

# CLASSIFICAÇÃO DE IMAGENS DE ULTRA-SOM DE TUMOR DE MAMA ATRAVÉS DE TÉCNICAS DE INTELIGENCIA COMPUTACIONAL

José L.R. Macrini<sup>1</sup>, André V. Alvarenga<sup>2</sup>, Wagner C.A. Pereira<sup>2</sup>, Antonio Fernando C. Infantsi<sup>2</sup> e Carlos E. Pedreira<sup>3</sup>

<sup>1</sup>Instituto de Estudos de Saúde Coletiva – (IESC) da Universidade Federal do Rio de Janeiro UFRJ

<sup>2</sup>Programa de Engenharia Biomédica, COPPE - Universidade Federal do Rio de Janeiro UFRJ

<sup>3</sup>Faculdade de Medicina e COPPE-PEE, Universidade Federal do Rio de Janeiro UFRJ

E-mails: macrini@centroin.com.br, victor@peb.ufrj.br, wagner@peb.ufrj.br, afci@peb.ufrj.br, pedreira@ufrj.br

**Abstract**— Mammography is the most important screening exam for early diagnosis of breast cancer. However, it still presents a lack of specificity, since a considerable number of suspicious lesions have to undergo surgical biopsy. Therefore, ultrasound (US) breast images have been used to improve diagnostics and decrease the number of unneeded biopsies for patients with palpable mass and inconclusive mammograms. Malignant breast tumors tend to present irregular and blurred contours while benign ones are usually round, smooth and well-defined. Accordingly, investigating the tumor contour may decisively help in establishing the diagnosis. On this basis, a non-linear method, Mutual Information (MI), and a linear one (Linear Discriminant Analysis - LDA) were implemented to rank morphological features in discriminating breast tumors in US images. Compared to LDA, the MI was able to identify the diagnostic contribution of *Mshape* feature. A feedforward neural network was implemented using as input seven morphological parameters to classify the breast US images. The best performance (88 %) was obtained with the five first parameters selected by MI.

**Keywords**— Mutual Information, parameters selection, ultrasound images, breast cancer, neural networks.

**Resumo** – A mamografia tem sido apontada como a única técnica diagnóstica que comprovadamente contribui, através de um programa de acompanhamento periódico, para a detecção precoce e redução da mortalidade por câncer de mama. Porém, uma quantidade considerável de lesões suspeitas ainda tem sido encaminhadas para biópsia. Imagens de ultra-som (US) têm sido usadas para melhorar o diagnóstico e diminuir o número de biópsias desnecessárias em pacientes com massas palpáveis e mamografias inconclusivas. Tumores malignos de mama tendem a apresentar contornos irregulares e borrados, enquanto tumores benignos são normalmente redondos, lisos e bem-definidos. A investigação do contorno do tumor pode ajudar no diagnóstico. Neste artigo, um método linear e um não-linear foram utilizados para ordenar parâmetros morfométricos usados para quantificar o contorno e a forma de tumores de mama em imagens por US. Comparado com o método linear, o método não-linear pôde identificar a contribuição não-linear do parâmetro morfométrico *Mshape* em relação ao diagnóstico. Uma rede neural *feedforward* foi implementada, tendo como entrada sete parâmetros morfométricos para classificar as imagens. O melhor desempenho (88%) foi obtido com os primeiros cinco parâmetros selecionados pelo método não-linear.

**Palavras-chave** - Informação Mútua, seleção de parâmetros, imagens de ultra-som, câncer de mama, redes neurais.

## 1. Introdução

A mamografia tem sido apontada como a única técnica diagnóstica que comprovadamente contribui, através de um programa de acompanhamento periódico, para a detecção precoce e redução da mortalidade por câncer de mama. Porém, a acurácia deste exame depende da composição do parênquima mamário e das características tissulares dos tumores, pois um parênquima denso dificulta a localização de um tumor. Ainda, sua especificidade, isto é o percentual de falsos positivos, pode ser melhorada. Estes falsos positivos se refletem em um considerável número de massas suspeitas e não patológicas (70 a 90% do total) (Horsh *et al*, 2002), que são recomendadas à biópsia cirúrgica (Dennis *et al*, 2001). O complemento deste exame, através de imagem de ultra-som (US) é uma forma usada para reduzir o número de biópsias desnecessárias em pacientes com massas palpáveis e mamogramas inconclusivos

(Huber *et al*, 2000, Rahbar *et al*, 1999, Skaane, 1999).

Tumores malignos de mama tendem a apresentar contornos irregulares ou indefinidos devido à infiltração no tecido ao redor (Chou *et al*, 2001). Assim, o contorno do tumor é de grande importância na formação da hipótese de diagnóstica (Huber *et al*, 2000, Skaane, 1999, Chou *et al*, 2001). Porém, quantificar de maneira precisa as características de contorno é difícil, devido à natureza do problema e a limitações intrínsecas ao próprio processo de captura da imagem (Maes *et al*, 2003). A seleção do conjunto adequado de parâmetros para a quantificação do contorno do tumor ainda é um problema em aberto. Alguns trabalhos utilizam métodos estatísticos lineares, como a Análise Linear Discriminante ou Análise de Componente Principal, para avaliar o potencial de discriminação dos parâmetros (Lafuente J *et al*, 1997).

Neste artigo, aplicou-se uma metodologia não-linear, baseada em Informação Mútua (IM)

(Cover e Thomas, 1991), para estabelecer os parâmetros morfométricos mais adequados para distinguir tumores de mama em imagens por ultra-som. Tais parâmetros, calculados a partir das técnicas do Polígono Convexo (Alvarenga *et al*, 2004) e da Distância Radial Normalizada (Chou *et al*, 2001), foram usados como entrada de uma Rede Neural em camadas para classificar os tumores mama.

## 2. Materiais e Métodos

### 2.1 Banco de dados

O banco de dados consiste em 246 imagens de US de tumores de mama de pacientes do Instituto Nacional do Câncer (Brasil - Rio de Janeiro), sendo 177 tumores malignos e 69 benignos, todos histopatologicamente provados. As imagens foram adquiridas com um transdutor linear de 7.5 MHz com resolução axial e lateral de 0,45 mm e 0,49 mm, respectivamente (*Sonoline Sienna* - Siemens, Erlangen, Germany). Na determinação do contorno do tumor foi aplicado um método de segmentação semi-automático (SAC), baseado em operadores morfológicos (Alvarenga *et al*, 2003).

### 2.2 Parâmetros Morfométricos

Os seguintes parâmetros morfométricos foram calculados a partir dos contornos determinados pelo SAC:

- \* Valor médio quadrático residual normalizado (*nrv*)
- \* Circularidade (*C*)
- \* Razão de superposição (*RS*)
- \* Índice de rugosidade (*R*)
- \* Razão de área (*RA*)
- \* Desvio padrão da distância radial normalizada (*D<sub>NRL</sub>*)
- \* Razão entre a área do tumor e seu respectivo fechamento morfológico (*Mshape*)

As definições detalhadas destes parâmetros encontram-se no Apêndice.

### 2.3 Análise Discriminante Linear

A Análise Discriminante Linear (ADL) foi aplicada aos sete parâmetros calculados (normalizados entre -1 e 1) para identificar os mais relevantes na discriminação de tumores de mama, entre malignos ou benignos, em imagens de ultra-som. Esta técnica é comumente usada para classificação e redução de dimensionalidade de dados, e maximiza a

relação de discrepância entre classes e dentro da classe, garantindo a separação máxima dos dados. Detalhes sobre esta técnica podem ser encontrados, por exemplo, em Johnson & Wichern (1998). Dado o número de observações disponível, foi usado o processo *leave-one-out* (Bishop, 1995) para estimar a performance fora da amostra. O desempenho dos parâmetros foi avaliado usando-se a área Az ( $\pm$  erro padrão) sob a curva ROC (*Receptor Operator Characteristic*).

### 2.4 Informação mútua

Considere duas variáveis aleatórias (VA),  $X$  e  $Y$ . A quantidade de informação que  $X$  fornece sobre  $Y$  é a Informação Mútua entre estas duas VA's e pode ser expressa como (Cover e Thomas, 1991):

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

onde  $p(x, y)$  é a distribuição de densidade de probabilidade conjunta das VA's  $X$  e  $Y$ . Note-se que duas VA's  $X$  e  $Y$  são estatisticamente independentes se e somente se  $p(x, y) = p(x)p(y)$ . Assim, se  $X$  e  $Y$  são independente temos que  $I(X, Y) = 0$ . Um valor alto (ou pequeno) de informação mútua significa que as variáveis são muito (ou pouco) relacionadas.

Considerando que correlação mede relações lineares, a IM é mais geral, no sentido de que avalia relações não lineares. Medindo a IM entre um parâmetro e o resultado de um processo (ou sistema), esta se avaliando quanto (de um modo não linear) o resultado é afetado por este parâmetro em particular. Da mesma maneira, com a meta de ter acesso ao grau de redundância, pode-se medir a IM entre dois parâmetros. Quanto maior é a IM, mais importante é a relação entre os parâmetros indicando redundância. Visando esta meta, teríamos que verificar a IM em um conjunto de todas as combinações de todos os parâmetros candidatos, o que é obviamente um retrocesso. Na maioria das aplicações reais, isto produziria um número infundável de possíveis associações. Um modo de se avaliar essas relações vem de um algoritmo chamado MIFS-U (Seleção de Característica de Informação Mútua sob Distribuição de Informação Uniforme), originalmente proposto em Kwak *et al* (2002). Segue uma descrição breve do algoritmo de MIFS-U:

Seja  $F$  o conjunto de todas (digamos  $n$ ) os parâmetros candidatos.

Passo 1– Inicialize  $S$  como um conjunto vazio (i.e.  $S = \emptyset$ ) e calcule a IM entre cada um dos parâmetros candidatos (i.e. cada  $f_i \in F$ ,  $i=1, \dots, n$ ) com o rótulo de saída, chame esta IM de “ $I(f_i;saída)$ ”.

Faça  $k = 1$ ;

Passo 2 – Selecione, entre todo  $f_i \in F$ , o parâmetro (chame de  $f_k$ ) que forneceu o valor máximo com  $I(f_i,saída)$ . Remova  $f_k$  de  $F$  e ponha em  $S$ , i.e.,  $F \leftarrow F - \{f_k\}$  e  $S \leftarrow f_k$

Passo 3–selecione, entre todos os parâmetros em  $F$ , o que, juntamente com os que já estão em  $S$ , maximiza a IM conjunta com a saída, i.e., selecione um  $f_k \in F$  tal que  $I(f_i,S; saída)$  seja maximizado para  $f_i \in F$ .

Se  $k < n$ , faça  $k = k + 1$  e volte ao passo 2, caso contrário, pare.

Todavia, o cálculo de  $I(f_i,S; saída)$  no passo 3 é tecnicamente difícil uma vez que envolve a maximização da probabilidade condicional (encontrar o parâmetro que maximiza a IM com relação a saída, dado que um conjunto de parâmetros já escolhidos). O algoritmo MIFS-U propõe uma simplificação que dá uma boa aproximação para  $I(f_i,S; saída)$  como detalhado em Kwak *et al* (2002). O algoritmo não só leva em consideração a IM entre cada parâmetro e a saída, como também a IM entre todos os parâmetros no conjunto  $S$ . Mais uma vez foi utilizado o método *leave-one-out* para avaliar o desempenho fora da amostra.

### 2.5 Rede Neural

Os parâmetros selecionados por MIFS-U foram usados como entrada de uma *Feedforward Neural Network (RN)*. Foi empregado o método de Regularização Bayesiana (MacKay, 1992) com o intuito de diminuir a arbitrariedade na especificação da arquitetura neural. Neste método, uma função objetivo é otimizada forçando a poda de pesos menos relevantes. A princípio utilizamos uma camada escondida e a rede foi inicializada com 10 neurônios. Utilizamos uma saída com ativação logística produzindo saídas zero ou um, representando as duas classes dos tumores, maligno ou benigno. O procedimento *leave-one-out* foi executado como uma estimativa do desempenho fora da amostra.

## 3. Resultados e Discussão

Entre todos os parâmetros, tanto ADL e IM, indicaram o valor médio quadrático residual normalizado ( $nrv$ ) e a circularidade ( $C$ ) como os

parâmetros mais relevantes para distinguir tumores malignos de benignos. Apesar desses parâmetros apresentarem alta IM entre si (Tabela 2), há uma quantidade apreciável de informação na variável  $C$  que não está incluída em  $nrv$ , como é indicado pelo alto valor de IM com a saída (Tabela 1). Ambos os parâmetros,  $nrv$  (Alvarenga *et al*, 2004) e  $C$  (Chou *et al*, 2001), foram citados previamente como relevantes para distinguir tumores de mama malignos de benignos.

Tabela 1. Desempenho individual dos parâmetros morfométricos na distinção de tumores malignos e benignos a partir de imagens de US. Valor médio ( $\mu$ ) e coeficiente de variação (CV) para ADL (Az sob a curva ROC) e para a IM, usando-se o método *leave-one-out*.

Parâmetros	$\mu_{Az}$	CV(%)	$\mu_{IM}$	CV(%)
$nrv$	0,86	2,68	0,37	3,39
$C$	0,78	3,84	0,28	1,87
$RS$	0,75	4,23	0,20	2,83
$R$	0,64	5,84	0,12	2,84
$Mshape$	0,60	6,48	0,17	2,73
$RA$	0,58	6,86	0,06	6,99
$D_{NRL}$	0,57	6,90	0,11	3,91

Os resultados para as principais parâmetros e os valores obtidos para IM e ADL encontram-se na Tabela 1. É interessante notar que a característica  $RS$  foi ordenada na terceira posição (Tabela 1) em relação à saída, embora só tenha sido classificada como a quinta mais importante através da ordenação do MIFS-U (Tabela 4). Isto pode ser explicado pelo alto valor de IM entre  $RS$  e os dois parâmetros mais bem ordenados,  $nrv$  e  $C$  (Tabela 2). Isto significa que a informação adicional que poderia ser fornecida por  $RS$ , já foi incorporada em  $nrv$  e  $C$ .

Analisando-se o parâmetro  $Mshape$ , a IM entre este  $nrv$ , e também entre  $Mshape$  e  $C$  (Tabela 2), nota-se que estes não apresentam valores altos. Isto significa que a informação contida em  $Mshape$  não está presente em  $C$  e nem tão pouco em  $nrv$ . Adequadamente, o algoritmo MIFS-U escolhe  $Mshape$  para a terceira posição e difere do procedimento ADL que ordenou esta variável como a quinta mais relevante. Além disso, na Tabela 3, a correlação, que é uma medida linear, entre  $Mshape$  e a saída é a mais baixa entre todos os parâmetros selecionados, em contraste com a IM (Tabela 1). Dessa forma, aplicando o algoritmo MIFS-U, a relação não linear entre  $Mshape$  e a saída foi revelada.

Tabela 2. Informação Mutua entre parâmetros.

	1	2	3	4	5	6	7
1	-	1.17	0.96	0.57	0.67	0.61	0.72
2	1.17	-	1.25	0.60	0.60	0.89	0.87
3	0.96	1.25	-	0.43	0.45	0.84	0.77
4	0.57	0.60	0.43	-	0.55	0.50	0.50
5	0.67	0.60	0.45	0.55	-	0.48	0.56
6	0.61	0.89	0.84	0.50	0.48	-	2.14
7	0.72	0.87	0.77	0.50	0.56	2.14	-

1 – *nrv*, 2 – *C*, 3 – *RS*, 4 – *R*, 5 – *Mshape*, 6 – *RA*, 7 – *D<sub>NRL</sub>*.

A característica rugosidade (*R*) subiu uma posição comparada com o valor da IM e com a ordenação final do MIFS-U (Tabela 4). Isto provavelmente é devido ao baixo valor relativo da IM com todos os outros parâmetros (Tabela 2) refletindo que a informação contida em *R* não está presente em quaisquer das outras parâmetros. Em trabalho recente feito por Chou *et al.* (2001) os parâmetros *R*, *C* e *D<sub>NRL</sub>* foram usados como as melhores para classificar tumores de mama maligno ou benigno em imagens de US. No presente estudo os autores aplicaram a regressão logística *stepwise*, i.e., uma aproximação linear generalizada.

Tabela 3. Correlação entre parâmetros.

	1	2	3	4	5	6	7	8
1	1.00	-0.61	0.77	-0.21	-0.11	-0.56	-0.58	0.52
2	-0.61	1.00	-0.81	0.18	0.08	0.52	0.51	-0.45
3	0.77	-0.81	1.00	-0.21	-0.10	-0.72	-0.70	0.44
4	-0.21	0.17	-0.21	1.00	-0.05	-0.13	-0.06	-0.28
5	-0.11	0.08	-0.10	-0.05	1.00	0.22	0.23	-0.10
6	-0.56	0.52	-0.72	-0.13	0.22	1.00	0.98	-0.16
7	-0.58	0.51	-0.70	-0.06	0.23	0.98	1.00	-0.17
8	0.52	-0.45	0.44	-0.28	-0.10	-0.16	-0.17	1.00

1 – *nrv*, 2 – *C*, 3 – *RS*, 4 – *R*, 5 – *Mshape*, 6 – *RA*, 7 – *D<sub>NRL</sub>*, 8 – Saída.

Os parâmetros selecionados pelo MIFS\_U foram usados alimentando uma RN através de cinco experiências diferentes cujos resultados se encontram expostos na Tabela 5. O melhor desempenho obtido foi alcançado utilizando os parâmetros *nrv*, *C*, *Mshape*, *R* e *RS*, como entrada, obtendo um percentual de acerto médio de 88%, 92.1% e 78.3% de acurácia, sensibilidade e especificidade, respectivamente. Cabe salientar que com o acréscimo dos parâmetros *D<sub>NRL</sub>* e *RA* na entrada da rede houve piora no desempenho. Este comportamento coincide com os resultados mostrados pela IM que indicam que estes dois parâmetros apresentam baixa informação mútua com o desfecho.

#### 4. Conclusão

Neste artigo buscou-se mostrar, através da implementação do método MIFS-U o potencial da utilização de técnicas baseadas em IM no estabelecimento dos parâmetros morfométricos mais relevantes para discriminação de tumores

de mama em imagens de ultra-som. Das sete parâmetros, o valor médio quadrático residual normalizado (*nrv*) e a circularidade (*C*) foram escolhidas tanto por MIFS-U como por ADL como as mais relevantes. A contribuição mais proeminente de IM esta na identificação do comportamento não linear da variável *Mshape* e revelando esta característica como a terceira em importância, enquanto ADL a classificou em quinto lugar.

Uma segunda contribuição é dada pela implantação de uma rede neural em camadas para classificação dos tumores como benigno e maligno, chegando-se a resultados com acurácia superior a 88%.

Tabela 4. Parâmetros relevantes ordenadas pelo algoritmo MIFS-U.

Ordem de Relevância	Saída MIFS-U
1	<i>nrv</i>
2	<i>C</i>
3	<i>Mshape</i>
4	<i>R</i>
5	<i>RS</i>
6	<i>D<sub>NRL</sub></i>
7	<i>RA</i>

Tabela 5. Resultados das RN considerando os parâmetros ordenados pelo algoritmo MIFS-U. Os números na primeira linha indicam os parâmetros usados na entrada da rede, de acordo com a Tabela 4.

Parâmetros	1 a 3	1 a 4	1 a 5	1 a 6	1 to 7
Acurácia (%)	79.3	84.2	<b>88.2</b>	82.5	84.6
Sensibilidade (%)	81.4	88.1	<b>92.1</b>	86.4	89.3
Especificidade (%)	73.9	73.9	<b>78.3</b>	72.5	72.5

#### Referências Bibliográficas

- Dennis M A, Parker S H, Klaus A J *et al.* (2001) Breast biopsy avoidance: the value of normal mammograms and normal sonograms in the setting of a palpable lump. *Radiology* 219:168–191
- Horsh K, Giger M L, Venta L A *et al.* (2002) Computerized diagnostic of breast lesions on ultrasound. *Med Phys* 29:157-164
- 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 Med & Biol* 26:1243–1252
- Rahbar G, Sie A C, Hansen G C *et al.* (1999) Benign versus malignant solid breast masses: US differentiation, *Radiology* 213:889-894
- Skaane P (1999) Ultrasonography as adjunct to mammography in the evaluation of breast tumours. *Acta Radiol. Supplementum* 40:1-47

- Chou Y H, Tiu C M, Hung G S *et al.* (2001) Stepwise logistic regression analysis of tumour contour features for breast ultrasound diagnosis. *Ultrasound in Med & Biol* 27:1493-1498
- Maes F, Vandermeulen D, Suetens P (2003) Medical image registration using mutual information. *Proc of the IEEE* 91:1699-1722
- Lafuente R, Belda J M, Sanchez-Lacuesta J *et al.* (1997) Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis. *Clin Biomech* 13:216-229
- Cover T(1991) *Elements of information theory.* Wiley, New York
- Alvarenga A V, Infantosi A F C, Pereira W C A *et al.* (2004) Normalised radial length and convex polygons to classify breast tumour contours in ultrasound images. *IFMBE News* 69:44-48
- Castleman, K.N. (1996). *Digital Image Processing.* Prentice-Hall International, New Jersey, New York.
- Alvarenga, A.V., A.F.C. Infantosi, C.M. Azevedo, W.C.A. Pereira (2003). Application of Morphological Operators on the Segmentation and Contour Detection of Ultrasound Breast Images. *Brazilian Journal of Biomedical Engineering*, 19, 91-101.
- Horsh, K., M.L. Giger, L.A. Venta, and C.J. Vyborny (2002). Computerized Diagnostic of Breast Lesions on Ultrasound. *Medical Physics*, 29, 157-164.
- Soille, P. (1999). *Morphological Image Analysis,* Springer-Verlag, Heidelberg, Berlin.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition.* Oxford. Clarendon Press.
- Johson A. R. & Wichern D.W. (1998). *Applied Multivariate Statistical Analysis,* 4th . Ed, Prentice Hall.
- Cover TM, JA Thomas, *Elements of Information Theory;* New York: Wiley, 1991
- Kwak, N. and Choi, C. (2002). Input Feature Selection for Classification Problems. *IEEE Trans. Neural Networks*, 13, 143-159.
- MacKay D J C (1992) Bayesian interpolation. *Neural Computation* 4:415-447

## Apêndice

Distância Radial Normalizada (NRL) é calculado como (Chou *et al.*, 2001):

$$d(i) = \frac{\sqrt{(x(i) - X_0)^2 + (y(i) - Y_0)^2}}{\max(d(i))}, 1 \leq i \leq N \quad (2)$$

onde  $(X_0, Y_0)$  e  $(x(i), y(i))$  são respectivamente as coordenadas do centróide e do *pixel* de limite na *i*-ésima localização,  $N$  é o número de *pixels* de contorno e  $\max(d(i))$  é o valor máximo da distância radial. Foram calculados três parâmetros de  $d(i)$ : desvio padrão ( $D_{NRL}$ ), razão de área (RA) e o índice de rugosidade (R). O  $D_{NRL}$  é definido como (Chou *et al.*, 2001):

$$D_{NRL} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d(i) - d_m)^2} \quad (3)$$

onde é  $d_m$  o valor médio de  $d(i)$ . Este parâmetro fornece a medida das irregularidades macroscópicas do contorno (Chou *et al.*, 2001).

A razão de área (RA) é definida por (Chou *et al.*, 2001):

$$RA = \frac{1}{d(i) \cdot N} \sum_{i=1}^N (d(i) - d_m)^2 \quad (4)$$

onde

$$d(i) - d_m = 0 \forall d(i) \leq d_m \quad (5)$$

calcula a porcentagem do tumor fora da região circular definida por  $d_m$ . Quanto mais irregular é o contorno, maior é o valor de RA.

O índice de rugosidade (R) é definido como (Chou *et al.*, 2001):

$$R = \frac{1}{N} \sum_{i=1}^N |d_n(i) - d_n(i+1)| \quad (6)$$

que é a distância entre *pixels* vizinhos no contorno. Contornos irregulares apresentam valores altos de rugosidade.

O Polígono convexo (Castleman, 1996) é a forma geométrica que circunscreve o contorno estabelecido pelo SAC. Quanto mais irregular o contorno, maior sua diferença em relação ao polígono convexo. Esta diferença pode ser quantificada usando-se dois parâmetros: a razão de superposição (RS) e o valor médio quadrático residual normalizado (*nrv*).

O parâmetro RS é definido por (Horsch *et al.*, 2002):

$$RS = \frac{Area(S_m)}{Area(S_o)} \quad (7)$$

onde  $S_m$  é a imagem binária determinada pelo SAC e  $S_o$  é a imagem binária de seu respectivo polígono convexo. Então, se as áreas têm a mesma forma e tamanho e estão na mesma posição, RS será a unidade.

A aplicação do parâmetro *nrv* é baseada na determinação de um resíduo  $S_r$  definido como (Alvarenga *et al.*, 2004):

$$S_r = Area(S_o) - Area(S_m) \quad (8)$$

Se as áreas são idênticas (forma e tamanho) e estão na mesma posição,  $S_r = 0$ . Logo,  $nrv$  é definido por (Alvarenga *et al.*, 2004):

$$nrv = \frac{\psi_r^2}{\psi_o^2} \quad (9)$$

onde  $\psi_r^2$  é o quadrado de  $S_r$  e  $\psi_o^2$  é o quadrado do perímetro do contorno  $P_o$ .

Também foram calculados dois outros parâmetros: a circularidade ( $C$ ) e  $Mshape$ . O parâmetro  $C$  se mostrou importante para a classificação correta de tumores de mama (Chou *et al.*, 2001) e é definido como:

$$C = \frac{P^2}{A} \quad (10)$$

onde  $P$  é o perímetro e  $A$  a área do tumor. O perímetro foi medido somando-se o número de *pixels* no contorno do tumor e a área corresponde ao número de *pixels* no interior do contorno.

$Mshape$  é definido como a relação entre a área  $S_m$  e a área do fechamento morfológico de  $S_m$ . Este operador morfológico permite preencher pequenos buracos (possíveis dados perdidos) (Soille, 1999) do contorno definido por SAC. Conseqüentemente, quanto mais irregular é o contorno, menor é  $Mshape$ .