\partial^2}{\partial y^2} f(x, y), \quad (\text{IV.1})$$

sendo $f(x,y)$ uma função bidimensional. No caso da imagem binária obtida da segmentação, a fronteira determinada pelo *Laplaciano* é o contorno da região segmentada. A implementação digital da Eq. (IV.1) foi realizada como (CASTLEMAN, 1996):

$$\nabla^2 f(x, y) = A * f(x, y), \quad (\text{IV.2})$$

onde $*$ é a convolução da imagem com a matriz A .

$$A = \begin{bmatrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (\text{IV.3})$$

Esta matriz, clássica no processamento de imagem, é construída de modo que a convolução seja nula quando todos os valores da imagem sob A forem iguais, ou resulte em valores elevados quando houver transições bruscas entre os valores da imagem sob A . Logo, este operador é adequado para imagens binárias. O resultado final da aplicação do método de segmentação proposto é ilustrado na Figura IV.8d.

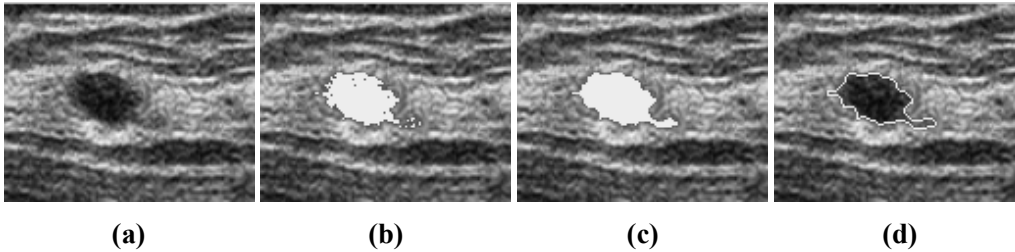
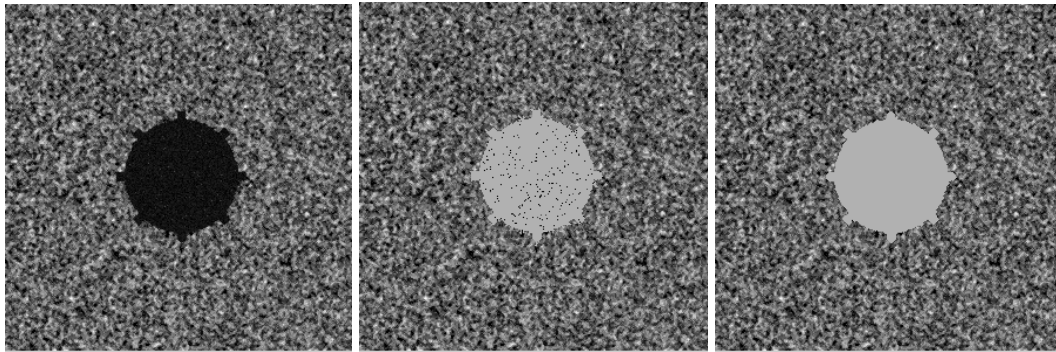


Figura IV.8. (a) Imagem original. (b) Após a aplicação da segmentação. (c) Após aplicação do operador de **fechamento por reconstrução** na imagem (b). (d) Contorno determinado após aplicação do *Laplaciano*.

IV.3 Segmentação de Imagens Simuladas

Para uma imagem simulada contendo figuras geométricas conhecidas (quadrado - 60x60 *pixels*, círculo - 40 *pixels* de diâmetro e rodas dentadas 8 e 16 dentes), a aplicação do SAC resultou em erro sempre inferior a 1% da área original (quadrado: +0,86%; círculo: +0,88%; roda de 8 dentes: 0,95%; roda de 16 dentes: +0,96%), conforme descrito nas publicações decorrentes deste trabalho (ALVARENGA *et al.*, 2002a, 2003). Exemplo da segmentação para a roda de 8 dentes é mostrado na Figura IV.9.



(a)

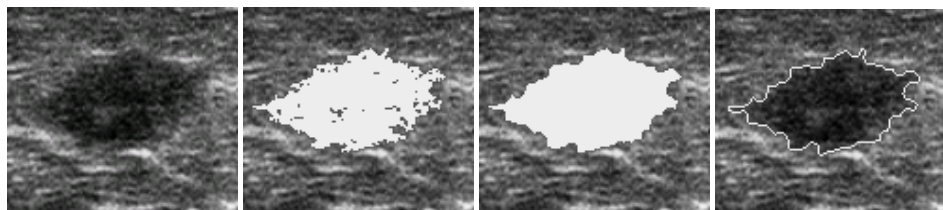
(b)

(c)

Figura IV.9. (a) Imagem original. (b) Após a aplicação da segmentação. (c) Após aplicação do operador de **fechamento por reconstrução** na imagem (b).

IV.4 Segmentação de Imagens de US de Mama usando SAC

A segmentação de uma imagem contendo tumor de mama de contorno irregular e ecotextura heterogênea (Figura IV.10a) resultou na área representada em cinza na Figura IV.10b. Nesta, nota-se a existência de pequenas falhas em seu interior, que foram corrigidas com o **fechamento por reconstrução**, resultando na Figura IV.10c. O resultado da detecção do contorno é ilustrado na Figura IV.10d. De maneira similar, a segmentação de outro tumor, também irregular (Figura IV.11a), resultou em área (cinza) com um número maior de falhas (Figura IV.11b). Estas, que aparentemente se devem a maior heterogeneidade presente no interior tumor, foram corrigidas com o **fechamento por reconstrução** (Figura IV.11c). O contorno resultante é ilustrado na Figura IV.11d.



(a)

(b)

(c)

(d)

Figura IV.10. (a) ROI de imagem de US de mama. (b) Após a realização da segmentação. (c) Após aplicação do operador de **fechamento por reconstrução** na imagem (b). (d) Contorno determinado após aplicação do *Laplaciano*.

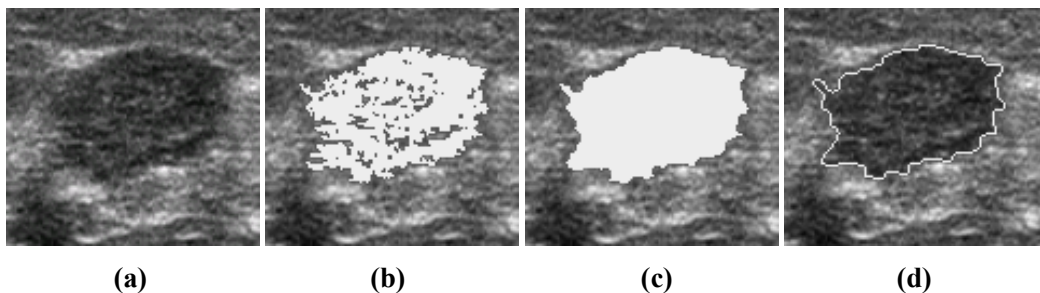


Figura IV.11. **(a)** ROI de imagem de US de mama. **(b)** Após a realização da segmentação. **(c)** Após aplicação do operador de **fechamento por reconstrução**. **(d)** Contorno determinado após aplicação do *Laplaciano*.

Aplicando-se o SAC aos 272 tumores do banco de imagens (seção IV.1), o mesmo foi capaz de segmentar 246 deles (Tabela IV.1), sendo todos malignos os 26 não segmentados. Características típicas deste tipo tumor, como ecotextura interna mista e altamente heterogênea (Figura IV.12), ou similar a do tecido adjacente (Figura IV.13), dificultaram a determinação do contorno. Assim, o SAC tendeu a falhar em casos onde não havia dúvidas quanto ao diagnóstico do tumor.

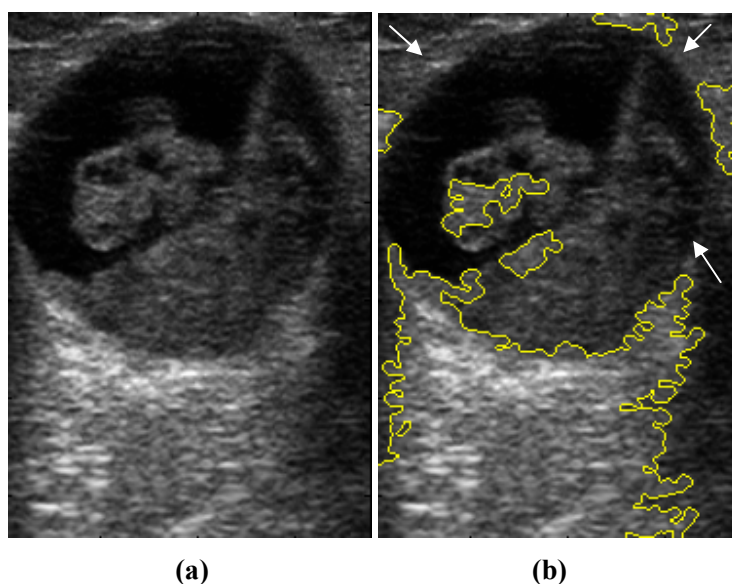


Figura IV.12. **(a)** ROI contendo tumor de ecotextura mista e elevada heterogeneidade e **(b)** respectivo contorno resultante da aplicação do SAC. Nota-se que o SAC não foi capaz de determinar completamente os limites entre o tumor e o tecido adjacente, principalmente na sua porção superior (seta).

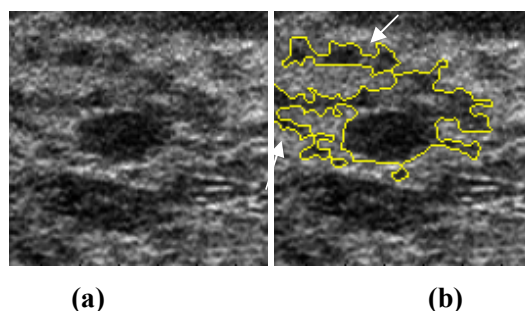


Figura IV.13. (a) ROI contendo tumor de ecotextura interna similar a do tecido adjacente. Nota-se que (b) seu respectivo contorno, resultante da aplicação do SAC, excedeu os limites do tumor (setas).

Com o objetivo de avaliar os resultados obtidos com a segmentação resultante da aplicação do SAC, os contornos foram comparados àqueles definidos, manualmente, por dois radiologistas. Como exemplificado na Figura IV.14, aplicando-se o SAC em tumor de contorno regular, hipocóico e homogêneo (Figura IV.14a), resultou em mais irregularidade (Figura IV.14d) do que as presentes nos contornos definidos por dois radiologistas experientes (Figura IV.14b e 14c). Da mesma maneira, o resultado da aplicação do SAC (Figura IV.15d) em tumor de contorno irregular e ecotextura heterogênea (Figura IV.15a) foi mais irregular que os contornos delineados por radiologistas experientes (Figura IV.15b e 15c).

Apesar destas diferenças, dois testes *t* pareados ($\alpha = 5\%$, $p = 0,06$ e $p = 0,13$), usando a razão de superposição (*RS*) (Eq.V.14) como medida de similaridade, indicaram não existir diferença estatisticamente significativa entre os contornos resultantes da aplicação do SAC e os definidos pelos radiologistas, conforme reportado por ALVARENGA *et al.* (2002b), em publicação relacionada à Tese. Este resultado sugere que, apesar do contorno resultante da aplicação do SAC tender a representar mais as irregularidades do que os definidos pelos radiologistas a forma geral do mesmo não é comprometida. Maiores detalhes podem ser encontrados em ALVARENGA *et al.* (2003a), publicação relacionada à Tese.

Conclui-se, então, que o método de segmentação proposto foi capaz de segmentar mais de 90% das imagens, mesmo com um banco de imagens heterogêneo, sugerindo, então, a potencialidade dos métodos baseados em Morfologia Matemática na segmentação de imagens médicas por US.

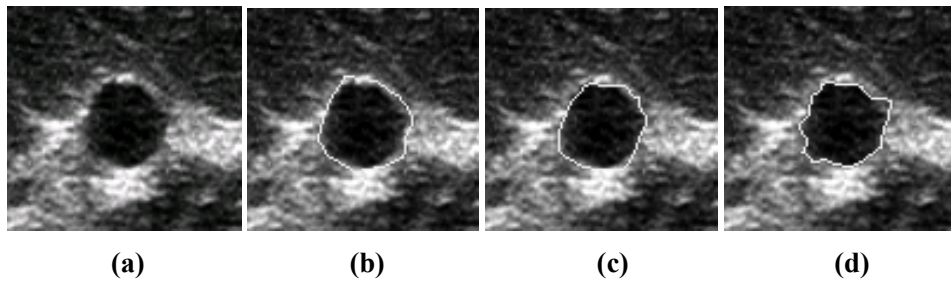


Figura IV.14. (a) ROI de imagem de tumor de mama de contorno regular. Contornos estabelecidos pelo (b) 1^o radiologista, (c) 2^o radiologista e (d) SAC.

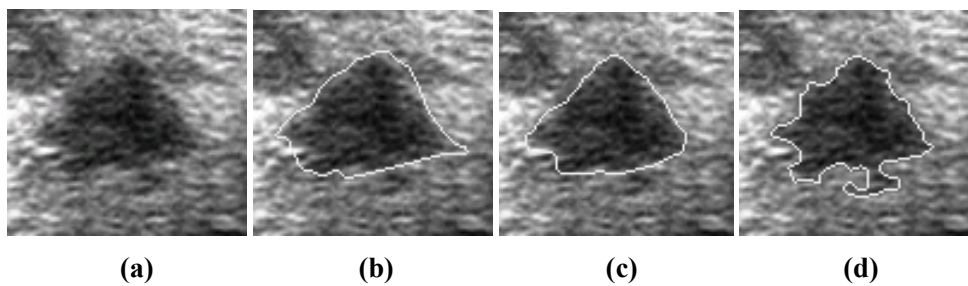


Figura IV.15. (a) ROI de imagem de tumor de mama de contorno irregular. Contornos estabelecidos pelo (b) 1^o radiologista, (c) 2^o radiologista e (d) SAC.

Tabela IV.1. Total de tumores segmentados com o SAC

Tipos de tumores	Total	Segmentados
Benignos	69	69 (100%)
Malignos	203	177 (87,2%)
Total	272	246 (90,4%)

V PARÂMETROS MORFOMÉTRICOS NA DISTINÇÃO DE TUMORES: CONTORNO E DIAGNÓSTICO

Com descrito no Capítulo II, a forma e o contorno de um tumor são características utilizadas pelos radiologistas para iniciar uma hipótese diagnóstica. Tumores benignos apresentam um contorno regular e bem definido, além de uma forma redonda ou oval. Entretanto, os malignos tendem a possuir um contorno irregular e algumas vezes indefinido. Estas são características que podem ser quantificadas e auxiliar o radiologista a definir quão irregular é um contorno. Assim, pesquisadores têm investigado diversos parâmetros morfométricos com vistas a estabelecer o desempenho em distinguir entre tumores benignos ou malignos.

As principais técnicas para se calcular os parâmetros morfométricos serão apresentadas neste Capítulo. Uma breve revisão da literatura aponta a potencialidade da aplicação deste tipo de parâmetro às imagens de mama. Tais parâmetros serão então aplicados às imagens US de mama (seção IV.1), e avaliados quanto ao seu desempenho em distinguir tumores, quanto ao contorno, e sua contribuição ao diagnóstico. Os resultados do uso dos parâmetros morfométricos serão comparados com os da literatura.

V.1 Técnicas para Quantificar as Características Morfométricas de Imagens

“... compare a definição de um círculo por Euclides com a de Descartes:”
(MLODINOW, 2004).

“Euclides - Um círculo é uma figura plana contida por uma linha [isto é, uma curva] tal que todas as linhas retas que vão até ela de um certo ponto de dentro do círculo – chamado de centro – são iguais entre si” (MLODINOW, 2004).

“Descartes - Um círculo é todo x e y que satisfaça a $x^2 + y^2 = r^2$ para algum número constante r ” (MLODINOW, 2004).

“Mesmo para aqueles que não sabem o que significa a equação, a definição de Descartes tem de parecer mais simples. A questão não é a interpretação da equação, mas meramente que, no método de Descartes, o círculo é definido por uma equação. Descartes traduziu o espaço em números e, mais importante ainda, usou sua tradução para descrever a geometria em termos de álgebra”
(MLODINOW, 2004).

A partir de uma imagem segmentada, podem-se analisar os objetos presentes na mesma, com vistas a se determinar forma, avaliar contorno e realizar medições. Para este objetivo ser alcançado, inúmeras técnicas têm sido propostas, porém, considerando o contexto deste trabalho, serão apresentadas apenas aquelas que foram aplicadas a imagens de mama.

V.1.1 Comprimento Radial Normalizado (Normalized Radial Length)

Usado para avaliar o contorno, o comprimento radial normalizado é uma função (Figura V.1b) que descreve a distância euclidiana do centróide do objeto aos *pixels* que pertencem ao seu contorno, e pode ser definido como (CHOU *et al.*, 2001):

$$d(p) = \frac{\sqrt{(x_p - X_0)^2 + (y_p - Y_0)^2}}{\max(d(p))}, p = 1, 2, \dots, P, \quad (\text{V.1})$$

onde (X_0, Y_0) são coordenadas do centróide do objeto, (x_p, y_p) , as coordenadas do p -ésimo *pixel* que pertence ao contorno, P é o número total de *pixels* que formam o contorno (perímetro) e $\max(d(p))$ é o valor máximo do comprimento radial calculado (Figura V.1a).

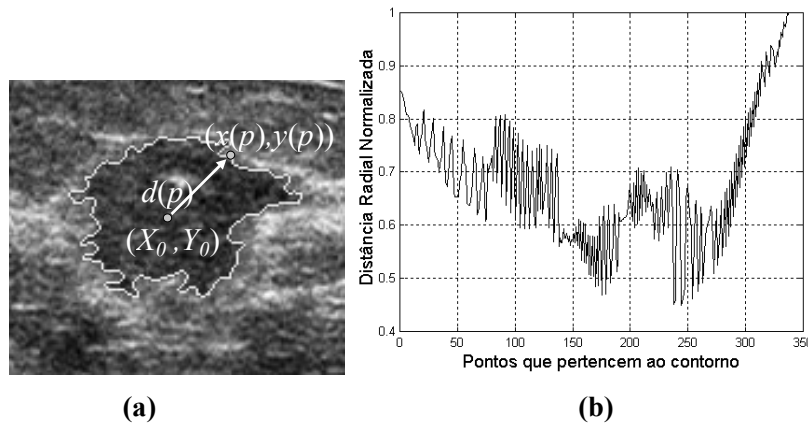


Figura V.1. Exemplo da determinação da **(b)** distância radial normalizada num **(a)** tumor previamente segmentado.

A partir de $d(p)$, pode-se calcular parâmetros para quantificar as características do contorno do objeto em análise. A média (Eq. V.2) e o desvio padrão (Eq. V.3) fornecem uma medida das variações macroscópicas do contorno, enquanto a razão de área (*area ratio*) (Eq. V.4) mede o percentual do contorno maior que a região circular de raio igual ao comprimento radial médio ($\overline{d(p)}$). Por outro lado, o índice de rugosidade (Eq. V.5) fornece informação sobre as pequenas variações presentes no contorno.

1 – Média:

$$\bar{d} = \frac{1}{P} \sum_{p=1}^P d(p) \quad (\text{V.2})$$

2 – Desvio padrão:

$$D_{NRL} = \sqrt{\frac{1}{P-1} \sum_{p=1}^P (d(p) - \bar{d})^2} \quad (\text{V.3})$$

3 – Razão de área:

$$RA = \frac{1}{\bar{d} \cdot P} \sum_{p=1}^P \beta(p), \text{ onde } \beta(p) = \begin{cases} d(p) - \bar{d}, & d(p) > \bar{d} \\ 0, & d(p) \leq \bar{d} \end{cases} \quad (\text{V.4})$$

4 – Rugosidade:

$$R = \frac{1}{P} \sum_{p=1}^P |d(p) - d(p+1)| \quad (\text{V.5})$$

V.1.2 Razão entre largura e profundidade

A razão entre profundidade e largura de um objeto (Figura V.2) (*Depth-to-Width Ratio* - DWR) é um parâmetro específico à análise de tumores de mama em imagens por US. Tumores benignos tendem a ser mais largos do que profundos, em relação à pele, produzindo valores de DWR menores, o que contrasta com os tumores malignos que tendem a ser mais profundos (HORSCH *et al.*, 2002).

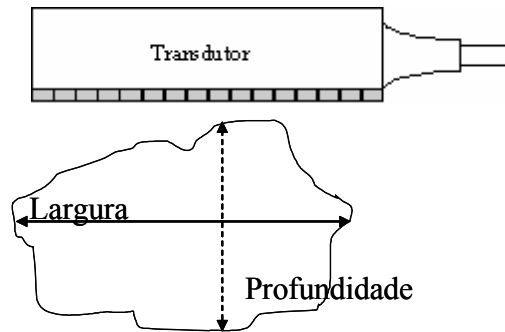


Figura V.2. Representação da largura (linha contínua) e profundidade (linha pontilhada) de um objeto em relação à posição do transdutor de US.

V.1.3 Circularidade

A circularidade fornece uma medida de quão circular é um objeto e relaciona o perímetro e a área do objeto, sendo definida por (CHOU *et al.*, 2001):

$$C = \frac{P^2}{A}, \quad (V.6)$$

onde P é o número total de *pixels* que formam o contorno (perímetro) e A é a área (total de *pixels*) do objeto sob análise. Caso o objeto, embora circular, possua uma grande quantidade de irregularidades em seu contorno, o valor da circularidade tende a crescer (CHOU *et al.*, 2001).

V.1.4 Número Substancial de Protuberâncias e Depressões

O Número Substancial de Protuberâncias e Depressões (NSPD) é um parâmetro que fornece o número total de pontos de mínimo e máximo no contorno de um objeto, em relação a um polígono convexo (CHEN *et al.*, 2003b), definido como o menor contorno convexo que envolve o objeto, conforme ilustrado na Figura V.3. Quanto mais irregular o contorno do objeto, maior será o valor de NSPD, pois, maior será o número de pontos de mínimo e máximo em seu contorno.



Figura V.3. Exemplo de objeto (contorno com protuberâncias e depressões) e seu respectivo polígono convexo (menor contorno convexo).

V.1.5 Índice de Lobulação

O índice de lobulação (*Lobulation Index* – LI) foi criado para caracterizar a distribuição de tamanho dos lóbulos presentes em um objeto. Um lóbulo é definido como a região formada pelo contorno do tumor e a linha que une os pontos de mínimo de duas concavidades adjacentes, sendo o tamanho do lóbulo a área A_i desta região, mostrado na Figura II.10 em cinza. Os pontos de mínimos $[w_i]$ das concavidades presentes no contorno são os mesmos definidos em NSPD. Supondo que o objeto possua N_l lóbulos e A_{max} e A_{min} são, respectivamente, as áreas do maior e do menor lóbulo, pode-se definir LI como (CHEN *et al.*, 2003b):

$$LI = \frac{A_{\max} - A_{\min}}{\frac{1}{N_l} \sum_{i=1}^{N_l} A_i} \quad (V.7)$$

Tumores malignos tendem a possuir uma distribuição irregular dos lóbulos, produzindo valores elevados de LI . Além disso, este parâmetro pode caracterizar corretamente tumores benignos que possuam vários lóbulos grandes e de tamanhos similares (CHEN *et al.*, 2003b).

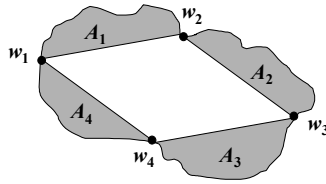


Figura V.4. Desenho representando os lóbulos em um objeto, onde $[w_1, w_2, w_3, w_4]$ são quatro pontos de mínimos das concavidades presentes no contorno e $[A_1, A_2, A_3, A_4]$ são as áreas dos lóbulos.

V.1.6 Perímetro Normalizado por Elipse Equivalente (Elliptic-Normalized circumference)

Para quantificar a sinuosidade do contorno de um objeto CHEN *et al.*, (2003b) propuseram a utilização de um parâmetro que relaciona os perímetros do contorno do objeto e de sua respectiva elipse equivalente, definido como:

$$ENC = \frac{P_O}{P_E}, \quad (V.8)$$

onde P_O é o perímetro (*pixels* pertencentes ao contorno) do objeto e P_E é o perímetro da elipse equivalente. Quanto mais irregular o contorno do objeto maior o valor de ENC .

Considerando uma elipse, com centro em $(0,0)$, definida por:

$$R = \{(x, y) : a \times x^2 + 2b \times xy + c \times y^2 = 1\}, \quad (V.9)$$

onde a , b e c são coeficientes, pode-se determinar a elipse equivalente de um objeto calculando-se momentos de segunda ordem, ou seja, as variâncias nas direções x (S_{xx}) e y (S_{yy}) e a orientação do objeto (S_{xy}), definidos como:

$$s_{xx} = \frac{1}{A} \sum_{(x,y) \in A} (x - X_o)^2 \quad (V.10)$$

$$s_{yy} = \frac{1}{A} \sum_{(x,y) \in A} (y - Y_o)^2 \quad (V.11)$$

$$s_{xy} = \frac{1}{A} \sum_{(x,y) \in A} (x - X_o)(y - Y_o) \quad (V.12)$$

onde (X_o, Y_o) são coordenadas do centróide do objeto e A é a área do objeto.

A partir destes momentos de segunda ordem do objeto, pode-se determinar os coeficientes $(a, b$ e $c)$ de sua elipse equivalente usando-se:

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \frac{1}{4s_{xx}s_{yy} - s_{xy}^2} \begin{pmatrix} s_{yy} & -s_{xy} \\ -s_{xy} & s_{xx} \end{pmatrix}, \quad (V.13)$$

V.1.7 Esqueleto Normalizado por Elipse Equivalente (Elliptic-Normalized Skeleton)

Considere uma nova imagem onde cada ponto que compõe um objeto (Figura V.5a) seja substituído pelo valor de sua distância, em *pixels*, até o ponto mais próximo do contorno (Figura V.5b). O seu “esqueleto” (Figura V.5d) será formado pelo o conjunto de pontos, da nova imagem, que são um máximo local (Figura V.5c) em relação aos pixels da vizinhança definida por uma dada janela (JAIN, 1995). O “esqueleto” de um objeto é sensível à sinuosidade do seu contorno, pois quanto mais sinuoso, maior o número de pontos que formam o “esqueleto” (Figura V.6). Logo, o mesmo pode ser utilizado como um parâmetro para quantificar a irregularidade do contorno. CHEN *et al.* (2003b) propõem como parâmetro a razão entre o número total de pontos que formam o “esqueleto” do objeto pelo o número de pontos que compõem a elipse equivalente do objeto. Esta normalização é realizada para tornar o parâmetro independente de variações no tamanho do objeto.

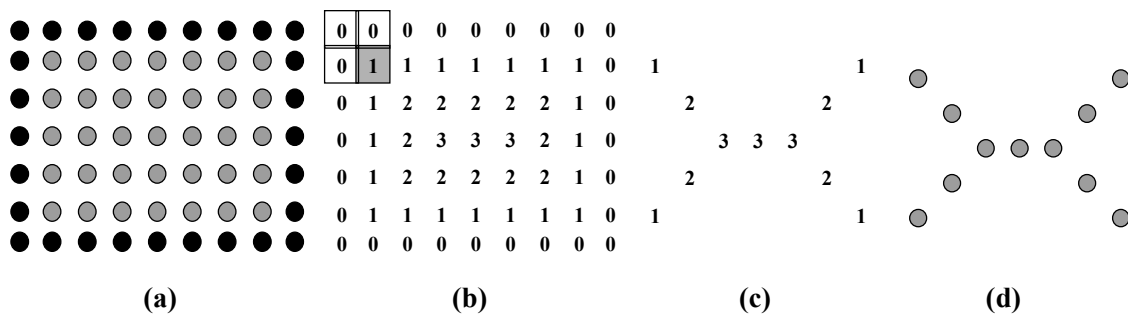


Figura V.5. Exemplo de **(a)** um objeto quadrado medindo 6 x 8 (em preto os pontos do contorno), **(b)** o objeto após a transformação de distância, onde cada valor indica a distância, em *pixels*, até o ponto mais próximo do contorno. Nota-se ainda a janela que define a vizinhança para a determinação do máximo local. **(c)** representação com os valores máximos locais. **(d)** esqueleto (em cinza) do objeto **(a)**.

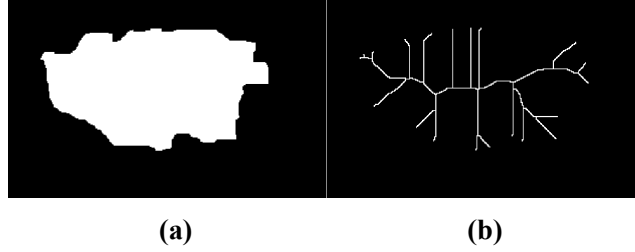


Figura V.6. Um **(a)** objeto e contorno irregular e seu respectivo **(b)** esqueleto. Note que devido à irregularidade do contorno, o esqueleto é mais complexo que o obtido na Figura V.5.

V.1.8 Razão de Superposição

A razão de superposição (RS) é definida como (HORSCH *et al.*, 2001):

$$RS = \frac{Area(S_i \cap S_o)}{Area(S_i \cup S_o)}, \quad (V.14)$$

onde S_o é o objeto sob avaliação e S_i o objeto de referência (que pode ser seu polígono convexo ou sua elipse equivalente), sendo que os símbolos \cap e \cup indicam interseção e união, respectivamente. Quando os contornos são idênticos, $RS = 1$.

V.1.9 Valor Médio Quadrático Residual Normalizado

Este parâmetro fornece uma medida global de imprecisão entre o objeto sob análise e um objeto de referência (polígono convexo ou elipse equivalente). Sua aplicação a imagens baseia-se na determinação de um resíduo S_r definido como:

$$S_r = Area|(S_i \cup S_o) - (S_i \cap S_o)| \quad (V.15)$$

onde S_o é o objeto sob avaliação e S_i o objeto de referência. Esta diferença representa a área residual existente entre S_i e S_o . Se as áreas são idênticas e estão na mesma posição, $S_r = 0$.

O valor médio quadrático residual normalizado (nrv) é definido como (INFANTOSI *et al.*, 1998):

$$nrv = \frac{\psi_r^2}{\psi_o^2} \quad (V.16)$$

onde ψ_r^2 é o valor médio quadrático de S_r e ψ_o^2 é valor médio quadrático de S_o . Alternativamente, pode-se substituir S_o pelo seu perímetro P_o . Com esta modificação, nrv passa a representar uma relação entre área residual e o perímetro do objeto, tornando-o

mais sensível às variações presentes no contorno. O valor de nrv para dois contornos coincidentes é nulo. Maiores detalhes sobre o uso do perímetro no lugar da área podem ser encontrados na publicação relacionada à Tese ALVARENGA *et al.* (2004).

V.1.10 Mshape

Mshape, parâmetro baseado na utilização de operadores morfológicos, é definido como:

$$Mshape = \frac{Area(S_o)}{Area(S_{fechamento})} \quad (V.17)$$

onde S_o e $S_{fechamento}$ são as áreas do objeto sob avaliação e de seu fechamento morfológico, respectivamente. O *fechamento morfológico* tende a ser maior que o objeto original, quanto maior for a irregularidade do seu contorno (Figura V.7). Assim, quanto mais irregular o contorno, menor é o valor de *Mshape*.

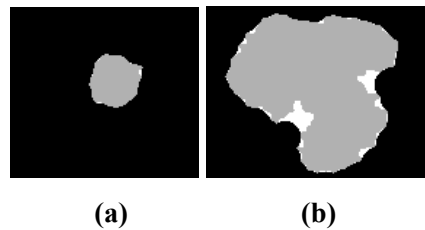


Figura V.7. Exemplo de **fechamento** morfológico (em branco) de um objeto (a) regular e outro (b) irregular.

V.2 Metodologia de Avaliação de Desempenho

Neste item, apresenta-se, de maneira sucinta, a Análise Discriminante Linear usada para selecionar os parâmetros mais significativos na distinção das características morfológicas dos tumores de mama, além da curva ROC usada como figura de mérito. A saber, as variáveis em **negrito** representam matrizes. Estas duas técnicas são aplicadas tanto nos parâmetros morfométricos, quanto nos parâmetros de textura (Capítulo VI).

V.2.1 Análise Discriminante Linear

A Análise Discriminante Linear (LDA), técnica amplamente utilizada para a classificação de dados, baseia-se na projeção das amostras a serem classificadas numa dada direção \mathbf{v} , que maximize a razão entre a dispersão interclasses (\mathbf{S}_b) e a dispersão intraclasses (\mathbf{S}_w) com base na função objetivo (PEDHAZUR, 1982):

$$J(\mathbf{v}) = \frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v}}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}} \quad (\text{V.18})$$

Onde o t sobrescrito indica matriz transposta. \mathbf{S}_b a dispersão entre as C classes a serem separadas é dada por:

$$\mathbf{S}_b = \sum_{c=1}^C n_c (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^t, \quad (\text{V.19})$$

onde n_c é o número de amostras da classe c , $\boldsymbol{\mu}_c$ é o valor médio da classe c e $\boldsymbol{\mu}$ é o valor médio de todas as amostras. A dispersão intraclasse \mathbf{S}_w é dada por:

$$\mathbf{S}_w = \sum_{c=1}^C \sum_{\mathbf{x}_c \in \text{classe } C} n_c (\mathbf{x}_c - \boldsymbol{\mu}_c)(\mathbf{x}_c - \boldsymbol{\mu}_c)^t, \quad (\text{V.20})$$

onde \mathbf{x}_c são as amostras de cada classe C .

Após algumas considerações, a direção ótima para projeção que maximiza $J(\mathbf{v})$ é obtida a partir de:

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{v} = \lambda \mathbf{v}. \quad (\text{V.21})$$

Detalhes da dedução do Discriminante Linear para duas classes são apresentados no Apêndice I.

V.2.2 A Curva ROC

A análise da curva característica de resposta do observador (*Receiver Operating Characteristic* – ROC), conhecida como curva ROC, é uma medida amplamente utilizada para avaliar o desempenho de testes diagnósticos. Para o seu completo entendimento, faz-se necessário definir alguns termos da literatura médica que definem o desempenho de um determinado teste diagnóstico em relação ao real diagnóstico do paciente (HUDSON e COHEN, 2000):

- Verdadeiro-Positivo (*VP*): teste diagnóstico é positivo e o paciente realmente possui a patologia;
- Falso-Positivo (*FP*): teste diagnóstico é positivo e o paciente não possui a patologia;
- Verdadeiro-Negativo (*VN*): teste diagnóstico é negativo e o paciente não possui a patologia;
- Falso-Negativo (*FN*): teste diagnóstico é negativo e o paciente possui a patologia.

A partir destes valores, é possível definir cinco figuras de mérito importantes para a análise detalhada do comportamento de um teste diagnóstico: acurácia, sensibilidade e especificidade.

A acurácia (Ac) dada pela equação (HUDSON e COHEN, 2000):

$$Ac = \frac{VP + VN}{VP + FN + VN + FP}, \quad (V.22)$$

indica o percentual total de acertos do teste diagnóstico.

A sensibilidade (S) definida como (HUDSON e COHEN, 2000):

$$S = \frac{VP}{VP + FN} \quad (V.23)$$

A especificidade (E) definida como (HUDSON e COHEN, 2000):

$$E = \frac{VN}{VN + FP}. \quad (V.24)$$

A resposta de um teste diagnóstico pode ser modelada por uma variável aleatória binomial, cujas funções de densidade de probabilidade para as duas hipóteses do teste apresentam uma distribuição normal (AZEVEDO-MARQUES, 2001). Estas distribuições podem estar superpostas (Figura V.8) implicando um compromisso entre os resultados falso-positivos e falso-negativos, sendo que uma mudança no limiar de decisão altera os valores de sensibilidade e especificidade (AZEVEDO-MARQUES, 2001) do teste.

A curva ROC relaciona a sensibilidade (e , portanto, considera os FN) com a especificidade (logo, dos FP), mais precisamente com a expressão $(1 - E)$ para qualquer valor do limiar de decisão, conforme exemplificado na Figura V.9. A área sob a curva ROC, medida usualmente empregada, varia de 0,5 (comportamento completamente aleatório) a 1,0 (discriminação perfeita). Com esta definição de curva ROC, quanto mais a curva se aproxima do canto superior esquerdo do gráfico (maior S e menor $1 - E$), melhor o desempenho do teste diagnóstico.

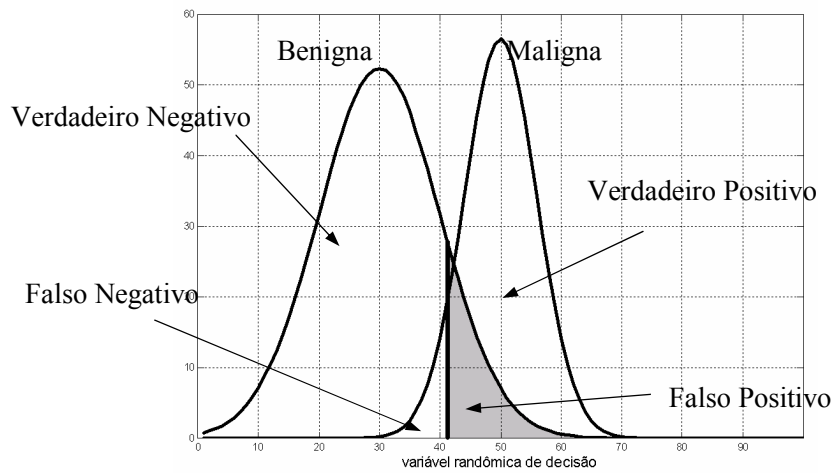


Figura V.8. Exemplo das distribuições de probabilidade das amostras benignas e malignas para análise da curva ROC, supondo um possível limiar. Nota-se a superposição entre as distribuições das amostras benignas e malignas, região onde ocorrem as classificações Falso-Positivo e Falso Negativo.

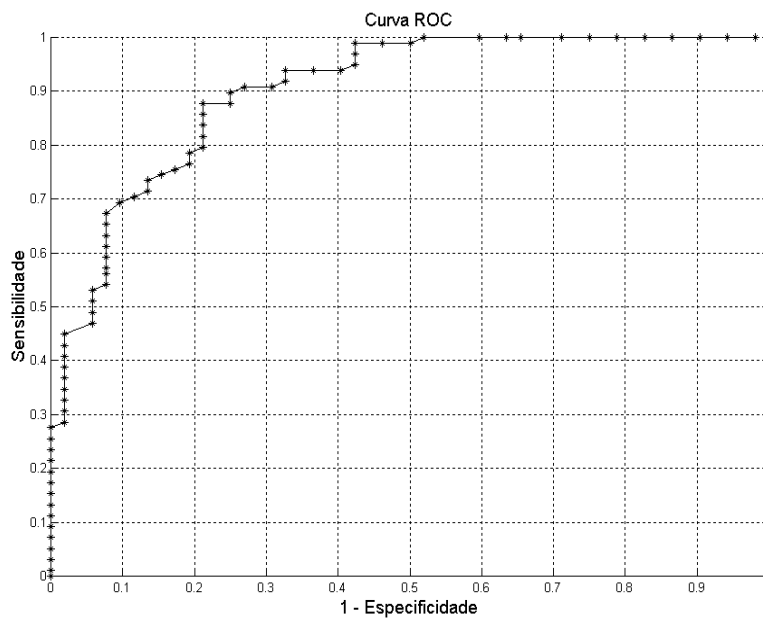


Figura V.9. Exemplo de curva ROC. Os (*) representam os pontos de operação sensibilidade x (1-especificidade) sobre a curva.

V.3 Aplicação de Parâmetros Morfométricos para Classificação de Tumores de Mama em Imagens por US

A razão entre a largura e a altura (*Depth-to-Width Ratio – DWR* – item V.1.2) é o parâmetro mais utilizado para auxiliar na distinção entre tumores de mama, em malignos ou benignos, em imagens por US (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002; ZHENG *et al.* 1997). Porém, sempre em conjunto com outros parâmetros que quantificam as irregularidades presentes no contorno dos tumores.

Uma das técnicas utilizadas para quantificar irregularidade no contorno de um tumor é a determinação do seu *gradiente radial normalizado* (*Normalized Radial Gradient – NRG*) que é uma medida da orientação média do gradiente dos níveis de cinza ao longo das margens do tumor. O *NRG* aparece na literatura combinado ao *DWR* e a um terceiro parâmetro, a *autocovariância* da ROI que contém o tumor, responsável pela quantificação das características de textura para separar tumores de mama, em imagens por US, entre malignos ou benignos (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002).

A determinação de parâmetros baseados no *comprimento radial normalizado* (*Normalized Radial Length – NRL* – item V.1.1) é outra técnica utilizada para quantificar a irregularidade. O *desvio padrão* (Eq. V.3) e a *rugosidade* (Eq. V.5) do *NRL* são apontados (CHOU *et al.*, 2001) como os mais significativos para classificar tumores de mama entre benignos ou malignos usando imagens por US. Combinando estes dois parâmetros a um terceiro, a *circularidade* (Eq. V.6) do tumor, CHOU *et al.* (2001) obtiveram uma sensibilidade de 97,0%, para uma especificidade de 80,0%. Neste estudo os tumores foram segmentados manualmente por um radiologista experiente.

CHEN *et al.* (2003b) propuseram quatro novos parâmetros para quantificar irregularidades no contorno de tumores de mama em imagens por US: o *número de depressões e protuberâncias substanciais* (*Number of Substantial Protuberances and Depressions - NSPD*), *índice de lobulação* (*Lobulation Index - LI*) (Eq. V.7), *perímetro normalizado por elipse equivalente* (*Elliptic-Normalized Circunference - ENC*) (Eq. V.8) e o *esqueleto normalizado por elipse equivalente* (*Elliptic-Normalized Skeleton – ENS*). A utilização destes quatro parâmetros na discriminação entre tumores de mama em imagens por US forneceu uma sensibilidade de 90,6% e uma especificidade de 97,2%.

A razão de superposição (RS) (Eq. V.14), que é um parâmetro adequado para comparar a forma de um objeto em relação à outra forma usada como referência, foi empregada (HORSCH *et al.*, 2001; ALVARENGA *et al.*, 2002b – publicação relacionada à Tese) para comparar o resultado da segmentação automática de tumores de mama em imagens por US com o contorno obtido por radiologistas experientes.

Neste trabalho, investigando o desempenho de diferentes parâmetros como indicadores de irregularidade do contorno dos objetos, propôs-se a utilizar o valor médio quadrático residual normalizado (nrv) (Eq. V.15). Este parâmetro, usado originalmente para a avaliação do desempenho de interpoladores no mapeamento cerebral através de EEG (INFANTOSI *et al.*, 1998), forneceu mais detalhes no contorno de tumores de mama em imagens por US do que o parâmetro RS . Maiores detalhes sobre o desempenho de nrv frente a RS podem ser obtidos na publicação relacionada à Tese ALVARENGA *et al.* (2004).

V.4 Cálculo dos Parâmetros Morfométricos para Imagens de US de mama

Os sete parâmetros morfométricos, a saber, desvio padrão (D_{NRL}) (Eq. V.3), razão de área (RA) (Eq. V.4), rugosidade (R) (Eq. V.5), razão de superposição (RS) (Eq. V.14), do valor médio quadrático residual normalizado (nrv) (Eq. II.16), $Mshape$ e circularidade (c) (Eq. V.6) foram avaliados quanto à capacidade de distinguir os tumores de mama no que se refere ao tipo de contorno (irregular ou regular) e de diagnóstico (maligno ou benigno), usando contornos obtidos pelo SAC. Os parâmetros D_{NRL} , RA e R foram calculados a partir da distância radial normalizada.

Como visto no item V.1, para o cálculo dos parâmetros RS e nrv , é necessário utilizar uma imagem de referência. Neste trabalho, optou-se por se determinar o polígono convexo (CASTLEMAN, 1996) de cada contorno como tal. Na Figura V.10, nota-se que quanto mais irregular é o contorno, maior a diferença para seu polígono convexo.

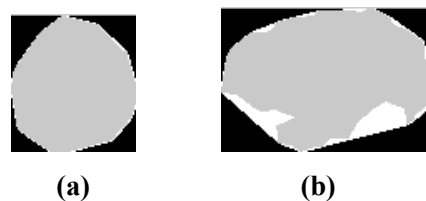


Figura V.10. Polígono convexo (em branco) de dois tumores de mama (a) regular e (b) irregular, segmentados usando SAC.

A todas as combinações dos sete parâmetros citados anteriormente aplicou-se a análise discriminante linear (LDA), sendo o desempenho avaliado com base nos valores da área (A_z) sobre a curva ROC, acurácia (Ac), sensibilidade (S) e especificidade (E). Tais medidas de desempenho foram calculadas, para o teste de distinção quanto ao contorno, considerando-se os tumores irregulares como casos positivos e regulares como negativos. Critério similar foi utilizado na distinção quanto ao diagnóstico, sendo tumores malignos considerados casos positivos e benignos negativos.

Para a aplicação da LDA foram utilizadas 152 imagens ($\approx 60\%$ das 246 segmentadas) de tumores de mama classificados pelo radiologista, com base no contorno, em 92 irregulares e 60 regulares. Quanto ao diagnóstico, 100 eram malignos e 52 benignos (*intranet* – seção IV.1). A utilização de somente parte das imagens teve como objetivo a seleção dos parâmetros morfométricos de melhor desempenho, de modo que o conjunto total de imagens pudesse ser utilizado na análise do desempenho combinando tais parâmetros aos de textura.

V.5 Avaliando os Parâmetros Morfométricos na Distinção dos Tumores: Contorno e Diagnóstico

V.5.1 Irregular x regular

O desempenho dos sete parâmetros morfométricos, tomados separadamente, em distinguir o contorno dos tumores, com base em A_z (Tabela V.1), indicou *nrv* (0,97) como o parâmetro de melhor desempenho e *Mshape* (0,55) como o pior. Os maiores valores de acurácia (93,4%) e de especificidade (94,6%) foram obtidos por *nrv*, embora *Mshape* tenha resultado na maior sensibilidade (95,7%), porém com baixa acurácia (60,5%) e muito baixa especificidade (6,7%).

Entre os parâmetros calculados a partir da NRL (D_{NRL} , RA , R), a maior Ac foi obtida por D_{NRL} e RA , com valores de S e E da mesma ordem de grandeza, enquanto R apresentou a segunda pior acurácia entre os sete parâmetros. O parâmetro C , segundo melhor com base em A_z , apresentou um desempenho equilibrado, com valores de Ac , S e E em torno de 88%. O parâmetro RS , terceiro melhor desempenho em termos de A_z , apresentou o menor valor de S e a terceira especificidade entre os todos os parâmetros.

Tabela V.1. Desempenho individual dos parâmetros na classificação dos tumores, quanto ao contorno, em irregular ou regular

Parâmetros	A_z	Ac (%)	S (%)	E (%)
<i>nrv</i>	0,97	93,4	94,6	91,7
<i>C</i>	0,93	88,2	88,0	88,3
<i>RS</i>	0,91	83,6	81,5	86,7
<i>RA</i>	0,71	71,7	87,0	48,3
D_{NRL}	0,70	71,7	85,9	50,0
<i>R</i>	0,61	65,8	87,0	33,3
<i>Mshape</i>	0,55	60,5	95,7	6,7

As curvas ROC (Figura V.11) evidenciam que o melhor desempenho de *nrv*, em comparação a *C* e *RS*, pode ser alcançado quando a especificidade varia de 65,5% a 98,0% (faixa indicada por setas).

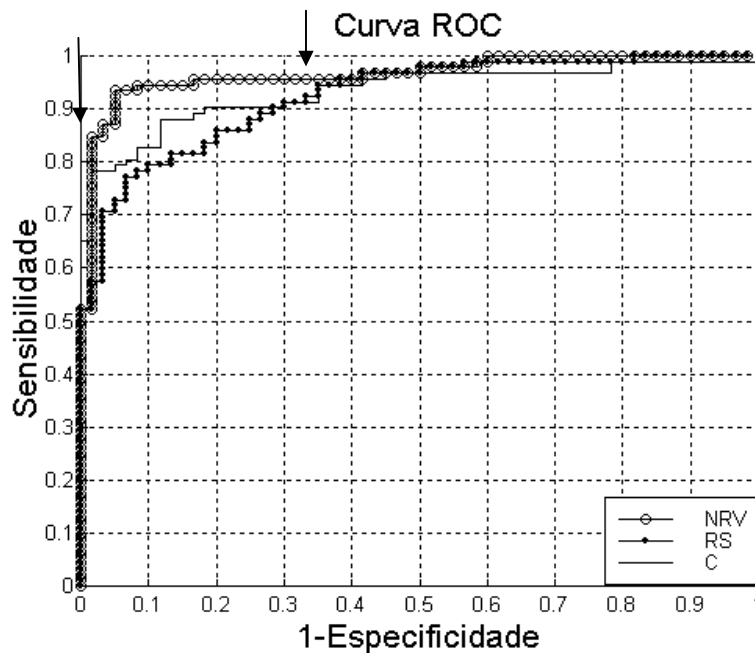


Figura V.11. Curva ROC dos três parâmetros de melhor desempenho analisados individualmente. As setas indicam a região (E entre 65,5% e 98,0%) onde se obtém o melhor desempenho pelo parâmetro *nrv*.

Usando-se dois parâmetros em conjunto, o melhor desempenho foi alcançado por *nrv* e *C* (Tabela V.2), por ter resultado em aumento de *Ac* para 94,7%, de *S* (95,7%) e de *E* (93,3%). No diagrama de espalhamento de *nrv* x *C*, mostrado na Figura V.12, nota-se que o parâmetro *nrv* tende a se concentrar para tumores irregulares ($\mu = 0,71$, $\sigma = 0,17$) e a se espalhar para os regulares ($\mu = 0,08$, $\sigma = 0,44$), enquanto que o parâmetro *C* possui comportamento oposto (irregulares: $\mu = -0,45$, $\sigma = 0,37$; regulares: $\mu = -0,87$, $\sigma = 0,08$). A função linear que melhor separa os dois grupos ($y = -2,26x + 0,38$) é também mostrada na Figura V.12 e, a respectiva curva ROC, é apresentada na Figura V.14.

Tabela V.2. Desempenho das melhores combinações entre os parâmetros na classificação dos tumores, quanto ao contorno, em irregular ou regular

Parâmetros	A_z	<i>Ac</i> (%)	<i>S</i> (%)	<i>E</i> (%)
<i>nrv</i> e <i>C</i>	0,97	94,7	95,7	93,3
D_{NRL} , <i>C</i> e <i>R</i>	0,91	87,5	81,5	96,7
<i>nrv</i> , <i>C</i> e <i>R</i>	0,97	96,1	95,7	96,7

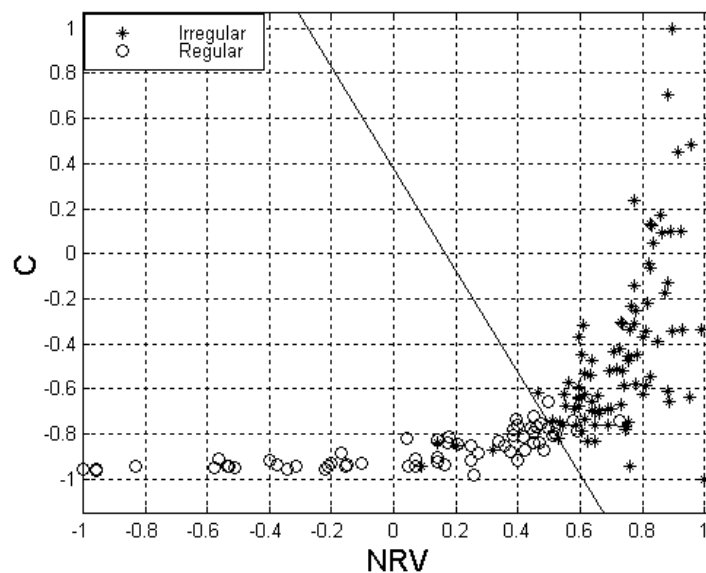


Figura V.12. Diagrama de espalhamento *nrv* x *C* e a função linear que melhor separa os tumores irregulares dos regulares, reta $y = -2,26x + 0,38$.

Considerando-se três parâmetros, o melhor desempenho foi obtido como o conjunto nrv , C e R , que resultou em aumento da Ac para 96,1% e da E (96,7%), embora S (95,7%) não tenha sido alterada (Tabela VIII.2). No diagrama de espalhamento de $nrv \times C \times R$ (Figura V.13), nota-se que, apesar do fraco desempenho individual do parâmetro R , quando combinado com nrv e C , este auxilia na distinção dos tumores quanto ao contorno. O plano que melhor separa os dois grupos ($z = -2,48x - 3,45y - 0,62$) é também mostrado na Fig V.13.

Aumentando-se o número de parâmetros, tomados em conjunto, não se obteve melhora no desempenho. As curvas ROC das melhores combinações de dois e três parâmetros (Figura V.14), indica desempenhos similares exceto no que concerne à sensibilidade 95,7% (indicada por seta), valor este para o qual pode-se alcançar uma maior especificidade desde de que se considere o conjunto de três parâmetros formado por nrv , C e R .

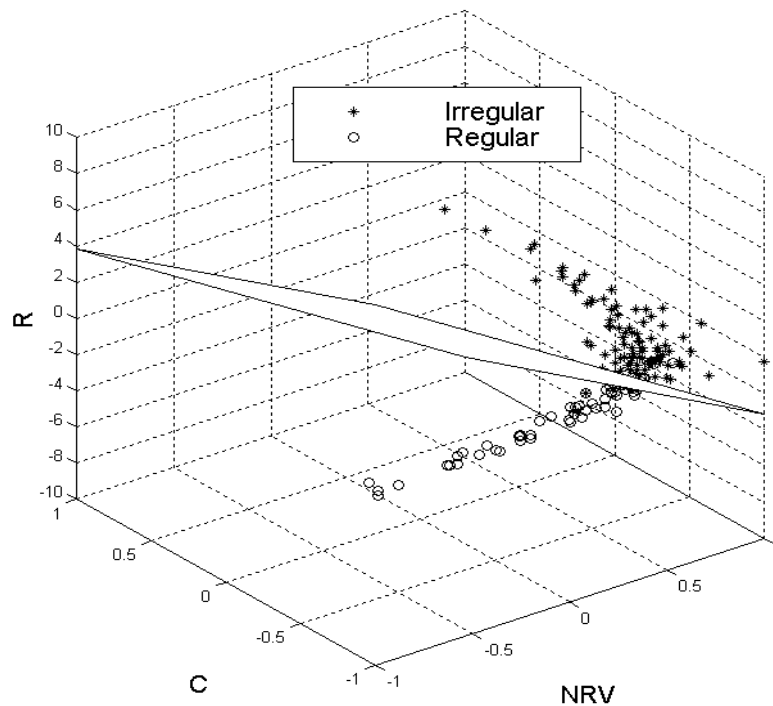


Figura V.13. Diagrama de espalhamento de $nrv \times C \times R$ e o plano ($z = -2,48x - 3,45y - 0,62$) que melhor separa os tumores irregulares dos regulares.

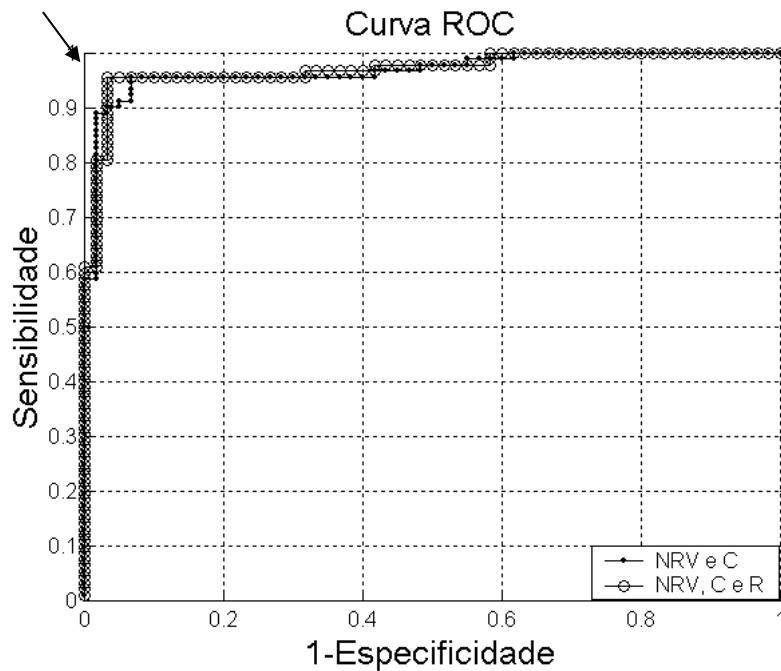


Figura V.14. Curvas ROC das combinações de dois e três parâmetros de melhor desempenho na separação entre tumores quanto ao contorno. A seta indica o ponto da curva onde se obtém o melhor desempenho ($S = 95,7\%$ e $E = 96,7\%$).

V.5.2 Maligno x benigno

Similarmente ao resultado obtido na distinção dos tumores quanto ao contorno, o melhor desempenho em distinguir tumores quanto ao diagnóstico (maligno ou benigno) foi obtido por nrv ($A_z = 0,91$) e pior por $Mshape$ ($A_z = 0,57$) (Tabela V.3). Os maiores valores de acurácia (88,2%) e de especificidade (92,3%) foram também obtidos pelo parâmetro nrv , enquanto RA apresentou a maior sensibilidade (93,0%), porém com baixa especificidade (36,5%).

O desempenho dos parâmetros calculados a partir da NRL (D_{NRL} , RA , R) foram, mais uma vez, similares aos da distinção do contorno. D_{NRL} e RA resultaram em maior Ac , com valores de S e E da mesma ordem de grandeza, enquanto R apresentou a segunda pior acurácia entre os sete parâmetros. O segundo melhor parâmetro com base em A_z foi C , que, novamente, apresentou um desempenho equilibrado no que tange a Ac , S e E ($\approx 80\%$). O parâmetro RS , terceiro melhor desempenho em termos de A_z , resultou na segunda menor S e a segunda maior E entre os todos os parâmetros.

Tabela V.3. Desempenho individual dos parâmetros na distinção dos tumores entre benigno ou maligno

Parâmetros	A_z	Ac (%)	S (%)	E (%)
<i>nrν</i>	0,91	88,2	86,0	92,3
<i>C</i>	0,84	80,3	79,0	82,7
<i>RS</i>	0,81	74,3	68,0	86,5
<i>R</i>	0,64	69,1	82,0	44,2
D_{NRL}	0,62	71,7	90,0	36,5
<i>RA</i>	0,61	73,0	93,0	34,6
<i>Mshape</i>	0,57	56,6	42,0	84,6

O uso combinado de dois parâmetros não implicou em melhoria de desempenho. Entretanto, ao se tomar o conjunto formado por *nrν*, D_{NRL} e *R* (Tabela V.4), obteve-se o melhor resultado entre os demais conjuntos de três parâmetros, ou seja, aumento de Ac para 88,8% e de E para 94,2%, sem, contudo, alterar a S (86,0%). O diagrama de espalhamento de *nrν* x D_{NRL} x *R* mostrado na Figura V.15 indica que, apesar do fraco desempenho individual dos parâmetros *R* e D_{NRL} , quando combinados com *nrν*, resulta em melhor desempenho na distinção entre tumores malignos ou benignos. O plano que melhor separa os dois grupos ($z = -2,26x + 0,55y + 0,10$) é também mostrado na Figura V.15. Caso se utilizasse o parâmetro *C* ao invés de D_{NRL} em conjunto com *nrν* e *R*, a acurácia não se alteraria, porém haveria aumento de sensibilidade (88,0%) e redução de especificidade (90,4%), conforme Tabela V.4. Para este conjunto, o plano que melhor separa os dois grupos ($z = -1,43x - 0,99y - 0,05$) é mostrado na Figura V.16. O aumentando do número de parâmetros tomados em conjunto, não resultou em melhoria no desempenho do classificador. As curvas ROC das melhores combinações três parâmetros (Figura V.17) evidenciam o conjunto *nrν*, D_{NRL} , e *R* como o de melhor desempenho, em particular para especificidade superior a 94% (seta na Figura V.17).

Tabela V.4. Desempenho das melhores combinações entre os parâmetros na distinção dos tumores entre benigno ou maligno

Parâmetros	A_z	Ac (%)	S (%)	E (%)
nrv	0,91	88,2	86,0	92,3
$nrv, D_{NRL}, e R$	0,92	88,8	86,0	94,2
$nrv, C e R$	0,92	88,8	88,0	90,4

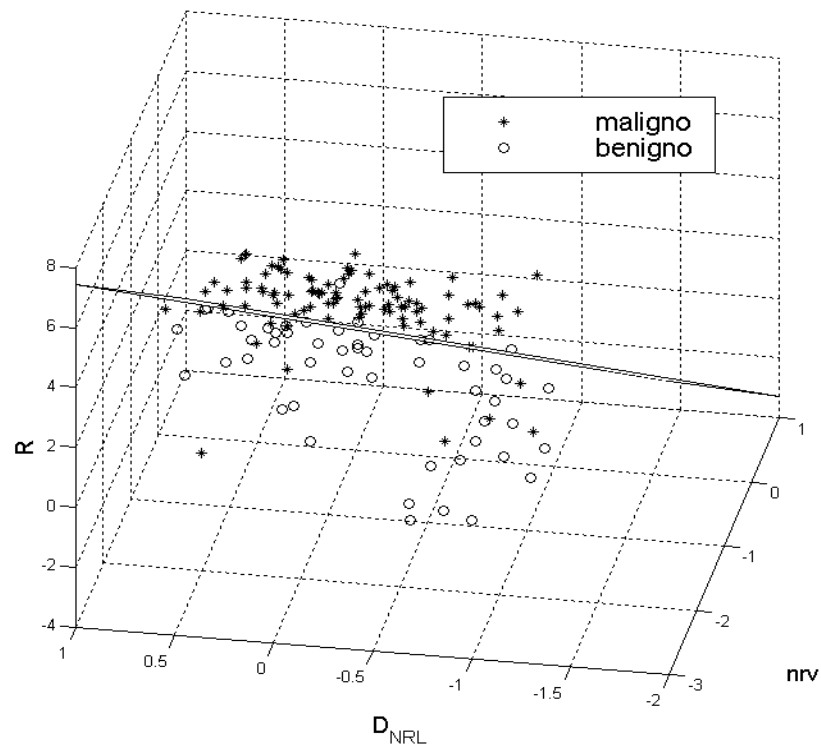


Figura V.15. Diagrama de espalhamento de $nrv \times D_{NRL} \times R$ e o plano ($z = -2,26x + 0,55y + 0,10$) que melhor separa os tumores malignos dos benignos.

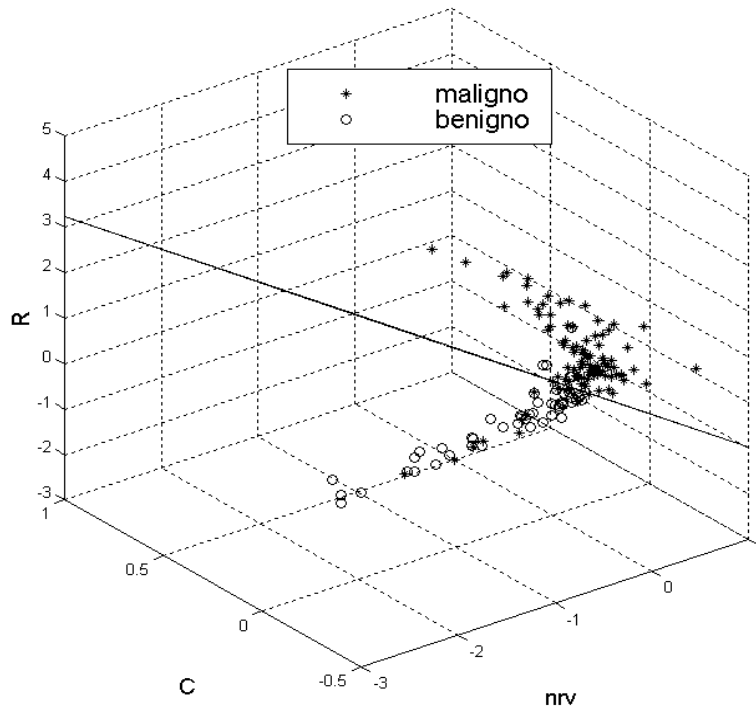


Figura V.16. Diagrama de espalhamento de $nrv \times C \times R$ e o plano ($z = -1,43x - 0,99y - 0,05$) que melhor separa os tumores malignos dos benignos.

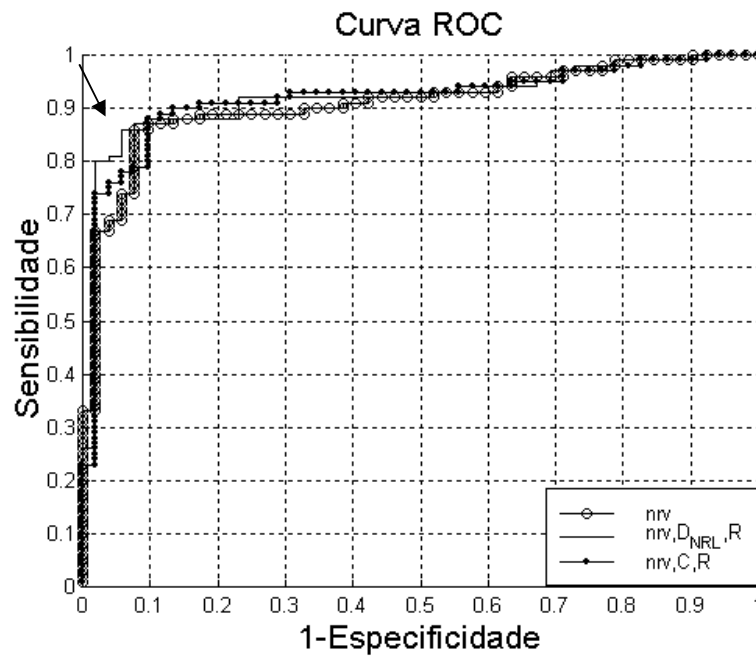


Figura V.17. Curvas ROC das combinações de três parâmetros de melhor desempenho na separação dos tumores quanto ao diagnóstico. A seta indica o ponto da curva onde se obtém o melhor desempenho ($S = 86,0\%$ e $E = 94,2\%$).

V.6 Discussão

Na distinção dos tumores de mama em imagens por US, quanto ao contorno, os três parâmetros que apresentaram melhor desempenho individual foram o valor médio quadrático residual normalizado (nrv) e a razão de superposição (RS), ambos calculados com base no polígono convexo, e a circularidade (C). Estes parâmetros também foram os de melhor desempenho na distinção dos tumores entre malignos ou benignos. Tal evidência sugere uma estrita ligação entre a irregularidade do contorno do tumor e o fato deste ser maligno. Porém, esta não pode ser tomada como uma única e definitiva característica para o diagnóstico de malignidade. Isto fica evidente ao se analisar a acurácia destes parâmetros na distinção dos tumores entre irregular ou regular ($Ac = 93,4\%$) e entre maligno ou benigno ($Ac = 88,2\%$). Logo, deve haver pelo menos 5,2% de tumores que dependem, além do contorno, de uma ou mais características (tal como a textura) para serem corretamente diagnosticados.

O parâmetro que mais se destacou em toda a análise foi nrv , pois seu desempenho individual na distinção entre tumores malignos ou benignos, em termos de acurácia, sensibilidade e especificidade, é da mesma ordem de grandeza que o desempenho obtido com a combinação de três parâmetros (Tabela V.4). Este desempenho superior pode estar relacionado ao fato de nrv ser uma relação entre a área residual, mais sensível a variações no contorno, e o contorno de referência, definido pelo polígono convexo.

Embora RS tenha resultado em terceiro melhor desempenho individual (Tabela V.1), quando em conjunto com os demais parâmetros este não contribui para a melhoria na distinção entre tumores malignos ou benignos. Por outro lado, a rugosidade (R) e o desvio padrão (D_{NRL}) que resultaram no quarto e quinto melhores desempenhos individuais, respectivamente, contribuíram para um melhor desempenho dos conjuntos de parâmetros (Tabela V.4). Segundo CHOU *et al.* (2001), entre os parâmetros calculados a partir do NRL, estes dois (R e D_{NRL}) foram apontados como os mais significativos para classificar tumores de mama, entre maligno ou benigno, em imagens por US, e, quando combinados com C , forneceram uma sensibilidade de 97,0%, para uma especificidade de 80,0%.

Neste trabalho, os conjuntos de três parâmetros formados por (nrv , D_{NRL} e R) e (nrv , C e R) resultaram em valores de especificidade (94,2% e 90,4%, respectivamente) superiores as obtidas de CHOU *et al.* (2001), embora com sensibilidade inferior (86,0% e

88,0%, respectivamente). Maior sensibilidade seria obtida apenas com uma redução expressiva da especificidade (Figura V.17).

Além das evidências clínicas de que o contorno regular tende a ser uma característica de elevada preditividade para benignidade (SKAANE, 1999; RAHBAR *et al.*, 1999; HUBER *et al.*, 2000), o desempenho dos parâmetros morfométricos na distinção dos tumores entre malignos ou benignos (superior a 90% - Tabela V.4), sugere que estes podem contribuir para a redução do número de biópsias em tumores benignos (elevado valor de especificidade), em particular com o parâmetro *nrv*.

VI PARÂMETROS DE TEXTURA NA DISTINÇÃO DOS TUMORES: ECOTEXTURA E DIAGNÓSTICO

Como visto no Capítulo II, os ecos provenientes das regiões interna e adjacente de um tumor maligno tendem a produzir uma textura heterogênea e desorganizada, características que podem diferenciar este tumor de um benigno. Desta forma, os parâmetros de textura são calculados com o intuito de quantificar as mesmas.

Neste Capítulo, as principais técnicas para se calcular os parâmetros de textura são apresentadas. Uma breve revisão da literatura aponta a potencialidade da aplicação destes parâmetros às imagens de mama. Em seguida, estes são aplicados ao banco de imagens de mama, apresentado no item IV.1, e avaliados quanto ao seu desempenho em distinguir tumores quanto a ecotextura e sua contribuição para o diagnóstico. Por fim, seus desempenhos são discutidos à luz da literatura.

VI.1 Técnicas para Quantificar as Características de Textura de Imagens

“A análise de textura é importante em vários problemas de segmentação e classificação de imagens. É uma tecnologia bem desenvolvida e tem sido usada para a análise de todos os tipos de imagens. Desde as microscópicas na área biomédica até as de satélite da superfície da terra”
(AL-JANOBI, 2001).

Uma imagem digital $f(x,y)$ pode ser representada por uma matriz, cujos elementos, conhecidos como *pixels*, são números que definem o valor do nível de cinza em pontos da imagem. A textura da imagem, definida como uma função da variação espacial na intensidade dos *pixels* (TUCERYAN & JAIN, 1998), pode apresentar um padrão periódico ou aleatório e ser quantificada através do cálculo de parâmetros de textura.

As técnicas para determinação dos parâmetros de textura baseiam-se na extração de informação da imagem por meio da manipulação numérica dos *pixels*. Geralmente, a informação é representada através de um valor que quantifica alguma característica referente à variação dos níveis de cinza e à posição relativa dos *pixels* (TOURASSI, 1999).

As principais técnicas usualmente aplicadas a imagens de mama são: Parâmetros estatísticos do histograma da imagem, Matriz de Co-ocorrência, *Run-Length Analysis*, Markoviano, Coeficientes de Autocovariância Bi-dimensional Normalizados, Curva de Complexidade e Diferença Mínima Lateral.

VI.1.1 Parâmetros estatísticos do histograma da imagem

Os parâmetros como a média, o desvio padrão e a variância, obtidos a partir do histograma de uma imagem, são os mais simples de serem calculados e representam a distribuição dos valores, em níveis de cinza, dos *pixels* da região em análise, porém sem levar em consideração suas posições relativas (HUBER *et al.*, 2000).

VI.1.2 Matriz de Co-ocorrência

Para inserir na análise estatística a informação da posição relativa entre os *pixels* da imagem, a abordagem mais utilizada na literatura é o cálculo da matriz de co-ocorrência (GLCM – *Grey Level Co-occurrence Matrix*). A GLCM é um histograma bi-dimensional ($G \times G$ níveis de cinza) de uma imagem $f(x,y)$, de tamanho $M \times N$, que descreve a ocorrência de pares de *pixels* de valores $\{i, j\} \in [0, 1, \dots, G - 1]$, separados por uma certa distância k , numa determinada direção θ (AL-JANOBI, 2001).

Dada uma imagem $f(x,y)$ com apenas dois níveis de cinza (e.g., 0 e 1) e dimensões 6×5 (Figura VI.1a), o procedimento para a obtenção da GLCM $m(i,j)$ (Figura VI.1b), para uma distância k de dois *pixels* e θ nulo, é ilustrado na Figura VI.1. Considerando-se a posição inicial $f(1,1)$ da imagem original, o primeiro par formado é $\{f(1,1) = 0, f(1,3) = 0\}$, resultando em $m(0,0) = 1$. Deslocando-se para a próxima posição à direita, ou seja, $f(1,2)$, o par formado é $\{f(1,2) = 1, f(1,4) = 1\}$, logo resultando em $m(1,1) = 1$. Similarmente, para $f(1,3)$, tem-se $m(0,0) = 2$. Procedendo deste modo para qualquer $f(x,y)$, os pares serão $\{f(x,y), f(x,y+k)\}$ para $x = \{1..6\}$ e $y = \{1..5\}$, obtendo-se, conforme ilustrado em Figura VI.1b, $m(0,0) = 10$, $m(0,1) = 0$, $m(1,0) = 0$ e $m(1,1) = 10$.

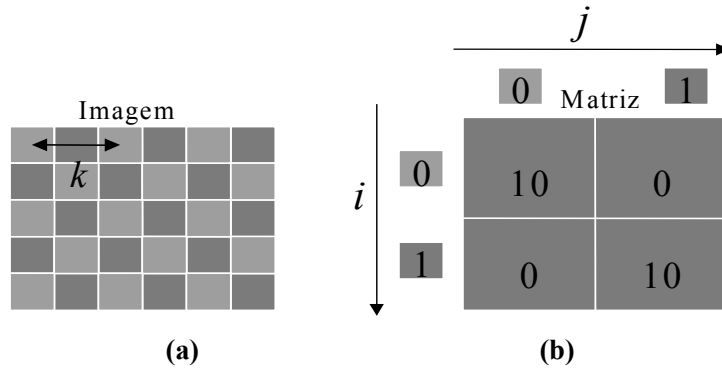


Figura VI.1. Determinação da GLCM **(b)** $m(i,j)$, a partir da imagem **(a)** $f(x,y)$ de dois níveis de cinza (0 e 1, sendo o último mais escuro) para distância k de dois *pixels* e direção θ nula. A matriz $m(i,j)$ resulta em 10 pares $\{0,0\}$ e em $\{1,1\}$, e nenhum par em $\{0,1\}$ e em $\{1,0\}$.

Originalmente, HARALICK *et al.* (1973) propuseram 14 parâmetros para quantificar textura usando a GLCM. Os parâmetros usualmente empregados (GOLDBERG *et al.*, 1992; GARRA *et al.*, 1993; DUTTA *et al.*, 1995) são definidos por AL-JANOBI (2001) como:

1 - Entropia (*coo*)

$$coo = \sum_{i,j} m(i,j) \cdot \log[m(i,j)], \quad (VI.1)$$

que mede quão aleatória é a textura da imagem. Valores elevados de *coo* indicam uma distribuição aleatória de níveis de cinza da GLCM.

2 - Segundo momento angular (*asm*), medida da homogeneidade na transição entre níveis de cinza na textura da imagem, pode ser expresso por:

$$asm = \sum_{i,j} m(i,j)^2. \quad (VI.2)$$

Uma maior homogeneidade ou um padrão regular de textura, implica geralmente valores elevados de *asm*.

3 - Desvio padrão (*std*) ou grau de espalhamento dos níveis de cinza entre pares de *pixels* é dado por:

$$std = \sqrt{\frac{1}{G-1} \sum_{i,j} (m(i,j) - \bar{m})^2}, \quad (VI.3)$$

onde G é o número de níveis de cinza presentes na imagem e \bar{m} o valor médio da matriz $m(i,j)$.

4 - Contraste (*con*), medida da dispersão ou da variabilidade local da imagem, é definido como:

$$con = \sum_{i,j} (i - j)^2 \cdot m(i, j). \quad (VI.4)$$

Imagens com muitas variações locais de níveis de cinza tendem a fornecer valores elevados de contraste.

5 - Correlação (*cor*), medida da dependência local entre os níveis de cinza em uma textura, pode ser expressa por:

$$cor = \frac{\sum_{i,j} i \cdot j \cdot m(i, j) - m_x m_y}{\sqrt{\sigma_x \cdot \sigma_y}}, \quad (VI.5)$$

onde

$$\bar{m}_x = \sum_i i * (\sum_j m(i, j)|_{i=0, \dots, G-1}) \text{ e } \bar{m}_y = \sum_j j * (\sum_i m(i, j)|_{j=0, \dots, G-1}) \quad (VI.6)$$

$$\sigma_x^2 = \sum_i (\sum_j m(i, j)|_{i=0, \dots, G-1}) * (i - \bar{m}_x)^2 \text{ e} \quad (VI.7)$$

$$\sigma_y^2 = \sum_j (\sum_i m(i, j)|_{j=0, \dots, G-1}) * (j - \bar{m}_y)^2$$

Valores elevados de *cor* são obtidos para regiões com elevada similaridade entre os níveis de cinza.

VI.1.3 Run-Length Analysis

Outra abordagem que leva em consideração a posição relativa entre os *pixels* é a *run-length analysis*. Uma *run*, ou corrida, é um conjunto de *pixels* consecutivos que pertencem a um mesmo nível de cinza (i) numa dada direção (horizontal ou vertical), sendo o comprimento (l), o número de *pixels* da “corrida”. Assim, o total de corridas para diferentes comprimentos l e níveis de cinza i forma a matriz *run-length* $r(i,l)$. A Figura VI.2 ilustra a aplicação desta técnica a uma imagem $f(x,y)$ de três níveis de cinza (e.g. 0, 1 e 2) para formar a matriz *run-length* na direção horizontal. Neste exemplo são enfatizadas na

imagem original (Figura VI.2a) as corridas de comprimento $l = 2$ e nível de cinza $i = 1$, resultando na quantidade de corridas $r(1,2) = 4$ na matriz *run-length*.

0	0	1	1	0	0
2	2	0	0	1	1
0	0	1	2	0	0
2	0	1	1	0	0
0	0	2	1	1	1
1	2	0	0	1	1

(a)

	l	1	2	3	4	5	6
i							
0		1	8	0	0	0	0
1		2	4	1	0	0	0
2		4	1	0	0	0	0

(b)

Figura VI.2. **(a)** Imagem $f(x,y)$ e **(b)** sua respectiva matriz *run-length*, para a direção horizontal. Notam-se grifadas as corridas de comprimento $l = 2$, para o nível de cinza $i = 1$ (Figura II.2a) e sua respectiva quantidade corridas $r(1,2) = 4$.

A partir de $r(i,l)$, parâmetros de textura podem ser extraídos, sendo os cinco parâmetros clássicos definidos como (TANG, 1998; HADJIISKI *et al.*, 2001):

1 – *Short Run Emphasis (sre)*, que enfatiza a presença de corridas curtas:

$$sre = \frac{1}{n_r} \sum_{i=0}^{G-1} \sum_{l=1}^L \frac{r(i,l)}{l^2}, \quad (VI.8)$$

onde n_r é o número total de *runs*, G é o número de níveis de cinza e L o comprimento máximo de uma corrida. Valores elevados indicam que a textura é formada por pequenos padrões repetitivos na imagem, ou a imagem possui inúmeras variações de nível de cinza na região em análise.

2 – *Long Run Emphasis (lre)*:

$$lre = \frac{1}{n_r} \sum_{i=0}^{G-1} \sum_{l=1}^L r(i,l) \cdot l^2, \quad (VI.9)$$

ênfatisa a existência de corridas longas. Valores elevados indicam que a textura possui pouca variações de níveis de cinza dentro a região sob análise.

3 – *Gray-Level Nonuniformity (gl)*:

$$gl = \frac{1}{n_r} \sum_{i=0}^{G-1} \left(\sum_{l=1}^L r(i,l) \right)^2. \quad (VI.10)$$

Como o próprio nome indica, mede a não uniformidade dos níveis de cinza da imagem. Se o número de corridas é igualmente distribuído por todos os níveis de cinza, gl tende a fornecer valores baixos.

4 – *Run-Length Nonuniformity (rl)*:

$$rl = \frac{1}{n_r} \sum_{l=1}^L \left(\sum_{i=0}^{G-1} r(i,l) \right)^2. \quad (\text{VI.11})$$

Este parâmetro mede a não uniformidade das corridas. Se o número de corridas é igualmente distribuído pelos diferentes comprimentos de corrida calculados, rl tende a fornecer valores baixos.

5 – *Run Percentage (rp)*:

$$rp = \frac{n_r}{n_p}, \quad (\text{VI.12})$$

onde n_p é o número total de *pixels* na imagem. Se o número total de corridas e o número total de *pixels* da imagem forem bem próximos, então rp tende a ser unitário, refletindo a enorme variabilidade dos valores de nível de cinza na região em análise.

Segundo TANG *et al.* (1998), estes cinco parâmetros foram formulados intuitivamente na tentativa de capturar algumas das propriedades da distribuição da matriz *run-length*, e possuem como limitação uma elevada correlação entre si.

VI.1.4 Markoviano

Para o cálculo de parâmetros baseados na estatística Markoviana, deve-se primeiro determinar uma matriz Mk (similar a GLCM), na qual cada elemento é a probabilidade condicional $p_k(i/j)$ de um nível de cinza i ocorrer k *pixels* distante de um nível de cinza j , sendo definida como GOLDBERG, *et al.*, 1992):

$$Mk = \begin{bmatrix} p_k(0/0) & p_k(0/1) & \cdots & p_k(0/G-1) \\ p_k(1/0) & p_k(1/1) & \cdots & p_k(1/G-1) \\ \vdots & \vdots & & \vdots \\ p_k(G-1/0) & p_k(G-1/1) & \cdots & p_k(G-1/G-1) \end{bmatrix}, \quad (\text{VI.13})$$

onde $\{i, j\} \in [0, 1, \dots, G-1]$ para a imagem $f(x,y)$, de tamanho $M \times N$ com G níveis de cinza.

Similarmente a GLCM, vários parâmetros estatísticos podem ser determinados a partir de Mk , como o contraste inverso (ic), que tende a produzir valores menores quanto maior for o contraste, e pode ser definido como:

$$ic = \sum_{i,j} \frac{p_k(i/j)}{|i-j|} \quad (VI.14)$$

VI.1.5 Coeficientes de Autocovariância Bi-dimensional Normalizados

Os coeficientes de autocovariância de uma imagem refletem quão correlacionados são os *pixels*, podendo ser utilizados como parâmetros na avaliação da regularidade da textura da imagem. O coeficiente de autocovariância $\rho(p,q)$, entre os *pixels* (x,y) e $(x+p,y+q)$ de uma imagem $f(x,y)$ de tamanho $M \times N$ é definido como (CHEN, D. *et al.*, 1998):

$$\rho(p,q) = \frac{1}{(M-p) \cdot (N-q)} \cdot \frac{\sum_{x=0}^{M-1-p} \sum_{y=0}^{N-1-q} |(f(x,y) - \bar{f})(f(x+p,y+q) - \bar{f})|}{\sum_{x=0}^M \sum_{y=0}^N f^2(x,y)}, \quad (VI.15)$$

onde \bar{f} é o valor médio dos níveis de cinza da imagem $f(x,y)$, $p \in [0,1,\dots,M-1]$ e $q \in [0,1,\dots,N-1]$.

Os valores do par (p,q) são estabelecidos com base na distância entre os *pixels* que se deseja analisar. Logo, o número de coeficientes de autocovariância é de $p \times q$. Por exemplo, para os coeficientes calculados considerando-se todas as combinações de (p,q) entre $(0,0)$ e $(9,9)$ têm-se cem parâmetros.

Ao invés da autocovariância, a autocorrelação tem sido usada (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002), ou seja, remove-se \bar{f} da Eq. V.15. Entretanto, este procedimento torna a avaliação da textura suscetível a variações de brilho entre as imagens.

VI.1.6 Curva de Complexidade

Dada uma imagem representada por $f(x,y)$, para qualquer valor de limiar α é possível gerar uma imagem binária tal que (BAHEERATHAN *et al.*, 1999):

$$f^\alpha(x,y) = \begin{cases} 1, & f(x,y) \geq \alpha \\ 0, & \text{nos demais casos} \end{cases} \quad (\text{VI.16})$$

onde $\alpha \in [0,1,\dots,G-1]$, G é o número de níveis de cinza da imagem $f(x,y)$.

A Figura VI.3 ilustra a determinação de uma imagem binária $f^\alpha(x,y)$ para $\alpha = 12$, a partir de uma imagem $f(x,y)$ de 16 níveis de cinza. Para $f^\alpha(x,y)$ pode-se determinar o número transições de preto para branco (0 para 1) e de branco para preto (1 para 0), numa dada direção, horizontal ou vertical. Neste exemplo, tomando-se o total de transições ocorridas na direção horizontal ($T_{1,0}^{12}$) e na vertical ($T_{0,1}^{12}$), obtém-se o ponto da curva de complexidade para $\alpha = 12$, que pode ser expresso por:

$$C(12) = \frac{T_{1,0}^{12} + T_{0,1}^{12}}{M \times (N - 1) + N \times (M - 1)}, \quad (\text{VI.17})$$

onde $M \times N$ é tamanho da imagem.

Logo, generalizando para qualquer valor α , a Curva de Complexidade (CC) pode ser expressa como uma função de α , ou seja:

$$C(\alpha) = \frac{T_{1,0}^\alpha + T_{0,1}^\alpha}{M \times (N - 1) + N \times (M - 1)}, \quad \alpha \in [0,1,\dots,G-1], \quad (\text{VI.18})$$

onde $T_{1,0}^\alpha$ e $T_{0,1}^\alpha$ são os números totais de transições em $f^\alpha(x,y)$ com passos fixos (1,0) e (0,1), respectivamente, e $M \times N$ é o tamanho da imagem. Na Figura VI.4 é mostrado $C(\alpha)$ para o exemplo da Figura VI.3a.

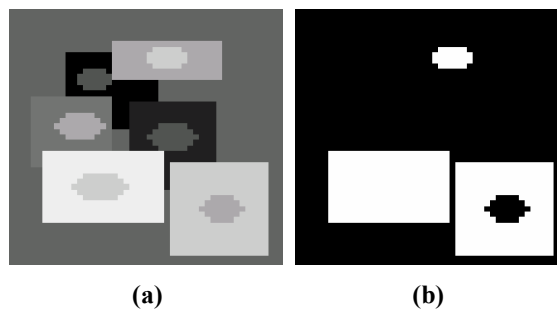


Figura VI.3. (a) Imagem $f(x,y)$ com 16 níveis de cinza e (b) sua respectiva imagem binária $f^\alpha(x,y)$ para um valor de limiar $\alpha = 12$.

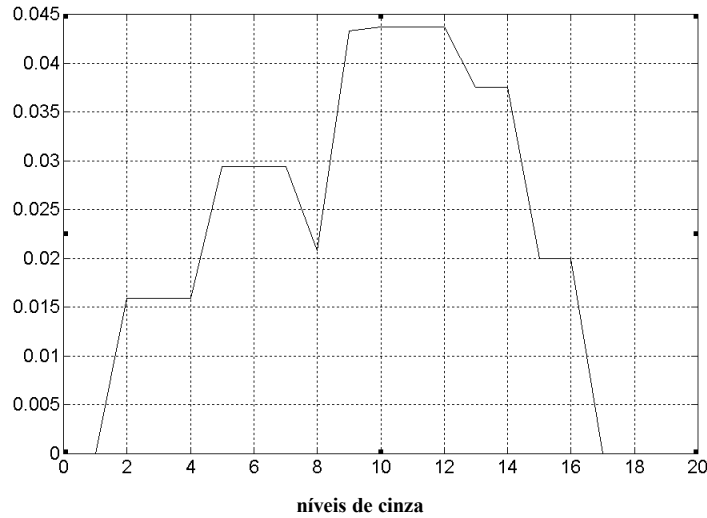


Figura VI.4. Curva de Complexidade $C(\alpha)$ para a imagem $f(x,y)$ da Figura II.3a.

A partir da CC, pode-se inferir que imagens que tenham pouca variação de níveis de cinza, resultarão em CC com pico marcante em torno do nível de cinza predominante. Por outro lado, imagens com inúmeras variações de níveis de cinza tenderão a apresentar uma curva mais suave com os valores distribuídos por toda a CC.

BAHEERATHAN *et al.* (1999) propõem o cálculo de cinco parâmetros para CC:

1 – Valor máximo de transições (mv) ocorridas na imagem:

$$mv = \max(C(\alpha)), 0 \leq \alpha \leq G - 1. \quad (\text{VI.19})$$

Este parâmetro tende a apresentar valores elevados para texturas homogêneas, devido à existência de um grande número transições para uma pequena variação de níveis de cinza.

2 – Valor médio de transições (av):

$$av = \frac{\sum_{\alpha=0}^{G-1} C(\alpha)}{G-1}. \quad (\text{VI.20})$$

Fornece o valor médio da distribuição das transições sobre o espectro de níveis de cinza. Valores elevados tendem a indicar a presença de inúmeras transições sobre todo o espectro, caracterizando texturas heterogêneas.

3 - Valor médio dos níveis de cinza ponderado pelo número de transições (sm):

$$sm = \frac{1}{\sum_{\alpha=0}^{G-1} C(\alpha)} \cdot \sum_{\alpha=0}^{G-1} \alpha \cdot C(\alpha). \quad (\text{VI.21})$$

Este parâmetro tende a privilegiar níveis de cinza para os quais ocorreu um maior número de transições. Imagens heterogêneas tendem a possuir inúmeras transições por todo o espectro resultando em valores elevados de sm .

4 - Desvio padrão dos níveis de cinza ponderado pelo número de transições (ssd):

$$ssd = \sqrt{\frac{1}{\sum_{\alpha=0}^{G-1} C(\alpha)} \cdot \sum_{\alpha=0}^{G-1} (\alpha - sm)^2 \cdot C(\alpha)}. \quad (VI.22)$$

Valores elevados de ssd tendem a indicar variação elevada entre os níveis de cinza presentes na imagem, privilegiando aqueles para os quais ocorre o maior número de transições. Este comportamento caracteriza texturas heterogêneas.

5 – Entropia do número de transições (ent):

$$ent = \sum_{\alpha=0}^{G-1} C(\alpha) \cdot \log_2[C(\alpha)], C(\alpha) > 0. \quad (VI.23)$$

Constitui-se em uma medida do grau aleatoriedade das transições ocorridas na imagem. Texturas heterogêneas tendem a resultar em valores elevados.

VI.1.7 Diferença Mínima Lateral

A *diferença mínima lateral* (*Minimum Side Difference - MSD*) é um parâmetro de textura específico para a avaliação do comportamento acústico da região posterior de um tumor numa imagem por US, sendo definida por HORSCH *et al.* (2002) como:

$$MSD = \min(\bar{g}_{posterior} - \bar{g}_{esquerda}, \bar{g}_{posterior} - \bar{g}_{direita}), \quad (VI.24)$$

e, portanto, como o próprio nome indica, é o mínimo das diferenças entre o valor médio dos níveis de cinza da região posterior ($\bar{g}_{posterior}$) ao tumor e das regiões adjacentes à direita ($\bar{g}_{direita}$) e à esquerda ($\bar{g}_{esquerda}$). Tais regiões são tomadas como retângulos, localizados na mesma profundidade e possuindo as mesmas dimensões da região de interesse (ROI) definida pelo radiologista (Figura VI.5). Na determinação da ROI, deve-se evitar a inclusão de sombras acústicas laterais (regiões delimitadas pelas linhas pontilhadas na Figura VI.5). Tumores malignos tendem a possuir uma sombra acústica na sua região posterior, esta característica produzirá valores negativos de MSD . Em contrapartida, o reforço acústico posterior, característica de tumores císticos, propiciará valores elevados.

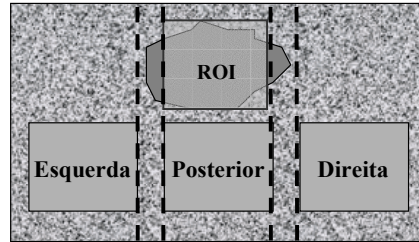


Figura VI.5. Representação das regiões utilizada para definir o parâmetro *MSD*.

VI.2 Aplicação de Parâmetros de Textura na Classificação de Tumores de Mama em Imagens por US

Entre as técnicas existentes, a mais citada na literatura para quantificar textura é a matriz de co-ocorrência (*Gray-Level Co-ocurrence Matrix* – GLCM – item VI.1.2) que é um histograma bi-dimensional ($N \times N$ níveis de cinza) que descreve a ocorrência de pares de *pixels* separados por uma distância determinada d , numa dada direção θ .

A partir da GLCM é possível determinar um conjunto de pelo menos 14 parâmetros (HARALICK *et al.*, 1973), porém apenas alguns se mostraram eficientes para classificar tumores de mama em imagens por US, com destaque para a *correlação*.

Segundo GARRA *et al.* (1993), os parâmetros que apresentam o melhor desempenho são o *contraste* (*con*) (Eq. VI.4) e a *correlação* (*cor*) (Eq. VI.5) da GLCM determinada com uma distância d de 4 *pixels*. Estes dois parâmetros combinados foram capazes de separar 78% dos fibroadenomas, 73% dos cistos e 91% dos tumores fibrocísticos, mantendo 100% da sensibilidade em detectar tumores malignos.

Em contra partida, para LEFEBVRE *et al.* (2000), os parâmetros que apresentam os melhores resultados são a *medida de informação da correlação* (IMC_2), *variância da diferença* (DV) e a *correlação* (*cor*) da GLCM. O melhor resultado obtido com estes parâmetros foi uma sensibilidade de 94,7% e uma especificidade de 80,4%.

A *correlação* da GLCM aparece novamente nos trabalhos de SIVARAMAKRISHNA *et al.* (2002) e HUBER *et al.* (2000), destacando-se como o parâmetro de textura que mais auxilia ao diagnóstico de tumores em imagens de mama por US.

Parâmetros calculados a partir do *Markoviano* e da técnica de *run-length* foram utilizados em conjunto para quantificar textura. GOLDBERG *et al.* (1992) citam o *segundo momento angular* (*asm*) (Eq. VI.2) e o *contraste inverso* (Eq. VI.14) do *Markoviano*, combinados com o *short run emphasis* (Eq. VI.8) calculado a partir do *run-length*, como a

combinação de parâmetros de melhor desempenho em seu estudo (considerando-se 100% de sensibilidade, obteve-se 40% de especificidade).

Estas mesmas técnicas foram utilizadas para identificar os tumores benignos em um grupo de tumores desconhecidos (ZHENG *et al.*, 1997). Vinte e oito parâmetros de textura foram calculados, considerando-se a técnica de *run-length* (7) e do *Markoviano* (21). O único parâmetro de textura que se destacou foi *medida da informação do Markoviano*.

Os coeficientes de autocovariância (Eq. VI.15) também foram utilizados como parâmetros para quantificar a textura de uma região de interesse (ROI) que contém o tumor (CHEN, D., *et al.*, 2000; CHANG *et al.*, 2003, KUO *et al.*, 2002).

CHEN D. *et al.* (2000) utilizaram 24 coeficientes de autocovariância para quantificar o seu grau de malignidade, obtendo uma sensibilidade de 97,6%, para uma especificidade de 79,5%. O uso deste parâmetro foi aprimorado (CHANG *et al.*, 2003) calculando-se os coeficientes de autocovariância apenas dos *pixels* que compunham o *speckle* da ROI sob análise, alcançando-se uma sensibilidade de 95,5% contra uma especificidade de 91,4%.

A autocorrelação aparece, no lugar da autocovariância, em alguns estudos (HORSCH *et al.*, 2002; DRUKKER *et al.*, 2002), porém sem apresentar grande importância discriminante.

Neste trabalho, parâmetros calculados a partir da curva de complexidade (CC) (BAHEERATHAN *et al.*, 1999) foram utilizados para separar imagens de US de regiões saudáveis da mama com predominância de tecido glandular, daquelas com predominância de tecido adiposo. A entropia da CC (*ent*) (Eq. VI.23) foi o parâmetro de maior capacidade discriminante, apresentando uma acurácia de 100%. Maiores detalhes podem ser obtidos em ALVARENGA *et al.* (2003a), publicação relacionada à Tese.

VI.3 Cálculo dos Parâmetros de Textura para Imagens de US de Mama

Os parâmetros de textura calculados a partir da GLCM (Eq. VI.1–5) e da CC (Eq. VI.19-23) foram avaliados quanto à sua capacidade de separar: regiões saudáveis da mama com predominância de tecido glandular ou adiposo, tumores quanto a sua ecotextura (homogênea ou heterogênea) e tumores quanto ao seu diagnóstico (benigno ou maligno). A GLCM foi escolhida por ser a técnica de quantificação de textura usada como referência na literatura, e a CC por não ter sido aplicada anteriormente em imagens médicas.

Para o cálculo da GLCM dos tumores, as imagens foram normalizadas para conter 32 níveis de cinza com o objetivo de aumentar o número de eventos na matriz. Este valor foi escolhido com base na literatura (DUTTA *et al.*, 1995). Cada matriz GLCM foi determinada usando-se a média de três matrizes, determinadas em ângulos θ específicos (0° , 45° e 90°) a uma distância d de três *pixels*. Todas as matrizes foram normalizadas pelo número de *pixels* presente na imagem. O cálculo de CC foi realizado nas imagens normalizadas, numa escala de 255 níveis de cinza.

A GLCM e a CC foram determinadas sobre toda a região de interesse (ROI) e apenas para a parte interna do tumor (Figura VI.6). Para identificar os parâmetros referentes à parte interna, acrescentou-se uma letra “i” subscrita. Os contornos que definem a região interna foram determinados usando o SAC.

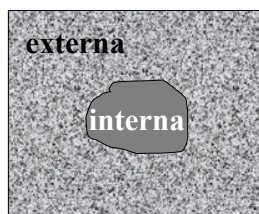


Figura VI.6. Diagrama esquemático de uma ROI, com o tumor em cinza, apontando as regiões contempladas pelos parâmetros de GLCM e CC.

Nota-se, nos exemplos apresentados nas Figuras VI.8 e VI.9, que a GLCM do tumor de ecotextura homogênea (Figuras VI.8c e VI.8d) tende a possuir um número menor de ocorrências de pares que a heterogênea (Figuras VI.9c e VI.9d). Isto se explica pelo fato do tumor heterogêneo possuir uma maior variação de níveis de cinza. Com relação à CC, observa-se que, para o tumor de ecotextura heterogênea, a curva se apresenta mais larga e com uma amplitude menor que a CC do tumor homogêneo. Este comportamento deve-se à distribuição do número de transições por uma faixa maior de níveis de cinza. A CC do tumor de ecotextura homogênea, mais estreita e de maior amplitude, tende a concentrar as transições dentro de uma faixa menor de níveis de cinza.

Para a avaliação inicial do desempenho dos parâmetros de textura, utilizando apenas regiões saudáveis da mama, foram calculados 10 parâmetros (5 GLCM e 5 CC), os quais foram aplicados individualmente à análise discriminante linear (LDA) e seu desempenho medido com base na área (A_z) sobre a curva ROC e no percentual de acertos na classificação.

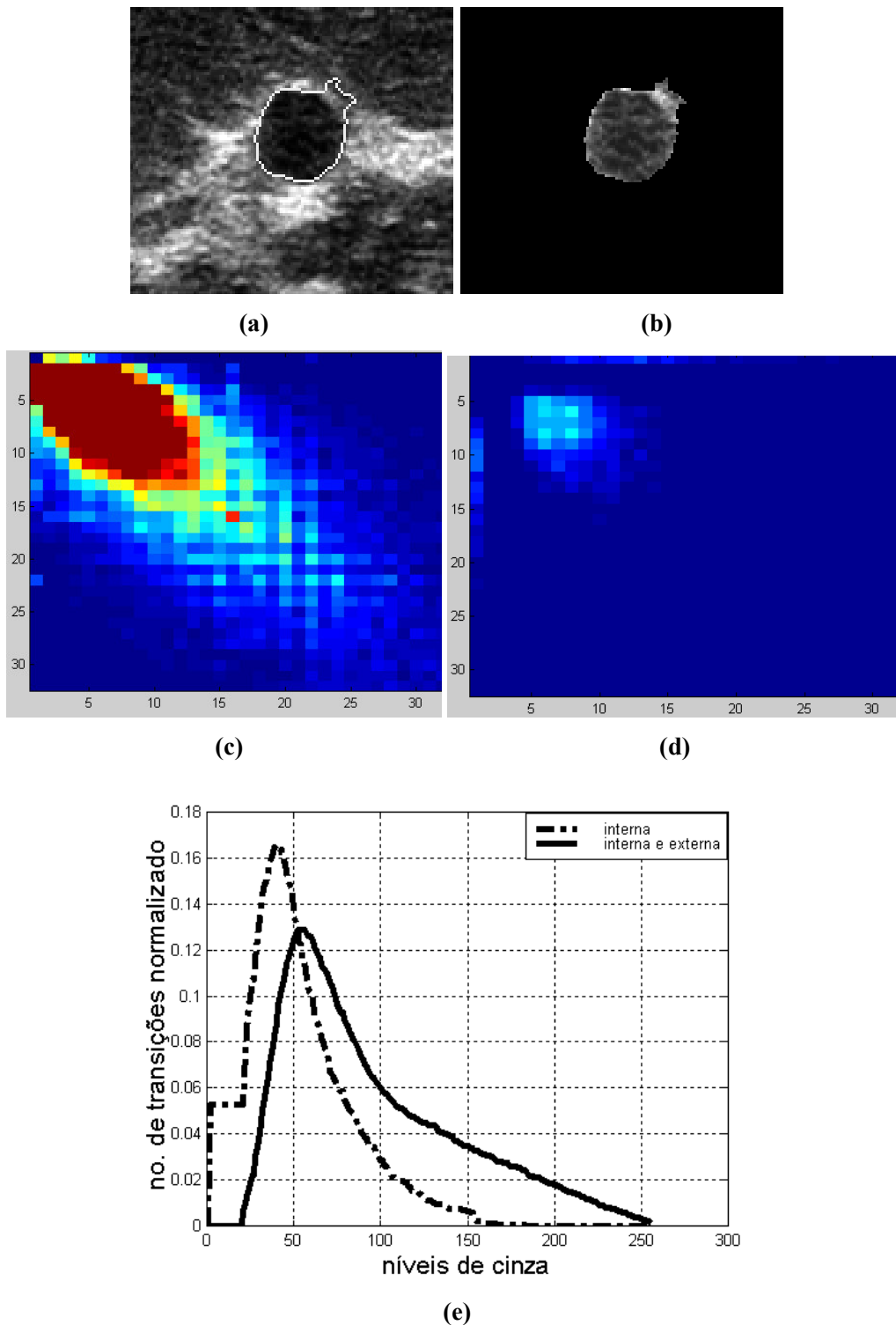


Figura VI.7. (a) ROI de um tumor, cuja (b) ecotextura interna é homogênea. Exemplos de GLCMs calculadas a partir da (c) ROI e da (d) região interna desta. Em (e) são mostradas as CCs.

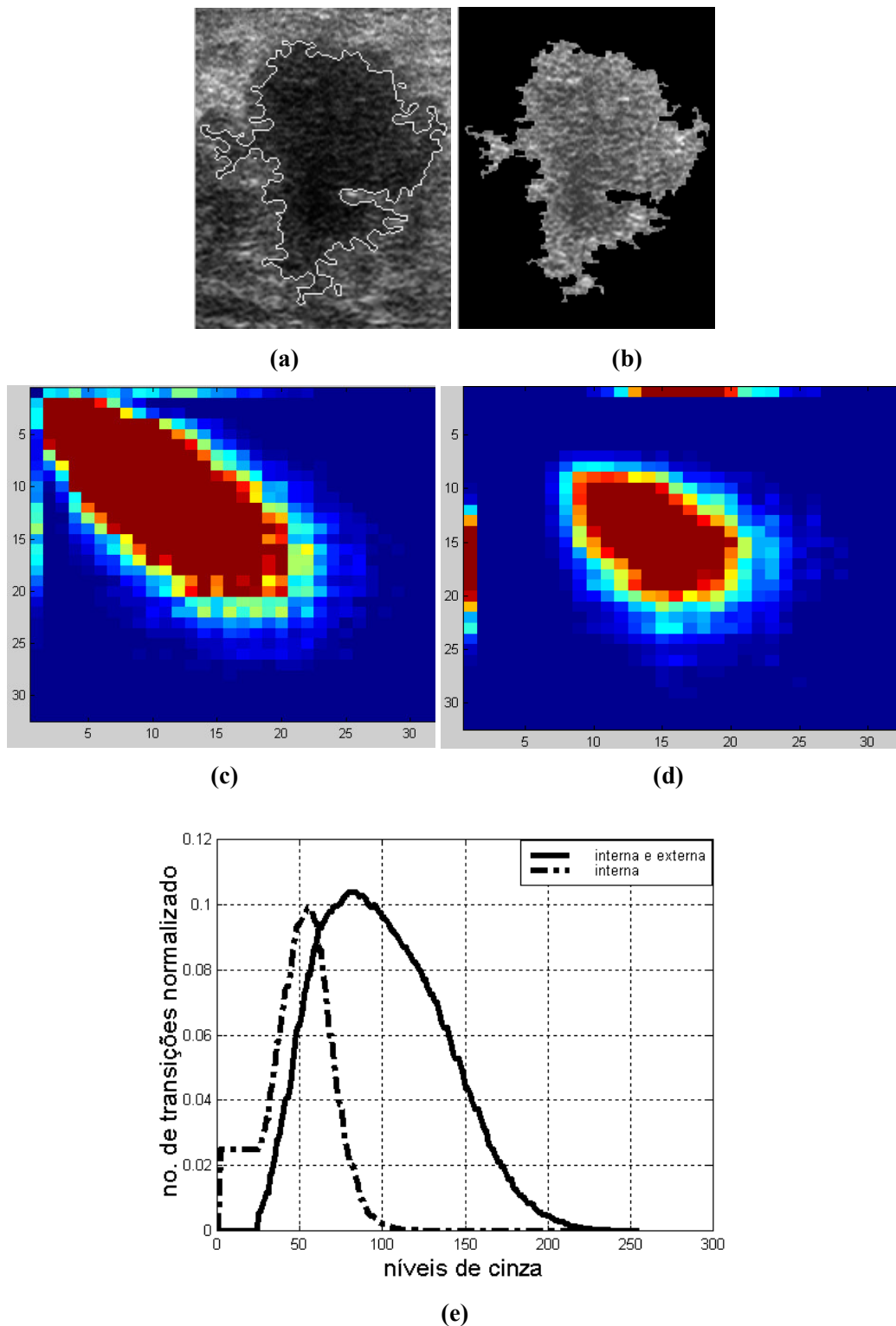


Figura VI.8. (a) ROI de um tumor, cuja (b) ecotextura interna é heterogênea. Exemplos de GLCMs calculadas a partir da (c) ROI e da (d) região interna desta. Em (e) são mostradas as CCs.

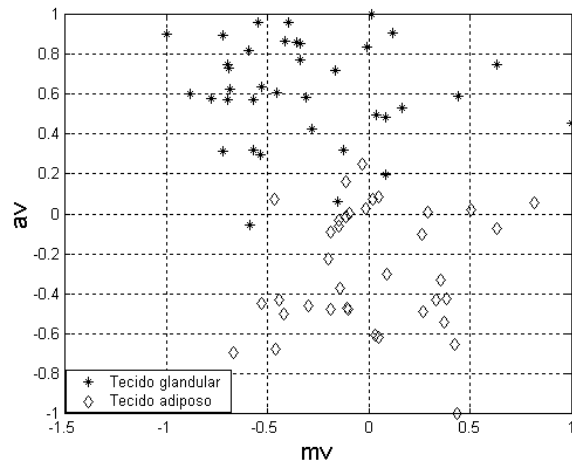
Na etapa de distinção dos tumores, a todas as combinações possíveis dos 20 parâmetros calculados (5 GLCM - ROI, 5 GLCM – interior, 5 CC - ROI e 5 CC - interior) aplicou-se a LDA, sendo então, o desempenho avaliado com base nos valores de A_z , acurácia (Ac), sensibilidade (S) e especificidade (E). Tais medidas de desempenho, quando da distinção de tumores quanto a ecotextura, foram calculadas considerando-se tumores heterogêneos como casos positivos e homogêneos como negativos. De maneira similar, tumores malignos foram considerados casos positivos, e benignos como negativos, na distinção quanto ao diagnóstico.

Para a aplicação da LDA foram utilizadas 152 imagens ($\approx 60\%$ das 246 segmentadas) de tumores de mama classificados pelo radiologista, com base no contorno, em 54 heterogêneos e 98 homogêneos. Quanto ao diagnóstico, 100 eram malignos e 52 benignos (*intranet* – seção IV.1). A utilização de somente parte das imagens teve como objetivo a seleção dos parâmetros morfométricos de melhor desempenho, de modo que o conjunto total de imagens pudessem ser utilizado na análise do desempenho combinando tais parâmetros aos de textura.

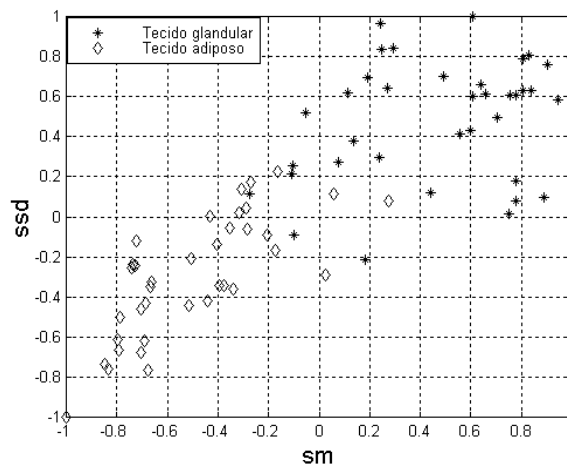
VI.4 Avaliando os Parâmetros de Textura na Distinção dos Tumores: Ecotextura e Diagnóstico

VI.4.1 Tecido adiposo x glandular

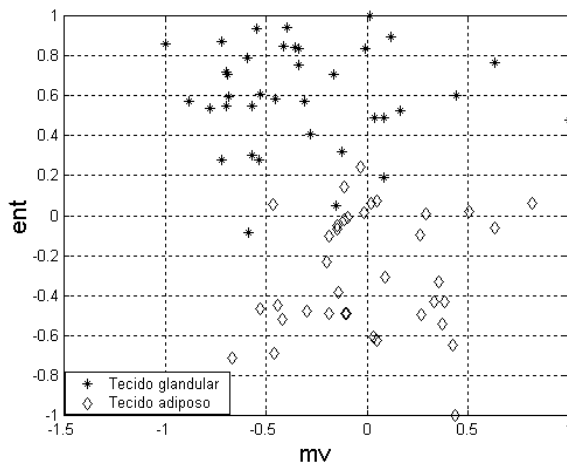
Ns diagramas de espalhamento dos parâmetros calculados a partir de CC nota-se que a entropia (*ent*) (Figura VI.9c) resultou na melhor separação entre regiões da mama com predominância de tecido glandular ou adiposo. Entre os calculados a partir da GLCM, o contraste (*con*) resultou na melhor separação (Figura VI.10b). De maneira geral, nota-se, ainda, que os parâmetros calculados a partir de CC apresentaram maior espalhamento intraclasse do que aqueles extraídos da GLCM.



(a)

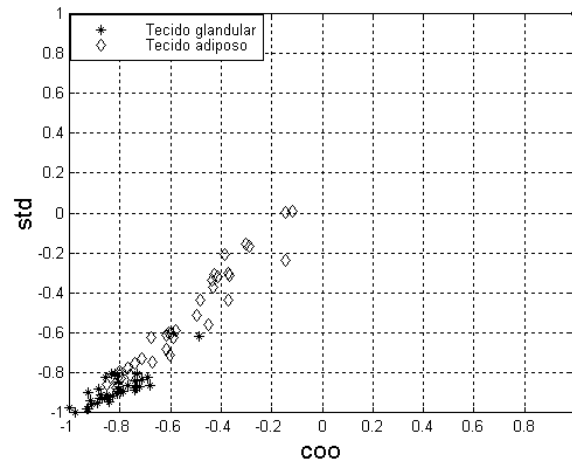


(b)

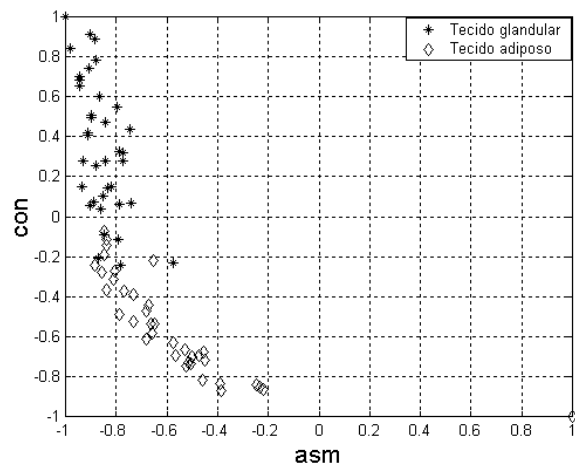


(c)

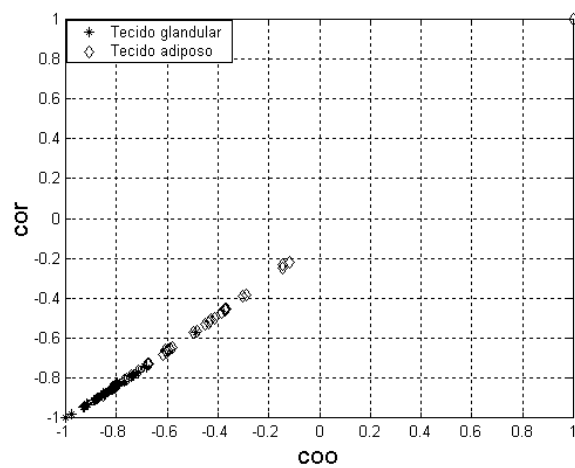
Figura VI.9. Diagrama de espalhamento dos parâmetros calculados a partir de CC: (a) $mv \times av$. (b) $sm \times ssd$. (c) $mv \times ent$.



(a)



(b)



(c)

Figura VI.10. Diagrama de espalhamento dos parâmetros calculados a partir de GLCM: (a) *std* x *coo*. (b) *con* x *asm*. (c) *cor* x *con*.

Entre os parâmetros calculados a partir da CC, aqueles que apresentaram maior acurácia foram *ent* e *av* (96,0%), enquanto *con* (93,2%) apresentou melhor desempenho entre os calculados a partir da GLCM (Tabela VI.1). Estes três parâmetros distinguiram corretamente todas amostras de regiões predominantes em tecido adiposo, enquanto *con* e *ent* identificaram, respectivamente, 86,5%, e 91,9% das amostras de regiões predominantes em tecido glandular.

Tabela VI.1. Desempenho individual de cada um dos 10 parâmetros de textura, ordenados por A_z , na distinção entre regiões saudáveis da mama com predominância de tecido glandular ou adiposo.

Parâmetros	A_z	Total de acertos (%)	Acertos Tecido Glandular (%)	Acertos Tecido Adiposo (%)
<i>ent</i>	0,99	96,0	91,9	100,0
<i>av</i>	0,99	96,0	91,9	100,0
<i>con</i>	0,99	93,2	86,5	100,0
<i>sm</i>	0,98	94,6	97,3	91,9
<i>std</i>	0,96	89,2	97,3	81,1
<i>ssd</i>	0,96	89,2	81,1	97,3
<i>cor</i>	0,90	83,8	97,3	70,3
<i>asm</i>	0,90	83,8	97,3	70,3
<i>coo</i>	0,89	83,8	97,3	70,3
<i>mv</i>	0,74	73,0	64,9	81,1

VI.4.2 Homogêneo x heterogêneo

O desempenho dos 20 parâmetros de textura, tomados separadamente, em distinguir os tumores em heterogêneos ou homogêneos variou em acurácia de 55,3% a 79,3%, enquanto os maiores valores de sensibilidade e especificidade foram de 86,5% e 86,7%, respectivamente. Para os parâmetros extraídos da GLCM, o maior valor de acurácia (79,3%) foi obtido por *con_i*, que apresentou especificidade de 83,7% e sensibilidade de 71,2%. Entre os calculados a partir da CC, os maiores valores de Ac (75,3%), S (71,2%) e E (77,6%) foram obtidos por *ent_i* e *av_i*, ambos determinados a partir da região interna do tumor.

Tabela VI.2. Desempenho individual de cada um dos 20 parâmetros de textura parâmetros de textura na separação entre tumores heterogêneos ou homogêneos, ordenados por A_z

Parâmetros	A_z	Ac (%)	S (%)	E (%)
<i>con_i</i>	0,82	79,3	71,2	83,7
<i>con</i>	0,80	77,3	67,3	82,7
<i>ent_i</i>	0,78	75,3	71,2	77,6
<i>av_i</i>	0,78	75,3	71,2	77,6
<i>ent</i>	0,74	70,7	73,1	69,4
<i>av</i>	0,74	70,7	73,1	69,4
<i>coo</i>	0,74	68,0	73,1	65,3
<i>cor</i>	0,73	70,0	63,5	73,5
<i>asm</i>	0,73	70,0	63,5	73,5
<i>mv_i</i>	0,73	68,0	67,3	68,4
<i>std</i>	0,73	67,3	71,2	65,3
<i>sm</i>	0,68	64,7	65,4	64,3
<i>ssd_i</i>	0,66	59,3	86,5	44,9
<i>sm_i</i>	0,60	58,7	75,0	50,0
<i>mv</i>	0,59	62,0	73,1	56,1
<i>ssd</i>	0,58	55,3	69,2	48,0
<i>cor_i</i>	0,53	67,3	30,8	86,7
<i>asm_i</i>	0,53	67,3	30,8	86,7
<i>std_i</i>	0,53	64,7	32,7	81,6
<i>coo_i</i>	0,52	64,7	28,8	83,7

Usando-se pares de parâmetros, o melhor desempenho foi alcançado por *con_i* e *mv_i* (Tabela VI.3), que resultou em aumentos significativos nos valores de acurácia (84,0%) e especificidade (94,9%), em detrimento de redução significativa da sensibilidade (63,5%). No diagrama de espalhamento de *mv_i* x *con_i*, mostrado na Figura VI.11, nota-se que os tumores heterogêneos tendem a estar mais agrupados (*con_i*: $\mu = 0,59$, $\sigma = 0,28$; *mv_i*: $\mu = -0,32$, $\sigma = 0,34$) que os homogêneos (*con*: $\mu = 0,21$, $\sigma = 0,35$; *mv_i*: $\mu = -0,02$, $\sigma = 0,37$).

Aumentando-se o número de parâmetros tomados em conjunto para três, não se obteve melhora no desempenho. Considerando-se quatro parâmetros, o melhor desempenho

foi obtido com o conjunto con_i , mv_i , coo e std , que resultou em aumento nos valores de Ac para 87,3% e de S para 75,0%, embora E tenha sido reduzida (93,9%).

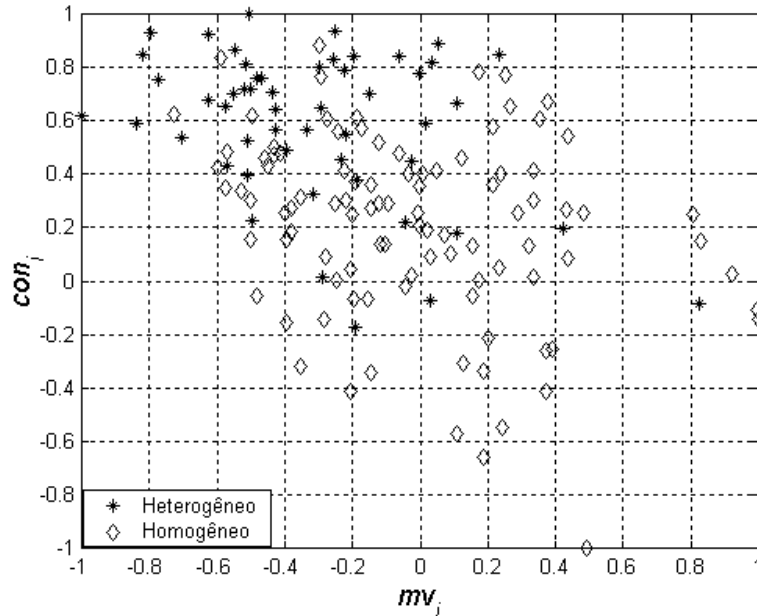


Figura VI.11. Diagrama de espalhamento dos parâmetros mv_i e con_i .

Tomando-se cinco parâmetros, o melhor desempenho foi obtido pelo conjunto con_i , coo , std , ent_i e av , que resultou, similarmente ao conjunto de quatro parâmetros, em aumento nos valores de Ac (88,0%) e S (76,9%), embora a especificidade (93,9%) não tenha sido alterada. O parâmetro mv_i não fez parte deste conjunto, apesar de contribuir para distinção nos apresentados anteriormente. A entropia (ent_i), melhor desempenho individual entre os parâmetros calculados a partir de CC, parece substituir mv_i , pois as dispersões de ent_i (Figura VI.12) são menores, tanto para tumores homogêneos ($\mu = -0,02$, $\sigma = 0,37$) quanto para heterogêneos ($\mu = -0,32$, $\sigma = 0,34$). Além disso, no diagrama de espalhamento de ent_i x av (Figura VI.12), nota-se que a dispersão de ent_i é maior nas amostras homogêneas ($\mu = 0,35$, $\sigma = 0,33$) do que nas heterogêneas ($\mu = 0,63$, $\sigma = 0,20$). Aumentando-se o número de parâmetros tomados em conjunto para seis, não se obteve melhora no desempenho.

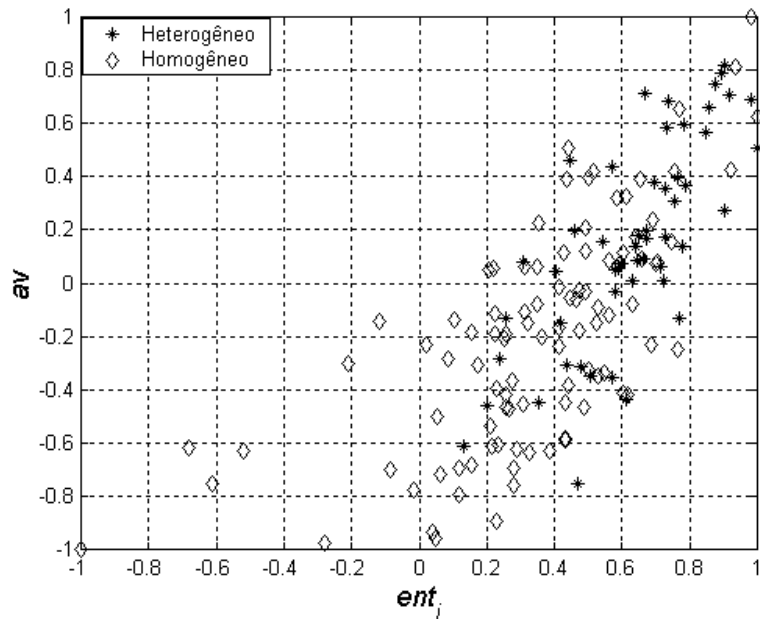


Figura VI.12. Diagrama de espalhamento dos parâmetros av e ent_i .

O maior valor de especificidade (98,0%) na distinção de tumores quanto a ecotextura obteve-se com o conjunto de sete parâmetros formado por con_i , coo , std , ent_i , sm , ssd e con , que resultou, também, em pequeno aumento da Ac (88,7%), embora com redução significativa da S (71,1%).

Conjuntos de oito e nove parâmetros não implicaram em melhora no desempenho. Todavia, este tornou a melhorar com o conjunto de dez parâmetros formado por con_i , mv_i , coo , std , av , ent , sm_i , ssd , asm , coo_i , que resultou em aumento de Ac para 89,3% e S para 75%, enquanto a especificidade reduziu (96,9%).

O máximo desempenho na distinção de tumores de mama, quanto à textura, entre heterogêneos e homogêneos ($Ac = 90,0\%$, $S = 76,9\%$ e $E = 96,9\%$) foi obtido com onze (con_i , mv_i , coo , std , av , ent , sm_i , ssd , asm , con , mv) dos vinte parâmetros. Deve-se, entretanto, notar que o desempenho obtido com apenas cinco parâmetros (Tabela VI.3) forneceu o mesmo valor de sensibilidade (76,9%), com pequenos decréscimos nos valores de Ac (88,0%) e E (93,9%). As curvas ROC para estes conjuntos de parâmetros, mostradas na Figura VI.13, evidenciam um comportamento similar, particularmente na região onde a sensibilidade é 76,9% e a especificidade varia de 86,7% a 91,8% (setas na Figura VI.13). Nesta região as curvas são coincidentes, o que justificaria o uso de cinco parâmetros ao

invés de onze. Entretanto, a partir da curva de onze parâmetros, pode-se alcançar maiores valores de sensibilidade como, por exemplo, de 80,8%, com E de 80,6% e Ac de 80,7%.

Tabela VI.3. Desempenho das melhores combinações de parâmetros na distinção entre tumores de mama de textura heterogênea ou homogênea

Parâmetros	A_z	Ac (%)	S (%)	E (%)
con_i	0,82	79,3	71,2	83,7
con_i, mv_i	0,84	84,0	63,5	94,9
con_i, mv_i, coo, std	0,86	87,3	75,0	93,9
$con_i, coo, std, ent_i, av$	0,86	88,0	76,9	93,9
$con_i, coo, std, ent_i, sm, ssd, con$	0,88	88,7	71,1	98,0
$con_i, mv_i, coo, std, ent, av, sm_i, ssd, asm, coo_i$	0,89	89,3	75,0	96,9
Onze parâmetros	0,89	90,0	76,9	96,9

11 parâmetros: $con_i, mv_i, coo, std, ent, av, sm_i, ssd, asm, con, mv$.

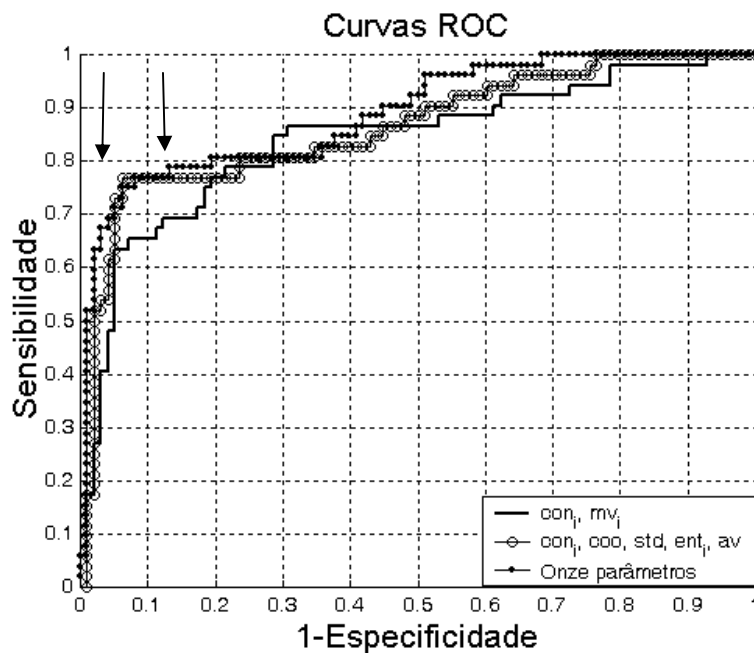


Figura VI.13. Curvas ROC do desempenho obtido considerando-se dois, cinco e onze parâmetros de textura na separação dos tumores quanto a ecotextura. As setas indicam a região (especificidade entre 86,7% e 91,8%) onde os parâmetros apresentam desempenho similar.

VI.4.3 Maligno x benigno

O desempenho dos 20 parâmetros de textura, tomados separadamente, em distinguir os tumores em heterogêneos ou homogêneos variou, em acurácia, de 55,3% a 79,3%, enquanto os maiores valores de sensibilidade e especificidade foram de 86,5% e 86,7%, respectivamente (Tabela VI.4). Para os parâmetros calculados a partir de CC, o maior valor de A_z (0,82) foi obtido por mv_i , que também resultou nos maiores valores de sensibilidade (87,0%) e acurácia (79,6%). Entre aqueles extraídos da GLCM, o maior valor de A_z (0,72) foi obtido usando-se con , que resultou, também, no maior valor de especificidade (86,5%).

Tabela VI.4. Desempenho individual de cada um dos vinte parâmetros de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z

Parâmetros	A_z	Ac (%)	S (%)	E (%)
mv_i	0,82	79,6	87,0	65,4
ent_i	0,75	70,4	70,0	71,2
av_i	0,75	70,4	71,0	69,2
con	0,72	64,5	53,0	86,5
av	0,69	67,1	66,0	69,2
ent	0,69	64,5	58,0	76,9
coo_i	0,66	66,5	66,0	67,3
coo	0,65	69,7	83,0	44,2
std	0,64	70,4	85,0	42,3
asm	0,64	65,8	73,0	51,9
cor	0,64	65,8	73,0	51,9
asm_i	0,64	64,5	62,0	69,2
cor_i	0,64	64,5	62,0	69,2
sm	0,64	59,2	54,0	69,2
con_i	0,64	59,2	48,0	80,8
ssd	0,62	62,5	59,0	69,2
std_i	0,62	61,8	62,0	61,5
ssd_i	0,61	67,8	77,0	50,0
sm_i	0,59	63,8	71,0	50,0
mv	0,51	56,5	58,0	53,9

Similarmente ao obtido na distinção quanto a ecotextura, o par de parâmetros que resultou em melhor desempenho foi mv_i e con (Tabela VI.5), que alcançou valores de acurácia (80,3%), sensibilidade (80,0%) e especificidade (80,8%) superiores a 80,0%. O diagrama de espalhamento de mv_i x con , mostrado na Figura VI.14, evidencia que tumores malignos tendem a apresentar valores de contraste (con) maiores ($\mu = 0,27$) que os benignos ($\mu = -0,17$), enquanto mv_i apresenta comportamento oposto (benignos: $\mu = 0,05$; malignos: $\mu = -0,25$). Ainda no diagrama, tumores malignos parecem estar mais agrupados (con : $\sigma = 0,29$; mv_i : $\sigma = 0,39$) do que os benignos (con : $\sigma = 0,38$; mv_i : $\sigma = 0,37$).

Tabela VI.5. Desempenho das melhores combinações de parâmetros de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z

Parâmetros	A_z	Ac (%)	S (%)	E (%)
mv_b, con	0,84	80,3	80,0	80,8
mv_b, con, ssd	0,84	80,9	79,0	84,6
mv_b, ssd, coo, cor	0,85	83,6	83,0	84,6
mv_b, con, ssd, coo, cor	0,85	84,9	85,0	84,6
$mv_b, con, coo, std, con_i, mv$	0,89	86,2	88,0	82,7
$mv_b, con, ssd, coo, std, con_i, mv, sm_i$	0,89	86,8	88,0	84,6
$mv_b, con, ssd, coo, std, con_i, cor, coo_i, ent$	0,89	87,5	87,0	88,5
Doze parâmetros	0,90	88,2	89,0	86,5

Doze parâmetros: $mv_b, con, ssd, coo, std, cor, con_i, ent, sm, ssd_i, std_i, cor_i$.

Tomando-se combinações de três parâmetros, o melhor desempenho foi obtido com o conjunto mv_i, con e ssd , que resultou em aumento nos valores de acurácia (80,9%) e especificidade (84,6%), contudo, a sensibilidade foi reduzida para 79,0%. Os conjuntos de melhor desempenho formados por quatro (mv_b, ssd, coo e cor) e cinco (mv_b, con, ssd, coo e cor) parâmetros, resultaram em aumento nos valores de acurácia e sensibilidade, embora a especificidade tenha se mantido inalterada (Tabela VI.5).

Considerando-se seis parâmetros, o melhor desempenho foi alcançado pelo conjunto $mv_b, con, coo, std, con_i, mv$, com o qual a especificidade (82,7%) foi reduzida, todavia com aumento significativo dos valores de Ac (86,2%) e S (88,0%). Aumentando-se para sete o número de parâmetros, não se obteve melhora no desempenho.

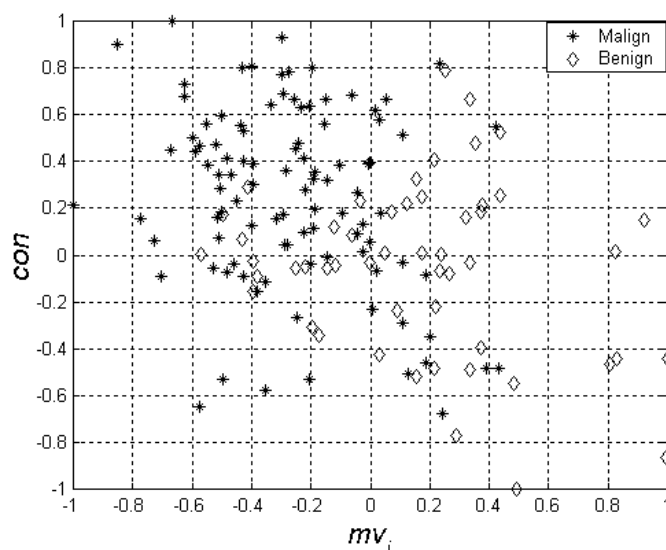


Figura VI.14. Diagrama de espalhamento entre os parâmetros mv_i e con . Nota-se que os tumores malignos tendem a apresentar valores elevados de contraste (con), enquanto que os valores de mv_i tendem a ser elevados em tumores benignos.

Assumindo conjuntos de oito parâmetros, o melhor desempenho foi alcançado por mv_i , con , ssd , coo , std , con_i , mv , sm_i , que resultou em aumento nos valores de Ac para 86,8% e E para 84,6%, enquanto a sensibilidade manteve-se inalterada 88,0%. O melhor conjunto de nove parâmetros (mv_i , con , ssd , coo , std , con_i , cor , coo_i , ent), resultou em redução no valor da sensibilidade (87,0%), com aumento, entretanto, da Ac (87,5%) e da E (88,5%).

Por fim, o desempenho máximo em distinguir tumores de mama entre malignos ou benignos, com base em parâmetros de textura, foi obtido tomando-se doze dos vinte parâmetros, e resultou em aumento da para Ac 88,2% e da S para 89,0%. Neste caso, os parâmetros mv , av , av_i , sm_i , ent_i , asm , coo_i , asm_i foram descartados. Nas curvas ROC dos conjuntos de nove e doze parâmetros de melhor desempenho, mostradas na Figura VI.15, nota-se que, na região onde a sensibilidade varia de 82,0 a 87,0% (setas na Figura VI.15), o conjunto de nove parâmetros apresenta uma especificidade maior (88,5%) que o de doze (86,5%). Entretanto, com este último é possível obter um valor de sensibilidade maior (89,0%) para a mesma especificidade (86,5%).

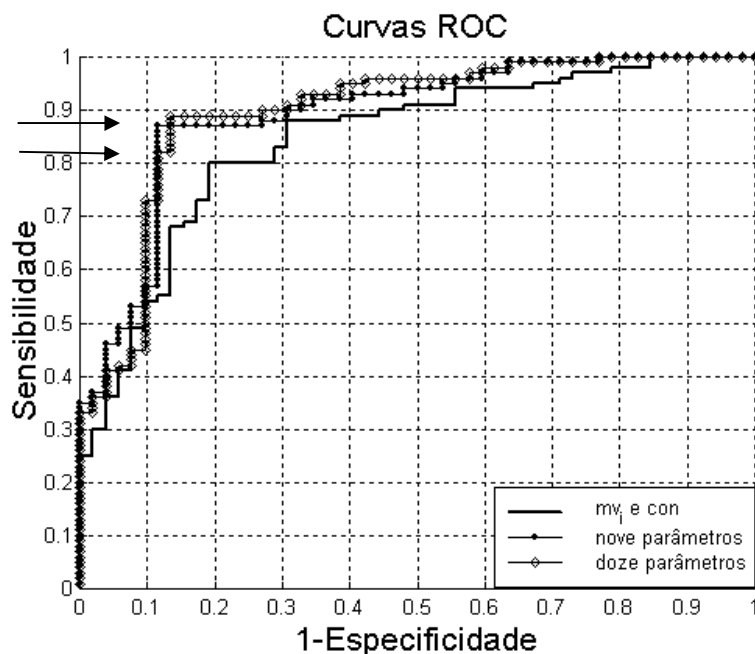


Figura VI.15. Curvas ROC do desempenho obtido considerando-se dois, nove e doze parâmetros de textura na separação dos tumores quanto ao diagnóstico. Nota-se que na região onde a sensibilidade varia de 82,0 a 87,0% (setas), o conjunto de nove parâmetros apresenta uma especificidade maior (88,5%) que o de 12 parâmetros (86,5%).

O melhor desempenho usando-se apenas parâmetros calculados a partir da CC (mv_i , ent , mv , sm) foi similar ao obtido com aqueles extraídos da GLCM (con , std , coo , coo_i , std_i and con_i), com valores de acurácia, sensibilidade e especificidade superiores a 80,0% (Tabela VI.6). A entropia (ent) da CC proporciona o maior valor de especificidade (76,9%), e o diagrama de espalhamento entre mv_i x ent (Figura VI.16) evidencia que tumores malignos tendem a apresentar valores de ent ($\mu = 0,03$) maiores que os benignos ($\mu = -0,25$). Entretanto, o desvio padrão (std) da GLCM, que resultou na sensibilidade (85,0%) (Tabela VI.6), tende a apresentar valores menores para tumores malignos, os quais, também, apresentam-se mais agrupados ($\mu = -0,61$; $\sigma = 0,37$) que os benignos ($\mu = -0,56$; $\sigma = 0,47$), conforme ilustrado na Figura VI.17.

Por fim, o desempenho obtido exclusivamente com parâmetros calculados sobre a ROI (mv e con) foi inferior ao alcançado por aqueles calculados sobre a região interna do tumor (mv_i , con_i , sm_i , ent_i , std_i) (Tabela VI.7).

Tabela VI.6. Desempenho, na distinção quanto ao diagnóstico, das melhores combinações de parâmetros de textura obtido usando-se, separadamente, parâmetros calculados a partir de CC e GLCM

Parâmetros	CC	GLCM
	mv_i, ent, mv, sm	$con, std, coo, con_i, std_i, coo_i$
Ac (%)	81,6	81,6
S (%)	82,0	80,0
E (%)	80,8	84,6
A_z	0,84	0,86

Tabela VI.7. Desempenho, na distinção quanto ao diagnóstico, das melhores combinações de parâmetros de textura obtido usando-se, separadamente, parâmetros calculados a partir da ROI e da região interna do tumor

Parâmetros	ROI	Região interna do tumor
	mv e com	$mv_i, con_i, sm_i, ent_i, std_i$
Ac (%)	72,4	84,9
S (%)	77,0	93,0
E (%)	63,5	69,2
A_z	0,73	0,60

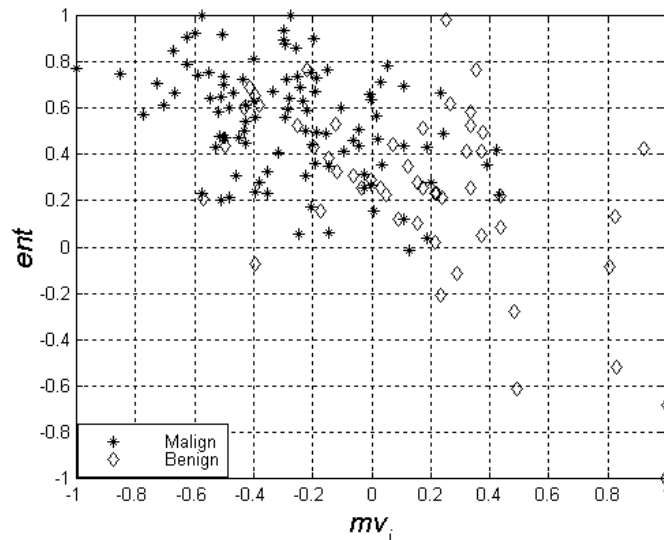


Figura VI.16. Diagrama de espalhamento entre os parâmetros mv_i e ent . Nota-se que os tumores malignos tendem a apresentar valores de ent mais elevados que os tumores benignos.

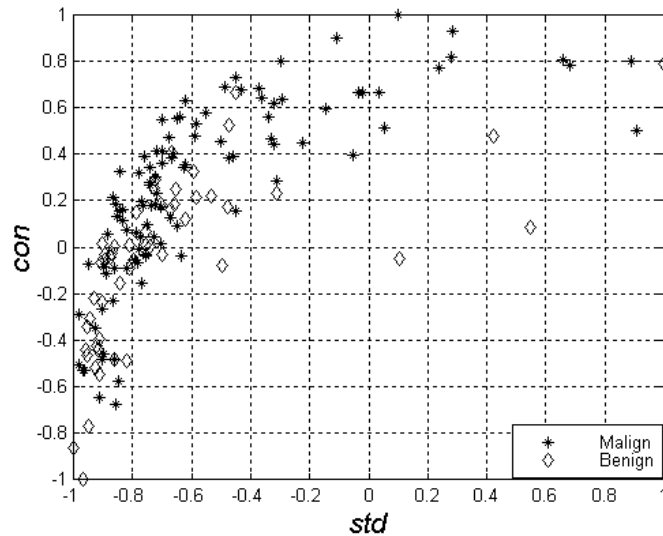


Figura VI.17. Diagrama de espalhamento entre os parâmetros *con* e *std*. Nota-se que os tumores malignos tendem a apresentar valores de *std* menores, além de estarem mais agrupados que os benignos.

VI.5 Discussão

A análise dos parâmetros de textura indicou o contraste (con_i e con) da GLCM e a entropia (ent e ent_i) de CC como os dois parâmetros de melhor desempenho individual na distinção dos tumores quanto a ecotextura (heterogênea ou homogênea). A variabilidade local da textura da imagem de US presente nos tumores heterogêneos e no tecido glandular resultou em valores elevados de con (Figura VI.10b) e con_i (Figura VI.11), enquanto tumores homogêneos e regiões predominantes em tecido adiposo, que tendem a apresentar uma menor variabilidade na textura, resultaram em valores menores. Comportamento similar foi observado nos parâmetros ent (Figura VI.10c) e ent_i (Figura VI.12), porém, estes estão relacionados à aleatoriedade espacial dos níveis de cinza presentes na textura.

Os resultados na distinção entre os tumores heterogêneos ou homogêneos obtidos, com os parâmetros tomados em conjunto, indicam quatro deles como base para a maioria dos conjuntos: con_i , mv_i , coo , std . Todavia, apesar de apresentar o terceiro melhor desempenho individual, a entropia da região interna do tumor (ent_i), faz parte apenas de conjuntos nos quais mv_i está ausente (Tabela VI.4). Este último, calculado a partir da CC, ao contrário do contraste (con e con_i) e da entropia (coo e coo_i) da GLCM, apresenta valores elevados para tumores homogêneos e baixos para heterogêneos (Figura VI.11). Este

comportamento sugere a potencialidade de mv_i para caracterizar a homogeneidade da ecotextura interna do tumor. A capacidade de mv em quantificar a homogeneidade em texturas é citada por BAHEERATHAN *et al.* (1999).

Na distinção entre tumores malignos ou benignos os conjuntos de parâmetros de melhor desempenho são formados, pelos quatro parâmetros citados anteriormente, além de con , mv e cor (Tabela VI.5). Estes três últimos fornecem informação sobre a variabilidade dos níveis de cinza em toda a ROI, sugerindo que as características do tecido adjacente ao tumor pode contribuir para a melhoria do desempenho na distinção dos tumores entre malignos ou benignos. A habilidade do contraste (con) e da correlação (cor) da GLCM, calculada sobre toda a ROI, em distinguir tumores de mama entre malignos ou benignos em imagens por US é citada por GARRA *et al.* (1993) e HUBER *et al.* (2000). Este último, reporta ainda a importância da análise da arquitetura do tecido adjacente ao tumor para o diagnóstico.

Neste trabalho, o desempenho alcançado usando-se apenas parâmetros calculados sobre a região interna do tumor, foi superior ao obtido com parâmetros calculados sobre toda a ROI (Tabela VI.7). Este resultado sugere que a ecotextura interna do tumor pode auxiliar no diagnóstico, apesar de LEFEBVRE *et al.* (2000) reportarem que a informação de ecotextura do centro do tumor, quantificada usando a GLCM e a *run-length analysis*, não auxilia no diagnóstico.

Por fim, o melhor desempenho ($Ac = 88,2\%$, $S = 89,0\%$ e $E = 86,5\%$ - Tabela VI.5) resultou da combinando dos parâmetros calculados a partir da região interna do tumor e de toda a ROI, sugerindo que ambas informações são relevantes para o diagnóstico.

Os desempenhos alcançados usando-se apenas parâmetros calculados a partir da GLCM ou da CC, em termos de acurácia, sensibilidade e especificidade foram bastante similares. Entretanto, o conjunto de parâmetros formado pelos calculados a partir de CC (4) foi menor do que o da GLCM (6) (Tabela VI.6). Entre os parâmetros calculados a partir de CC, ent , que apresentou elevada especificidade (Tabela VI.4), parece substituir o contraste da GLCM na tarefa de refletir a heterogeneidade dos tumores.

VII PARÂMETROS MORFOMÉTRICOS E DE TEXTURA: DISTINÇÃO DE TUMORES QUANTO AO DIAGNÓSTICO

Nesta etapa, os parâmetros morfométricos (7) e os de textura (20) foram combinados com o objetivo de separar os tumores entre malignos ou benignos. Novamente, todas as combinações possíveis de parâmetros foram apresentadas à análise discriminante (LDA) e seus desempenhos foram avaliados com base nos valores da área (A_z) sobre a curva ROC, acurácia, sensibilidade e especificidade.

Para a aplicação da LDA foram utilizados 152 tumores de mama ($\approx 60\%$ dos 246 segmentados), identificados, quanto ao diagnóstico, como 100 malignos e 52 benignos (*intranet* – seção IV.1). Após a seleção do melhor conjunto de parâmetros, todos os 246 tumores foram submetidos a LDA obtida, anteriormente, com os 152 tumores. Esta etapa teve o objetivo de avaliar a capacidade de generalização da LDA e dos parâmetros selecionados na distinção dos tumores.

VII.1 Aplicação dos Parâmetros Morfométricos e de Textura Tomados em Conjunto na Classificação de Tumores de Mama em Imagens por US

Como visto anteriormente, existem vários estudos que empregam parâmetros morfométricos ou de textura na classificação de tumores de mama em imagens por US, porém, poucos estudos relatam seu uso tomado em conjunto. LEFEBVRE *et al* (2000) ao empregar a LDA à razão entre a largura e a altura (*DWR*) do tumor em conjunto com 14 parâmetros calculados a partir da GLCM e 5 da técnica de *run-length*, observou que o conjunto de melhor desempenho ($A_z = 0,88$, $S = 94,7\%$ e $E = 80,4\%$) era formado apenas por 3 parâmetros, todos de textura, calculados a partir da GLCM (*medida de informação da correlação*, *variância da diferença* e a *correlação*). Contudo, *DWR* mostrou-se significativa ($p < 0,003$) na distinção dos tumores, apesar de não fazer parte deste conjunto. Além disso, HORSCH *et al.* (2002) reportaram que a aplicação da LDA a dois parâmetros morfométricos, entre eles *DWR*, (*gradiente radial normalizado* – *NRG* e *DWR*), e dois de textura (*autocovariância* da ROI e *diferença mínima lateral* - *MSD*), na distinção de tumores de mama em malignos ou benignos, resultou em $A_z = 0,92$.

VII.2 Avaliando Parâmetros Morfométricos e de Textura Tomados em Conjunto

O par de parâmetros que melhor separa os tumores, em malignos ou benignos, é formado por nrv e std_i (Tabela VII.1), com valores de acurácia (89,5%) e especificidade (92,3%) da ordem de 90%, porém com uma sensibilidade inferior (88,0%). No diagrama de espalhamento mostrado na Figura VII.1, nota-se que, apesar do fraco desempenho individual de std_i (Tabela VI.4), quando combinado com nrv (parâmetro morfométrico de melhor desempenho individual - Tabela V.3), apresenta um desempenho superior aos obtidos usando-se apenas parâmetros morfométricos (Tabela V.4) ou de textura (Tabela VI.5). A reta ($y = -3,3x + 0,83$) que melhor separa os tumores malignos dos benignos é também mostrada na Figura VII.1.

Considerando-se três parâmetros, o melhor desempenho foi obtido com o conjunto nrv , R e std , que resultou em aumento da Ac para 90,1% e da S para 96,2%, embora com pequeno decréscimo da sensibilidade (87,0%). No diagrama de espalhamento dos três parâmetros (Figura VII.2), mostra-se o plano que melhor separa os tumores ($z = -3,93x - 0,92y + 0,91$).

Conjuntos de quatro e cinco parâmetros não implicaram em melhoria de desempenho. Entretanto, ao se tomar o conjunto formado por nrv , R , RS , D_{NRL} , ent , sm , obteve-se de o melhor resultado entre os demais conjuntos de seis parâmetros, ou seja, Ac (91,5%), S (90,0%) e E (94,2%) superiores a 90,0%. Aumentando-se para sete o número parâmetros tomados em conjunto, não se obteve melhora no desempenho. Considerando-se o melhor conjunto de oito parâmetros, formado por nrv , R , RS , D_{NRL} , C , ent , mv_i , coo , o desempenho melhora com aumento da Ac (92,1%) e da S (92,0%), porém com redução da especificidade para 92,3%.

O desempenho máximo foi alcançado com o conjunto nrv , R , RS , D_{NRL} , C , std_i , ent , ent_i , sm , coo , cor , que resultou em aumento nos valores de acurácia (92,8%) e especificidade (94,2%), além da sensibilidade não ter sido alterada. Nas curvas ROC para os conjuntos de oito e onze parâmetros, mostradas na Figura VII.3, nota-se que, na região onde a especificidade varia de 75,0% a 92,3% (setas), as curvas são coincidentes, o que justificaria o uso de oito parâmetros ao invés de onze.

Tabela VII.1. Desempenho das melhores combinações de parâmetros morfométricos e de textura na separação dos tumores em malignos ou benignos, ordenados em função do valor de A_z

Parâmetros	A_z	Ac (%)	S (%)	E (%)
nrv e std_i	0,91	89,5	88,0	92,3
nrv, R, std	0,92	90,1	87,0	96,2
$nrv, R, RS, D_{NRL}, ent, sm$	0,93	91,5	90,0	94,2
$nrv, R, RS, D_{NRL}, C, ent, mv_i, coo$	0,93	92,1	92,0	92,3
Onze parâmetros	0,93	92,8	92,0	94,2

Onze parâmetros: $nrv, R, RS, D_{NRL}, C, std_i, ent, ssd_i, sm, coo, cor$.

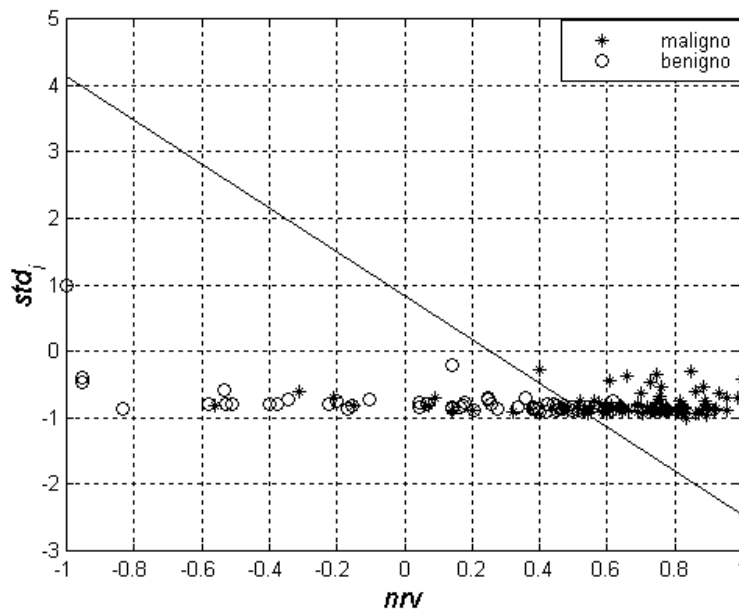


Figura VII.1. Diagrama de espalhamento de nrv x std_i e o plano ($y = -3,3x + 0,83$) que melhor separa os tumores malignos dos benignos.

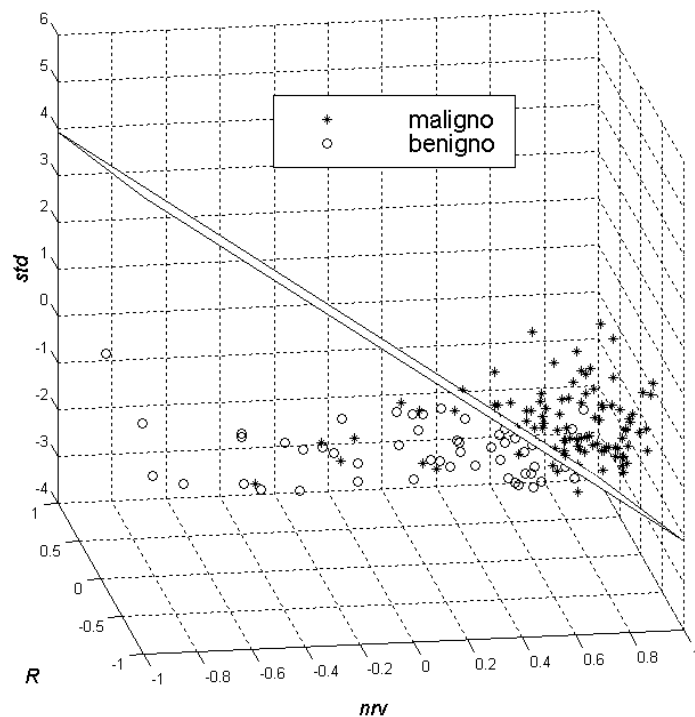


Figura VII.2. Diagrama de espalhamento de $nrv \times R \times std$ e o plano ($z = -3,93x - 0,92y + 0,91$) que melhor separa os tumores malignos dos benignos.

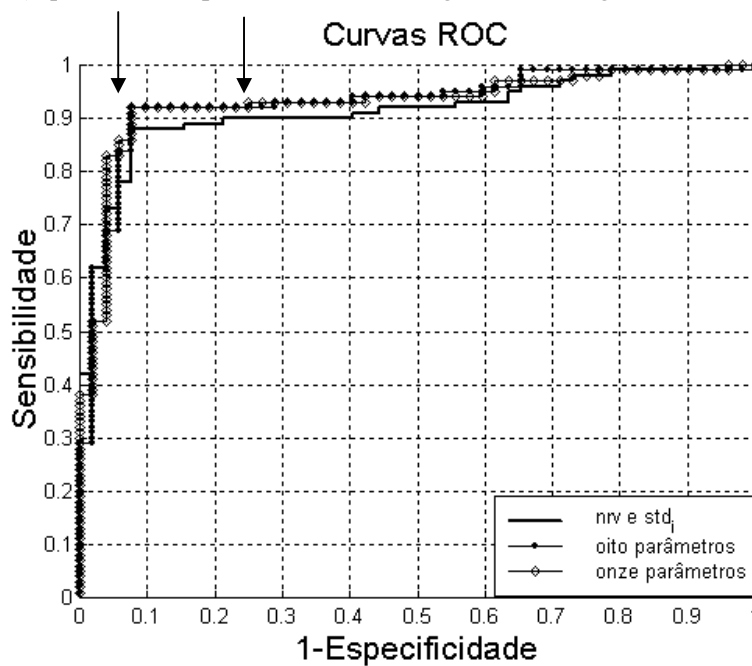


Figura VII.3. Curvas ROC do desempenho obtido considerando-se dois, oito e onze parâmetros de textura e morfométricos combinados na separação dos tumores quanto ao diagnóstico. Nota-se que na região onde a especificidade varia de 75,0% a 92,3% (setas) os conjuntos de 8 e 11 parâmetros mostram um comportamento idêntico.

A aplicação de todos os 246 tumores a LDA obtida com os 152 tumores, considerando-se o conjunto de parâmetros de melhor desempenho (*onze parâmetros* – Tabela VII.1), resulta em redução dos valores de acurácia ($\approx 5\%$), sensibilidade ($\approx 3\%$) e especificidade ($\approx 7\%$) (Tabela VII.2). A diferença entre a áreas das curvas ROC ($A_z=0,93$ e $A_z=0,90$, para 246 e 152 tumores, respectivamente) mostradas na Figura VII.4, evidencia a queda no desempenho da LDA.

Tabela VII.2. Desempenho da separação dos tumores em malignos ou benignos, entre das 246 amostras, quando aplicadas a LDA obtida com as 152 amostras considerando-se a melhor combinação entre parâmetros morfométricos e de textura (*onze parâmetros*)

Nº de amostras	A_z	Ac (%)	S (%)	E (%)
152	0,93	92,8	92,0	94,2
246	0,90	88,2	89,0	87,0

Onze parâmetros: nrv , R , RS , D_{NRL} , C , std_i , ent , ssd_i , sm , coo , cor .

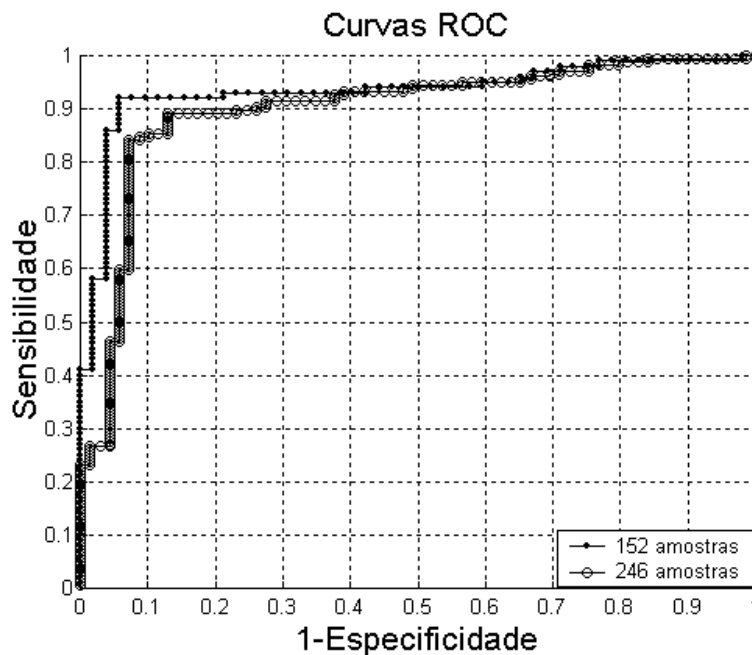


Figura VII.4. Curvas ROC do desempenho obtido na separação dos tumores em malignos ou benignos, entre das 246 amostras, quando aplicadas a LDA obtida com as 152 amostras considerando-se a melhor combinação entre parâmetros morfométricos e de textura (*onze parâmetros*).

VII.3 Discussão

O desempenho combinado dos parâmetros morfométricos e de textura na separação de tumores, quanto ao diagnóstico, foi superior ao obtido usando-se as informações de contorno e textura separadamente. Entre os onze parâmetros combinados, cinco foram morfométricos (nrv , R , RS , D_{NRL} , C), enquanto seis foram de textura (std_i , ent , ent_i , sm , coo , cor), e resultaram em valores de A_z (0,93), acurácia (92,8%), sensibilidade (92,0%) e especificidade (94,2%) superiores a 92%. Nota-se que existe um equilíbrio entre os valores de Ac , S e E quando comparado a outros estudos que empregam o LDA, sendo superior ao de LEFEBVRE *et al* (2000) ($A_z = 0,88$, $S = 94,7\%$ e $E = 80,4\%$) e similar, com base em A_z , ao reportado por HORSCH *et al.* (2002) ($A_z = 0,92$) (autor não fornece os valores de S e E).

Quando comparado a resultados obtidos em estudos clínicos onde o US de mama, interpretado por radiologistas, é utilizado como exame complementar a mamografia, o desempenho, mostrou-se superior ao reportado por SKAANE (1999) ($Ac = 81,3\%$, $S = 85,1\%$ e $E = 78,5\%$) e inferior ao de KOLB *et al.* (2002) em valores Ac (96,6%) e E (96,8%), embora este tenha reportado sensibilidade de apenas 75,3%. Além disso, o desempenho dos parâmetros de textura, neste trabalho, foi superior aos reportados por HUBER *et al.* (2000), RAHBAR *et al.* (1999) e SKAANE, (1999), que foram obtidos a partir da análise qualitativa da ecotextura das imagens por radiologistas. Este resultado sugere que a variabilidade existente na análise qualitativa da textura pode ser reduzida utilizando-se parâmetros quantitativos.

Em alguns estudos, utilizando-se apenas parâmetros de morfométricos (CHEN *et al.*, 2003b) ou de textura (CHANG *et al.*, 2003), valores maiores de S e E foram reportados. Todavia, estes estudos baseavam-se em classificadores que possuíam uma capacidade maior de generalização como RNA (CHEN *et al.*, 2003b) e *support vector machines* (CHANG *et al.*, 2003). Acredita-se, então, que o uso de um classificador não linear pode auxiliar na melhoria do desempenho dos parâmetros selecionados na distinção dos tumores de mama em imagens por US. Esta necessidade é evidenciada pelo resultado da aplicação dos 246 tumores a LDA obtida usando-se apenas 152 deles (Tabela VII.2). Todavia, constata-se, positivamente, que os valores de Ac , S e E reduziram proporcionalmente, o que sugere que, apesar da limitada capacidade de generalização da LDA, os parâmetros

seleccionados parecem representar, de maneira significativa, as características morfológicas e de textura dos tumores estudados.

VIII CONCLUSÃO

Com o objetivo de quantificar as características morfométricas e de textura de tumores de mama em imagens por ultra-som, sete parâmetros morfométricos e 20 de textura foram calculados, com base no resultado do método de segmentação proposto. Estes parâmetros foram aplicados à análise discriminante linear, e os de melhor desempenho selecionados e avaliados quanto ao seu potencial diagnóstico.

O método de segmentação, baseado em Morfologia Matemática, foi capaz de segmentar corretamente mais de 90% das imagens estudadas e realçar as irregularidades presentes no contorno dos tumores, mantendo contudo, forma similar àquelas realizadas por radiologistas.

Entre os parâmetros morfométricos estudados, dois foram propostos: valor médio quadrático residual normalizado e *Mshape*, enquanto cinco foram obtidos da literatura. O valor médio quadrático residual normalizado foi o parâmetro de melhor desempenho, sendo capaz de classificar corretamente, quanto ao diagnóstico, mais de 88% dos tumores de mama analisados. Quando combinado com a circularidade e a rugosidade, forneceu valores de sensibilidade e especificidade de 88,0% e 90,4% respectivamente. Este desempenho sugere que este parâmetro seja capaz de auxiliar na distinção de tumores de mama, apresentando, assim, potencial para a redução do número de biópsias em tumores benignos. Entretanto, o parâmetro *Mshape* não contribuiu para a distinção dos tumores.

Os parâmetros de textura foram calculados com base na matriz de co-ocorrência, clássica na literatura, e na curva de complexidade, aplicada, neste trabalho, pela primeira vez a imagens médicas. O contraste da matriz de co-ocorrência e o valor máximo da curva de complexidade foram os parâmetros de melhor desempenho individual, e quando combinados, classificaram corretamente, quanto ao diagnóstico, mais de 80,0% dos tumores de mama. O conjunto dos 12 melhores parâmetros de textura (sete calculados a partir da matriz de co-ocorrência e cinco da curva de complexidade) resultou em acurácia de 88,2%, sensibilidade de 89,0% e especificidade de 86,5%. Este resultado foi superior ao obtido a partir da análise qualitativa da textura das imagens, por radiologistas, de acordo com literatura já citada. Este fato sugere que a variabilidade existente na análise qualitativa da textura pode ser reduzida utilizando-se estes parâmetros.

Combinados, os melhores parâmetros morfométricos (nrv , R , RS , D_{NRL} , C) e de textura (std_i , ent , ent_i , sm , coo , cor) resultaram em desempenho diagnóstico superior ($A_z = 0,93$, acurácia = 92,8%, sensibilidade = 92,0% e especificidade = 94,2%) ao obtido usando-os separadamente. Assim, os resultados sugerem que estes parâmetros representam, de maneira significativa, as características morfométricas e de textura dos tumores estudados.

Uma continuação natural deste trabalho seria a construção de um classificador não linear tendo, como entrada, estes parâmetros. Em seguida, poder-se-ia criar um sistema de auxílio ao diagnóstico que seria colocado numa clínica para ser avaliado.

IX REFERÊNCIAS BIBLIOGRÁFICAS

- AL-JANOBI, A., 2001, “Performance Evaluation of Cross-Diagonal Texture Matrix Method of Texture Analysis”, *Pattern Recognition*, v.34, pp. 171-180.
- ALVARENGA, A. V., INFANTOSI, A. F. C., PEREIRA, W. C. A., 2002a, “Aplicação de Morfologia Matemática na Segmentação de Estruturas Hipoecóicas em Imagens Médicas por Ultra-som”, In: *Anais do XVIII Congresso Brasileiro de Engenharia Biomédica*, v. 5, pp. 281-285, CBEB’2002, São José dos Campos, Set.
- ALVARENGA, A. V., INFANTOSI, A. F. C., PEREIRA, W. C. A., *et al.*, 2002b, “Contour Detection of Breast Ultrasound Tumor Images Using Morphological Operators”, In: *Proceedings of the IFMBE*, v. 2, pp. 78-79, Reykjavik, Iceland, 18-22 June.
- ALVARENGA, A. V., INFANTOSI, A. F. C., AZEVEDO, C. M., PEREIRA, W. C. A., 2003a, “Aplicação de Operadores Morfológicos na Segmentação e Determinação do Contorno de Tumores de Mama em Imagens por Ultra-Som”, *Revista Brasileira de Engenharia Biomédica*, v. 19, n. 2, pp. 91-101
- ALVARENGA, A. V., INFANTOSI, A. F. C., AZEVEDO, C. M., PEREIRA, W. C. A., 2003b, “Classifying Fatty and Glandular Tissues from Breast Ultrasound Images with the Complexity Curve and Genetic Algorithm” (4398.pdf), In: *Proceedings of the IFMBE*, v. 4, pp. 01 – 04, Sydney, Australia, 24-29 Aug.
- ALVARENGA, A. V., INFANTOSI, A. F. C., AZEVEDO, C. M., *et al.*, 2004, “Morphologic Operators Applied to Breast Tumour Ultrasound Image Classification”, In: *27th International Symposium on Acoustical Imaging*, Saarbrucken, Germany, Proceedings of 27th International Symposium on Acoustical Imaging, New York: Kluwer Academic/Plenum Publishers, v. 27.
- AZEVEDO, C. M., 1994, *Manual de radiologia da mama*, Rio de Janeiro, INCa / Du Pont / Microservice, Rio de Janeiro.
- AZEVEDO-MARQUES, P. M., 2001, “Diagnóstico Auxiliado por Computador na Radiologia”, *Radiologia Brasileira*, v. 34, n. 5, pp. 285-293.

- BAHEERATHAN S., ALBREGTSEN F., DANIELSEN H. E., 1999, "New Texture Features Based on The Complexity Curve", *Pattern Recognition*, v. 32, n. 4, pp. 605-618.
- BARR, R. G., 2001, "Breast ultrasound: A bright future", *Medica Mundi*, v.45, n.2 (Jul.), pp. 8-13.
- BEASLEY, D., BULL, D. R., MARTIN, R. R., 1993, "An Overview of Genetic Algorithms: Part 1, Fundamentals", *University Computing*, v.15, n.2, pp.58-69.
- BERNSTEIN, P. L., 1997, *Desafio aos deuses: a fascinante história do risco*, Editora Campus, Rio de Janeiro.
- BETAL D., ROBERTS N., WHITEHOUSE G. H., 1997, "Segmentation and Numerical Analysis of Microcalcifications on Mammograms Using Mathematical Morphology", *British Journal of Radiology*, v.70, n.837, pp. 903-917.
- BOUKERROUI, D., BASKURT, A., NOBLE, J. A., *et al.*, 2003, "Segmentation of Ultrasound Images-Multiresolution 2D and 3D Algorithm Based on Global and Local Statistics", *Pattern Recognition Letters*, v.24, pp. 779-790.
- CASTLEMAN, K. N., 1996, *Digital Image Processing*, New Jersey, New York, Prentice-Hall International.
- CHANG, R. F., WU, W. J., MOON, W. K., *et al.*, 2003, "Improvement in Breast Tumor Discrimination by Support Vector Machines and Speckle-Emphasis Texture Analysis", *Ultrasound in Medicine and Biology*, v. 29, n. 5, pp. 679-686.
- CHANG, Y. H., ZHENG, B., GOD, W. F., *et al.*, 1998, "Identification of Clustered Microcalcification on Digitized Mammograms Using Morphology and Topology-Based Computer-Aided Detection Schemes", *Investigative Radiology*, v. 33, n. 10, pp. 746-751.
- CHEN, C. M., LU, H. H. S., HAN, K. C., 2001, "A Textural Approach Based on Gabor Functions for Texture Edge Detection in Ultrasound Images", *Ultrasound in Medicine and Biology*, v. 27, n. 4, pp. 515-534.
- CHEN, C. M., LU, H. H. S., LIN, Y. C., 2000, "An Early Vision-Based Snake Model for Ultrasound Image Segmentation", *Ultrasound in Medicine and Biology*, v. 26, n. 2, pp. 273-285.

- CHEN, C. M., LU, H. H. S., CHEN, Y. L., 2003a, "A Discrete Region Competition Approach Incorporating Weak Edge Enhancement for Ultrasound Image Segmentation", *Pattern Recognition Letters*, v. 24, pp. 693-704.
- CHEN, C. M., CHOU, Y. H., HAN, K. C., *et al.*, 2003b, "Breast Lesions on Sonograms: Computer-Aided Diagnosis with Nearly Setting-Independent Features and Artificial Neural Networks", *Radiology*, v. 226, pp.504-514.
- CHEN, D., CHANG, R., HUANG, Y., 2000, "Breast Cancer Diagnosis Using Self-Organizing Map for Sonography", *Ultrasound in Medicine and Biology*, v. 26, n. 3, pp. 405-411.
- CHENG H., LUI Y.M., FREIMANIS R.I., 1998, "A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique", *IEEE Transactions on Medical Imaging*, v. 17, n. 3, pp. 442-450.
- CHOU, Y., TIU, C., HUNG, G., *et al.*, 2001, "Stepwise Logistic Regression Analysis of Tumour Contour Features for Breast Ultrasound Diagnosis", *Ultrasound in Medicine and Biology*, v. 27, n. 11, p. 1493-1498.
- DONOHUE, K. D., HUANG, L., BURKS, T., *et al.*, 2001, "Tissue Classification with Generalized Spectrum Parameters", *Ultrasound in Medicine and Biology*, v. 27, n. 11, pp. 1505-1514
- DUTTA, S., BARBER, B. J., PARAMESWARAN, S., 1995 'Texture Analysis of Protein Distribution Images to Find Differences Due to Aging and Superfusion', *Annals of Biomedical Engineering*, v. 23, n. 6, pp.772-786.
- DRUKKER, K., GIGER, M. L., HORSCH, K., *et al.*, 2002, "Computerized Lesion Detection on Breast Ultrasound", *Medical Physics*, v. 29, n. 7, pp. 1438-1446.
- DUIJM, L. E. M., GUIT, G. L., ZAAT, J. O. M., *et al.*, 1997, "Sensitivity, Specificity and Predictive Values of Breast Imaging in The Detection of Cancer", *British Journal of Cancer*, v. 76, n. 3, pp. 377-381.
- EVANS III, W. P., MENDELSON, E., BASSETT, L., *et al.*, 2002, "Appropriate Imaging Work-up of Palpable Breast Masses", American College of Radiology Appropriateness Criteria™, pp. 961-964, Disponível em: <<http://www.acr.org>>
Acesso em: 30/10/2003

- GARRA, B. S., KRASNER, B. H., HORII, S. C., *et al.*, 1993, “Improving The Distinction Between Benign and Malignant Breast Lesions: The Value of Sonographic Texture Analysis”, *Ultrasonic Imaging*, v. 15, n. 4, pp. 267-285.
- GOLDBERG, V., MANDUCA A., EWERT D. L., *et al.*, 1992, “Improvement in Specificity of Ultrasonography for Diagnosis of Breast-Tumors by Means of Artificial-Intelligence”, *Medical Physics*, v. 19, n. 6, pp. 1475-1481.
- HADJIISKI L., SAHINER, B., CHAN, H., PETRICK, N., HELVIE, M. A., GURCAN, M., 2001, “Analysis of Temporal Changes of Mammographic Features: Computer-Aided Classification of Malignant and Benign Breast Masses”. *Medical Physics*, v. 28, n. 11, pp. 2309-2317.
- HAGEN-ANSERT, S. 1997, “Sonographic Evaluation of the Breast”, Ultrasound Online CME Courses, GE Medical Systems, Disponível em: <<http://www.gemedicalsystems.com/lapt/rad/us/education/msucmebr.html>>. Acesso em: 30/08/03.
- HARALICK, R. M., SHANMUGAM, K., DINSTEN, I., 1973, “Textural Features for Image Classification”, *IEEE Trans Syst Man Cybern*, v. 3, pp. 610–621.
- HAWKING, S., 2001, *O Universo Numa Casca de Noz*, São Paulo, Editora Mandarin.
- HAYKIN, S., 1999, *Neural Network – A comprehensive Foundation*, 2nd Ed., New Jersey, Prentice-Hall.
- HORSCH K., GIGER M.L., VENTA, L.A., *et al.*, 2001, “Automatic Segmentation of Breast Lesions on Ultrasound”, *Medical Physics*, v. 28, n. 8, pp. 1652-1659.
- HORSCH K., GIGER M. L., VENTA, L.A., *et al.*, 2002, “Computerized Diagnostic of Breast Lesions on Ultrasound”, *Medical Physics*, v. 29, n. 2, pp. 157-164.
- HOUSSAMI, N., IRWIG, L., LOY, C., 2002, “Accuracy of Combined Breast Imaging in Young Women”, *The Breast*, v. 11, pp. 36-40.
- HUANG, Y., CHEN, D., 2004, “Watershed Segmentation for Breast Tumor in 2-D Sonography”, *Ultrasound in Medicine and Biology*, v. 30, n. 5, pp. 625–632.
- HUBER, S., DANES, J., ZUNA, I., *et al.*, 2000, “Relevance of Sonographic B-Mode Criteria and Computer-aided Ultrasonic Tissue Characterization in Differential Diagnosis of Solid Breast Masses”, *Ultrasound in Medicine and Biology*, v. 26, n. 8, pp. 1243–1252.

- HUDSON, D.L., COHEN, M.E., 2000, *Neural Networks and Artificial Intelligence for Biomedical Engineering*, New York, NY, IEEE Press.
- INCa, 1997, *O Problema do Câncer no Brasil*, Quarta Edição, Instituto Nacional de Câncer / Pro-Onco, Rio de Janeiro, Brasil.
- INCa, 2002, *Falando Sobre Câncer de Mama*, MS / INCa, Rio de Janeiro.
- INCa, Rio de Janeiro, Instituto Nacional de Câncer, 2005, Disponível em: <http://www.inca.gov.br/cancer/mama/>. Acesso em: 14/02/2005
- INFANTOSI, A. F. C., SILVA JR, J. L., TIERRA-CRIOLLO, C. J., *et al.*, 1998, “Avaliação do Desempenho de Técnicas de Interpolação no Mapeamento Cerebral Utilizando Simulação”, *RBE Caderno de Engenharia Biomédica*, v. 14, n. 2, pp. 71-96.
- KAPLAN, S. S., 2001, “Clinical Utility of Bilateral Whole-Breast US in the Evaluation of Women with Dense Breast Tissue”, *Radiology*, v. 221, pp. 641-649.
- KOHONEN, T, 1998, “The Self-organizing Map”, *Neurocomputing*, n. 21, pp. 1-6.
- KOLB, T. M., LICHY, J., NEWHOUSE, J. H., 2002, “Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations”, *Radiology*, v. 225, pp. 165-175.
- KUO, W. J., CHANG, R. F., LEE, C. C., 2002, “Retrieval Technique for the Diagnosis of Solid Breast Tumors on Sonogram”, *Ultrasound in Medicine and Biology*, v. 28, n. 7, pp. 903-909.
- LEFEBVRE F., MEUNIER M., THIBAUT F., *et al.*, 2000, “Computerized Ultrasound B-Scan Characterization of Breast Nodules”, *Ultrasound in Medicine and Biology*, v. 26, n. 9, pp. 1421-1428.
- LEITE L., 2001, *Software Identificador e Classificador de Microcalcificações em Mamografias, Segundo a Classificação de Le Gal*, Projeto Final do Departamento de Engenharia Eletrônica – DEL / UFRJ, Rio de Janeiro, Brasil.
- MLODINOW L., 2004, “*A Janela de Euclides: a história da geometria, das linhas paralelas ao hiperespaço*”, São Paulo, Geração Editorial.

- MOJSILOVIC, A., POPOVIC, M., AMODAJ, N., *et al.*, 1997, "Automatic Segmentation of Intravascular Ultrasound Images: A Texture-based Approach", *Annals of Biomedical Engineering*, v. 25, n. 6, pp. 1059-1071.
- OLABARRIAGA, S. D., SMEULDERS, A. W. M., 2001, "Interaction in the Segmentation of Medical Images: A Survey", *Medical Image Analysis*, v. 5, pp. 127-142.
- RAHBAR, G., SIE, A. C., HANSEN, G. C., *et al.*, 1999, "Benign versus Malignant Solid Breast Masses: US Differentiation", *Radiology*, v. 213, pp. 889-894.
- ROBERTS, P., 1999, "The Agenda – Total Teamwork – Teams of doctors, nurses, and technicians at world-renowned Mayo Clinic bring new economy practices to "old-fashioned" medicine", *Fast Company*, v. 23, pp. 148-155.
- SIVARAMAKRISHNA, R., POWELL, K. A., LIEBER, M. L., *et al.*, 2002, "Texture Analysis of Lesions in Breast Ultrasound Images", *Computerized Medical Imaging and Graphics*, v. 26, pp. 303-307.
- SKAANE, P., 1999, "Ultrasonography as Adjunct to Mammography in The Evaluation of Breast Tumors", *Acta Radiologica Supplementum*, v. 40, suppl. 000, pp. 1-47.
- SOILLE, P., 1999, *Morphological Image Analysis*, Heidelberg, Berlin, Springer-Verlang.
- STAVROS, A. T., THICKMAN, D., RAPP, C.L., *et al.*, 1995, "Solid Breast Nodules: Use of Sonography to Distinguish Between Benign and Malignant Lesions", *Radiology*, v. 196, pp. 123-134.
- TANG, X., 1998, "Texture Information in Run-Length Matrices", *IEEE Transactions on Image Processing*, v. 7, n. 11, pp. 1602-1609.
- THOMAS, J. G., PETERS, R. A., JEANTY, P., 1991, "Automatic Segmentation of Ultrasound Images Using Morphological Operators", *IEEE Transactions on Medical Imaging*, v. 10, n. 2, pp. 180-186.
- TOURASSI, G. D., 1999, "Journey Toward Computer-Aided Diagnosis: Role of Image Texture Analysis", *Radiology*, v. 213, pp. 317-320.
- TUCERYAN, M., JAIN, A. K., 1998, "Texture Analysis". In: CHEN, C. H., PAU, L. F., WANG, P. S. P. (eds), *The Handbook of Pattern Recognition and Computer Vision*, 2 ed., chapter 2.1, New Jersey, USA, World Scientific Publishing Co.

- ZHENG, Y., GREENLEAF, J. F., GISVOLD, J.J., 1997, "Reduction of Breast Biopsies with a Modified Self-Organization Map", *IEEE Transactions on Neural Networks*, v. 8, n. 6, pp. 1386-1396.
- ZIMMER, Y., TEPPER, R., AKSELROD, S., 1999, "Computerized Quantification of Structures within Ovarian Cysts Using Ultrasound Images", *Ultrasound in Medicine and Biology*, v. 25, n. 2, pp 189-200.
- ZIMMER, Y., TEPPER, R., AKSELROD, S., 1996, "A Two-dimensional Extension of Minimum Cross Entropy Thresholding for the Segmentation of Ultrasound Images", *Ultrasound in Medicine and Biology*, v. 22, n. 9, pp. 1183-1190.
- ZONDERLAND, H. M., COERKAMP, E. G., HERMANS, J., *et al.*, 1999, "Diagnosis of Breast Cancer: Contribution of US as an Adjunct to Mammography", *Radiology*, v. 213, pp. 413-422.

X APÊNDICE I

A idéia principal do Discriminante Linear é encontrar a projeção sobre uma reta na qual amostras de diferentes classes estarão separadas (Figura A.1).

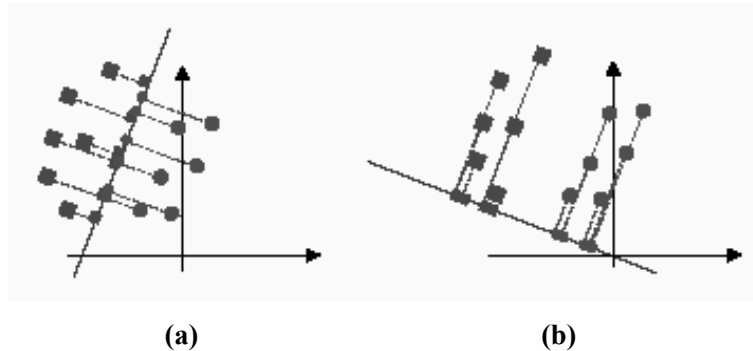


Figura A1. Exemplo da projeção das amostras de duas classes (quadrado e círculo) projetadas sobre (a) uma reta que não as separa e (b) outra que as separa.

Suponha duas classes e d -dimensional amostras x_1, x_2, \dots, x_n onde n_1 amostras são da classe 1 (quadrados) e n_2 amostras a classe 2 (círculos). Considerando a projeção sobre uma reta, a sua direção é dada pelo vetor unitário v .

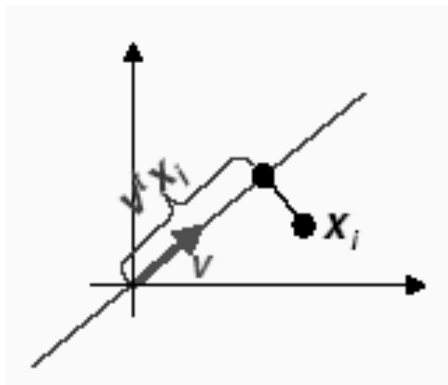


Figura A2. Projeção $v^T x_i$ da amostra x_i , na direção do vetor unitário v .

O escalar $v^T x_i$ é a distância da projeção de x_i com relação à origem. Então, $v^T x_i$ é a projeção de x_i em um subespaço unidimensional. Logo, a projeção da amostra x_i sobre uma reta na direção v é dada por $v^T x_i$. A separação entre a projeção de diferentes classes pode ser medida a partir da dispersão entre as classes (interclasse) calculada por:

$$\mathbf{S}'_b = \sum_{c=1}^C n_c (\boldsymbol{\mu}'_1 - \boldsymbol{\mu}'_2)(\boldsymbol{\mu}'_1 - \boldsymbol{\mu}'_2)^t, \quad (\text{A.1})$$

onde $\boldsymbol{\mu}'_1$ e $\boldsymbol{\mu}'_2$ são as médias das classes 1 e 2 projetadas, dadas por:

$$\boldsymbol{\mu}'_1 = \frac{1}{n_1} \sum_{x_i \in C_1} \mathbf{v}^t \mathbf{x}_i = \mathbf{v}^t \left(\frac{1}{n_1} \sum_{x_i \in C_1} \mathbf{x}_i \right) = \mathbf{v}^t \boldsymbol{\mu}_1, \text{ o mesmo para } \boldsymbol{\mu}'_2 = \mathbf{v}^t \boldsymbol{\mu}_2, \quad (\text{A.2})$$

Reescrevendo a matriz de dispersão inter-classes \mathbf{S}'_b tem-se:

$$(\boldsymbol{\mu}'_1 - \boldsymbol{\mu}'_2)^2 = (\mathbf{v}^t \boldsymbol{\mu}_1 - \mathbf{v}^t \boldsymbol{\mu}_2)^2 = \mathbf{v}^t (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \mathbf{v} = \mathbf{v}^t \mathbf{S}_b \mathbf{v}, \quad (\text{A.3})$$

Quanto maior for \mathbf{S}_b , melhor será a separação. Porém, ele não considera a dispersão de cada classe separadamente (intraclasse). Considerado $\mathbf{y}_i = \mathbf{v}^t \mathbf{x}_i$, sendo \mathbf{y}_i as amostras projetadas, A dispersão para as amostras projetadas das classes 1 e 2 são dadas por:

$$\mathbf{S}'_1 = \sum_{y_i \in \text{Classe 1}}^n (y_i - \boldsymbol{\mu}'_1)^2 \text{ e } \mathbf{S}'_2 = \sum_{y_i \in \text{Classe 2}}^n (y_i - \boldsymbol{\mu}'_2)^2. \quad (\text{A.4})$$

Normalizando \mathbf{S}_b por ambas dispersões tem-se:

$$J(\mathbf{v}) = \frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v}}{\mathbf{S}'_1 + \mathbf{S}'_2}. \quad (\text{A.5})$$

O discriminante deve determinar \mathbf{v} (direção da reta na qual as amostras serão projetadas) que maximize $J(\mathbf{v})$. As matrizes $\mathbf{S}_1 = \sum_{x_i \in \text{Classe 1}}^n (x_i - \boldsymbol{\mu}_1)(x_i - \boldsymbol{\mu}_1)^t$ e

$\mathbf{S}_2 = \sum_{x_i \in \text{Classe 2}}^n (x_i - \boldsymbol{\mu}_2)(x_i - \boldsymbol{\mu}_2)^t$ medem a dispersão das amostras originais x_i , definindo

$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$ como a dispersão intraclasse. Lembrando que $\mathbf{S}'_1 = \sum_{y_i \in \text{Classe 1}}^n (y_i - \boldsymbol{\mu}'_1)^2$, e

substituindo $\mathbf{y}_i = \mathbf{v}^t \mathbf{x}_i$ e $\boldsymbol{\mu}'_1 = \mathbf{v}^t \boldsymbol{\mu}_1$ tem-se:

$$\mathbf{S}'_1 = \sum_{x_i \in \text{Classe 1}}^n (\mathbf{v}^t x_i - \mathbf{v}^t \boldsymbol{\mu}_1)^2 \quad \text{A.6}$$

$$\mathbf{S}'_1 = \sum_{x_i \in \text{Classe 1}}^n (\mathbf{v}^t (x_i - \boldsymbol{\mu}_1))^t (\mathbf{v}^t (x_i - \boldsymbol{\mu}_1)) \quad \text{A.7}$$

$$\mathbf{S}'_1 = \sum_{x_i \in \text{Classe 1}}^n ((x_i - \boldsymbol{\mu}_1)^t \mathbf{v})^t ((x_i - \boldsymbol{\mu}_1)^t \mathbf{v}) \quad \text{A.8}$$

$$\mathbf{S}'_1 = \sum_{x_i \in \text{Classe 1}}^n \mathbf{v} (x_i - \boldsymbol{\mu}_1) (x_i - \boldsymbol{\mu}_1)^t \mathbf{v} = \mathbf{v}^t \mathbf{S}_1 \mathbf{v}, \quad \text{A.9}$$

então de maneira similar $\mathbf{S}'_2 = \mathbf{v}^t \mathbf{S}_2 \mathbf{v}$. Logo,

$$\mathbf{S}'_1 + \mathbf{S}'_2 = \mathbf{v}^t \mathbf{S}_1 \mathbf{v} + \mathbf{v}^t \mathbf{S}_2 \mathbf{v} = \mathbf{v}^t \mathbf{S}_w \mathbf{v}. \quad \text{A.10}$$

Normalizando-se \mathbf{S}_b pela dispersão intraclassa \mathbf{S}_w obtém-se a solução de Fisher dada

por $J(\mathbf{v}) = \frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v}}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}}$ e calculando-se a sua derivada e sua em relação a \mathbf{v} tem-se:

$$\frac{d}{d\mathbf{v}} J(\mathbf{v}) = \frac{\left(\frac{d}{d\mathbf{v}} \mathbf{v}^t \mathbf{S}_b \mathbf{v}\right) \mathbf{v}^t \mathbf{S}_w \mathbf{v} - \left(\frac{d}{d\mathbf{v}} \mathbf{v}^t \mathbf{S}_w \mathbf{v}\right) \mathbf{v}^t \mathbf{S}_b \mathbf{v}}{(\mathbf{v}^t \mathbf{S}_w \mathbf{v})^2} = \frac{(2\mathbf{S}_b \mathbf{v}) \mathbf{v}^t \mathbf{S}_w \mathbf{v} - (2\mathbf{S}_w \mathbf{v}) \mathbf{v}^t \mathbf{S}_b \mathbf{v}}{(\mathbf{v}^t \mathbf{S}_w \mathbf{v})^2} = 0. \quad \text{A.11}$$

Assim deve-se resolver $\mathbf{v}^t \mathbf{S}_w \mathbf{v} (\mathbf{S}_b \mathbf{v}) - \mathbf{v}^t \mathbf{S}_b \mathbf{v} (\mathbf{S}_w \mathbf{v}) = 0$. Para tal divide-se a equação por

$\mathbf{v}^t \mathbf{S}_w \mathbf{v}$ obtendo:

$$\frac{\mathbf{v}^t \mathbf{S}_w \mathbf{v} (\mathbf{S}_b \mathbf{v})}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}} - \frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v} (\mathbf{S}_w \mathbf{v})}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}} = 0 \therefore \mathbf{S}_B \mathbf{v} - \frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v} (\mathbf{S}_w \mathbf{v})}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}} = 0 \quad \text{A.12}$$

Fazendo

$$\frac{\mathbf{v}^t \mathbf{S}_b \mathbf{v} (\mathbf{S}_w \mathbf{v})}{\mathbf{v}^t \mathbf{S}_w \mathbf{v}}, \quad \text{A.13}$$

tem-se:

$$\mathbf{S}_b \mathbf{v} = \lambda \mathbf{S}_w \mathbf{v}. \quad \text{A.14}$$

Se \mathbf{S}_w possui uma inversa, pode-se converter a equação num problema padrão de autovalores $\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{v} = \lambda \mathbf{v}$. Este classificador é de fácil utilização e fornece, freqüentemente, resultados satisfatórios, não exigindo grandes recursos computacionais. Contudo, é limitado a situações onde as classes são linearmente separáveis.