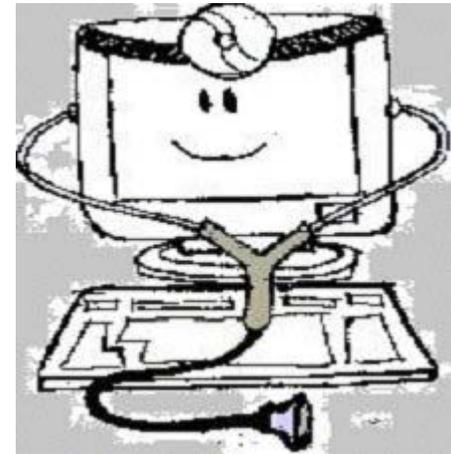


PUC-Rio / IMPA  
27 de novembro de 2002

## Medidas Globais em 3D para Diagnóstico de Nódulo Pulmonar Solitário

Aristófanês Corrêa Silva

Orientadores : Paulo Cezar P. Carvalho  
Marcelo Gattass



# Roteiro da Apresentação

- Motivação
- Nódulo pulmonar solitário
- Objetivo
- Coeficiente de atenuação
  - Textura – histograma, SGLDM, GLDM e GLRLM
- Forma
  - curvatura, tipos de superfícies e médias de curvedness
- Análise discriminante linear
- Testes e Resultados
- Conclusão

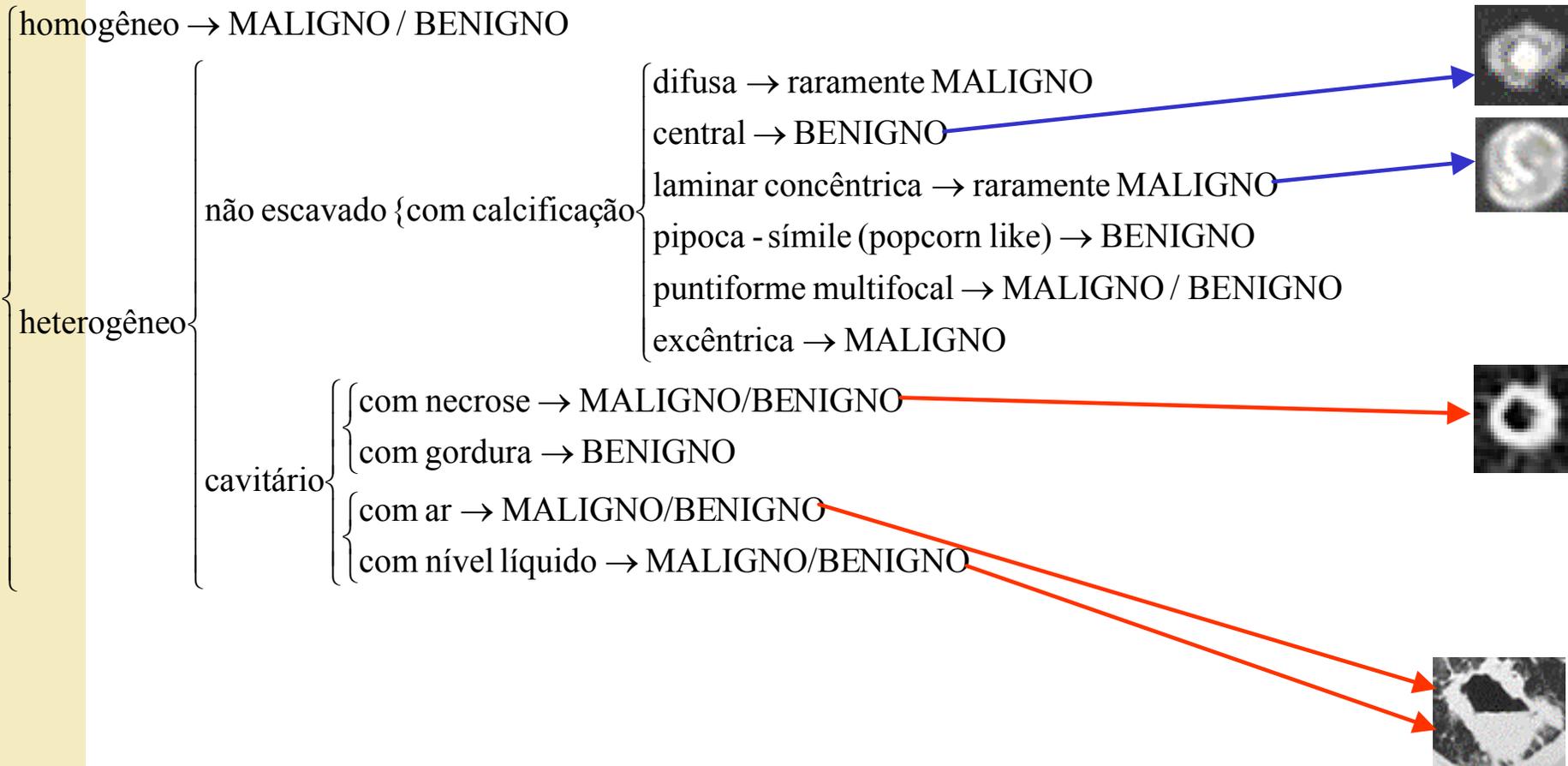
# Motivação

- O câncer de pulmão é o mais comum de todos os tumores malignos, apresentando um aumento por ano de 2% na sua incidência mundial
- O câncer de pulmão é conhecido como um dos cânceres de menor sobrevida a partir do diagnóstico
- O nódulo na fase inicial (pequena dimensão, tecido sem distinção e forma mal definida) é de difícil diagnóstico
- As medidas efetuadas pelos médicos para diagnóstico são de forma artesanal : a régua para medir o diâmetro e a jarra com água para medir o volume
- O cansaço visual e o fator emocional do médico podem influenciar no diagnóstico
- Diagnósticos errados causam intervenção cirúrgica desnecessária

# Nódulo Pulmonar Solitário - NPS

## Coeficiente de Atenuação

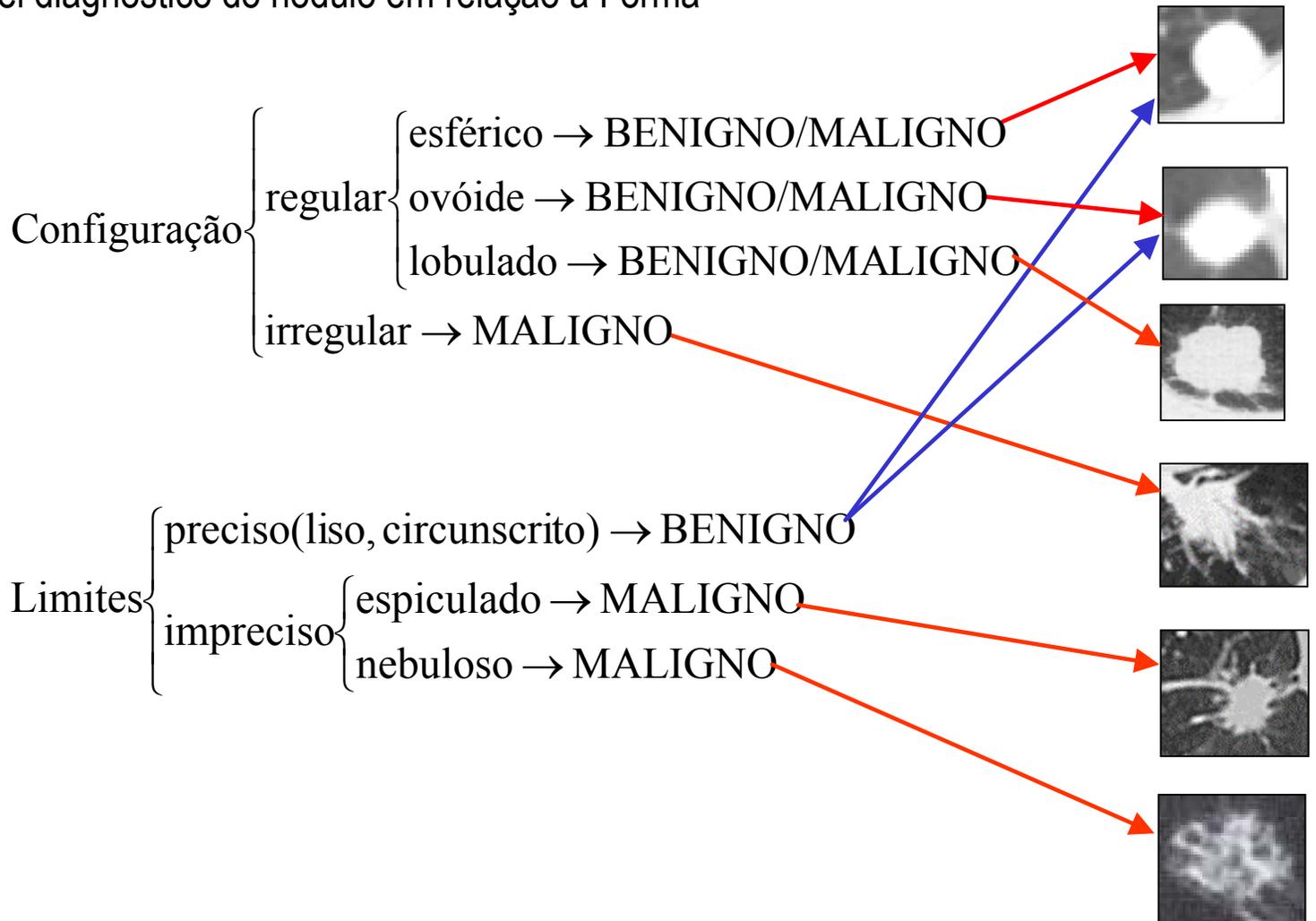
● Provável diagnóstico do nódulo em relação ao coeficiente de atenuação ( tecido na TC )



# Nódulo Pulmonar Solitário - NPS

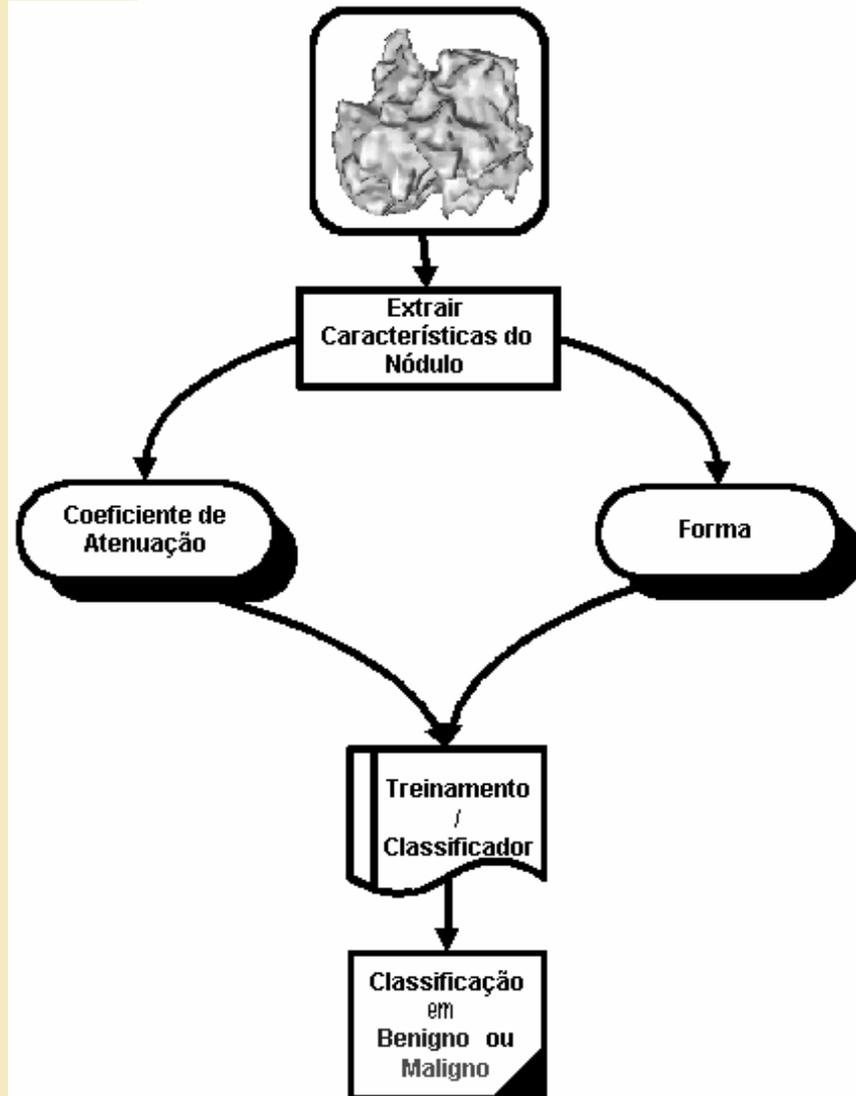
## Forma

● Provável diagnóstico do nódulo em relação à Forma

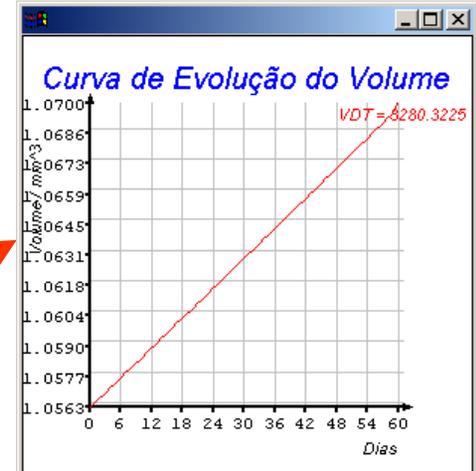
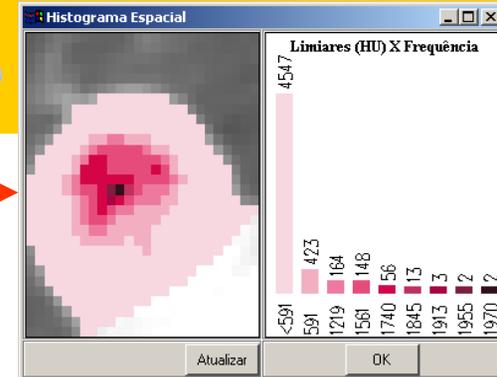
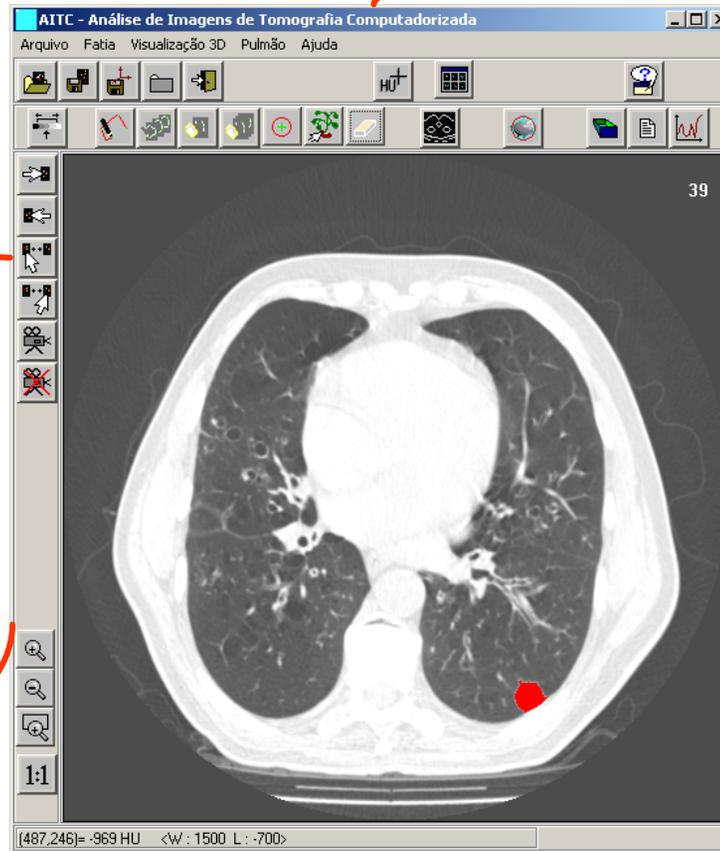
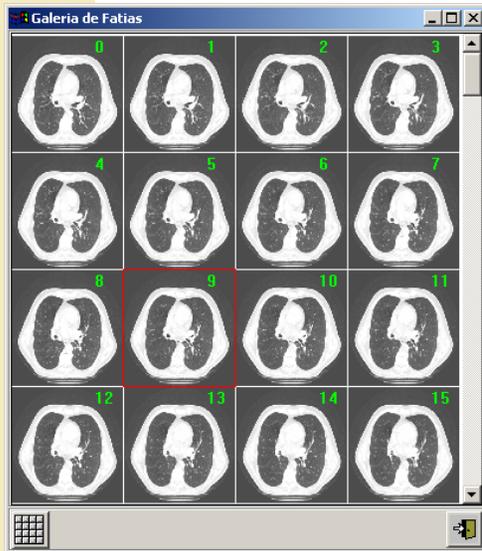


# Objetivos

- Coeficiente de atenuação em 3D
  - Textura : Histograma, SGLDM, GLDM e GLRLM)
- Forma em 3D
  - Curvatura média e gaussiana, tipo de superfície e curvedness
- Coeficiente atenuação + Forma
- Análise discriminante
  - Benigno ou Maligno



# Sistema de Análise de Nódulo Pulmonar - SANP



# Coeficiente de Atenuação (CA) Histograma

- Estatística de nível de cinza de primeira ordem
- Histograma do NPS
- Medidas
  - Média (MEN) – Mede a média do CA no NPS
  - Desvio Padrão (STD) – Mede a dispersão do CA no NPS
  - Simetria (SKW) – Mede a tendência da distribuição em relação à normal
  - Curtose (KUR) – Mede o achatamento da distribuição em relação à normal
  - Energia (ENE) – Mede a não uniformidade do NPS
  - Entropia (ENT) – Mede a uniformidade do NPS
  - Contraste (CNT) – Mede a quantidade de variação local no NPS
  - Quantidade de Coeficiente de Atenuação  $\geq 200$  UH (NCMAI)
  - Quantidade de Coeficiente de Atenuação  $< 200$  UH (NCMEN)

# Coeficiente de Atenuação (CA)

## SGLDM em 2D – Definições e Exemplo

- *Spatial Gray Level Dependence Method*
- Estatística de nível de cinza de segunda ordem
- Representa a frequência (ou probabilidade) de cada par de pixel no espaço em relação ao valor de nível de cinza
- Normalmente os ângulos são quantizados em  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  e simétricos
- Normalizado dividindo pelo número de pixels para produzir probabilidades
- Problemas : **distância? quantização?**
- Medidas : Halarick et al. (1973) extraiu 14 medidas. Ohanian et al. (1986) mostrou que um pequeno subconjunto é suficiente para caracterizar vários tipos texturas

1	1	2	2	2
1	1	2	2	2
1	3	3	3	3
3	3	4	4	4
3	3	4	4	4

	1	2	3	4
1	2	2	1	0
2	0	4	0	0
3	0	0	5	2
4	0	0	0	4

$$d = 1, \theta = 0$$

# Coeficiente de Atenuação (CA)

## SGLDM – Medidas

- Contraste (CON) – Mede a quantidade de variação local da imagem

$$CON = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{ij} (i - j)^2$$

- Homogeneidade (IDM) – Mede a homogeneidade da imagem

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} P_{ij}$$

- Segundo Momento Angular (ASM) – Mede a uniformidade da imagem

$$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{ij}^2$$

# Coeficiente de Atenuação (CA)

## SGLDM – Medidas

- Entropia (ENT) – Mede a quantidade de variação local da imagem

$$ENT = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{ij} \log(P_{ij})$$

- Variância (VAR) – Mede a dispersão em relação à média

$$VAR = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 P_{ij}$$

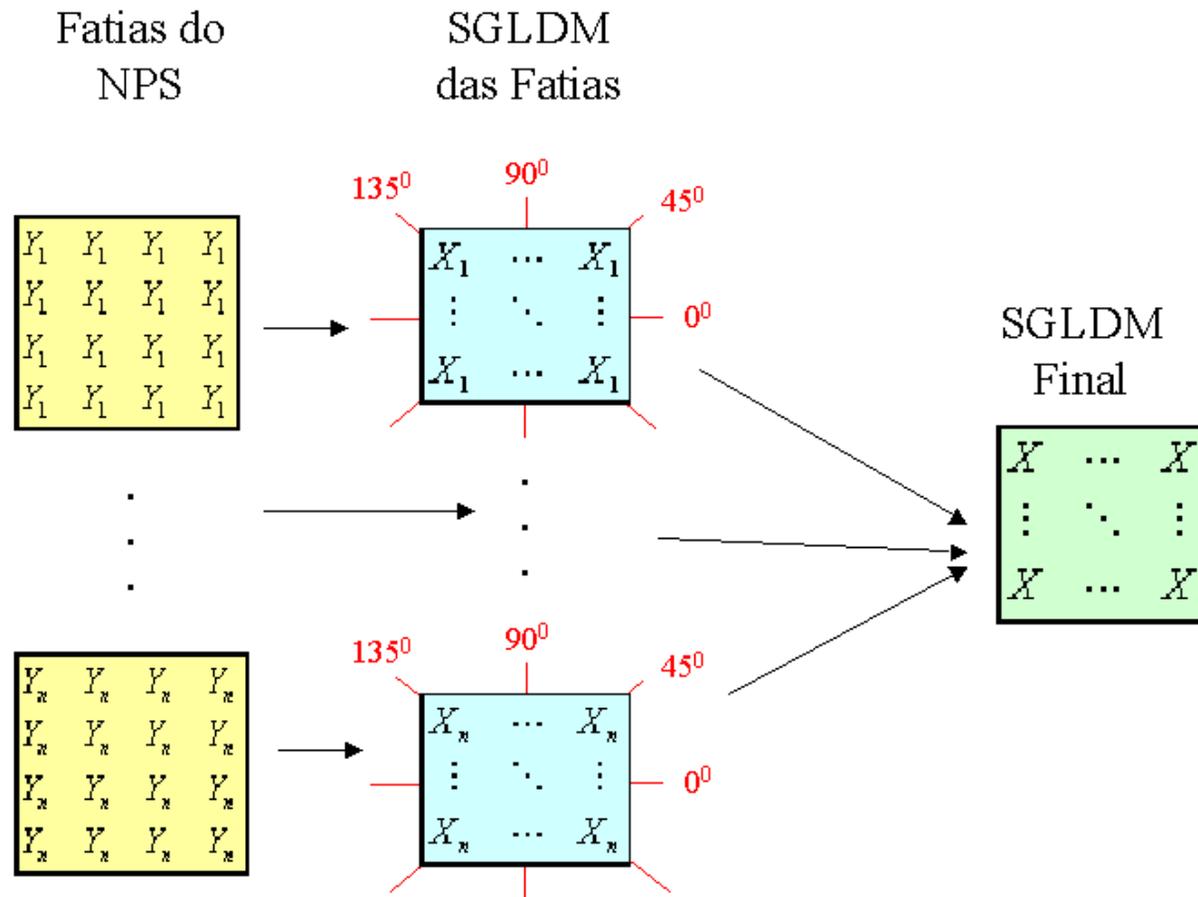
- Correlação (COR) – Mede a dependência linear de um nível de cinza em relação aos seus vizinhos

$$COR = \frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (ij) P_{ij} - \mu_i \mu_j}{\sigma_i \sigma_j}$$

# Coeficiente de Atenuação (CA) SGLDM em 3D – Métodos

## ● Método I

- Soma das matrizes de cada fatia da TC



# Coeficiente de Atenuação (CA) SGLDM em 3D – Métodos

## ● Método II

- 26 vizinhos a uma distância  $d$

$$P_{IK} = \left\| w(I(i), I(k), d(i, k)) \right\|, w(b_{Ii}, b_{Ik}, b_d) = \{(i, k) \in \mathfrak{R}^3 \mid i \neq k, \\ b_{Ii} = I(i), b_{Ik} = I(k), b_d = \text{round}(d(i, k)), x_k = (x_i + \Delta x), \\ y_k = (y_i + \Delta y), z_k = (z_i + \Delta z), -G \leq \Delta x \leq G, -G \leq \Delta y \leq G, \\ 0 \leq \Delta z \leq G\}$$

onde,

$$(i, k) \rightarrow \text{par de voxel} \quad i = (x_i, y_i, z_i) \quad k = (x_k, y_k, z_k)$$

$$d(i, k) \rightarrow \max(|x_1 - x_2|, |y_1 - y_2|, |z_1 - z_2|)$$

$I(i)$  e  $I(k)$   $\rightarrow$  intensidade dos voxels

# Coeficiente de Atenuação (CA) GLDM em 2D – Definições e Exemplo

- *Gray Level Difference Method*
- Estatística de nível de cinza de segunda ordem
- Histograma da diferença absoluta dos níveis de cinza de pares de voxels
- Normalmente os ângulos são quantizados em  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  e simétricos
- Normalizado dividindo pelo número de pixels para produzir probabilidades
- Problemas : **distância?** **quantização?**

$$P(k) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{ij}, \text{ onde } |i - j| = k$$

1	1	2	2	2
1	1	2	2	2
1	3	3	3	3
3	3	4	4	4
3	3	4	4	4

0	1	2	3	4
15	4	1	0	0

$$d = 1, \theta = 0$$



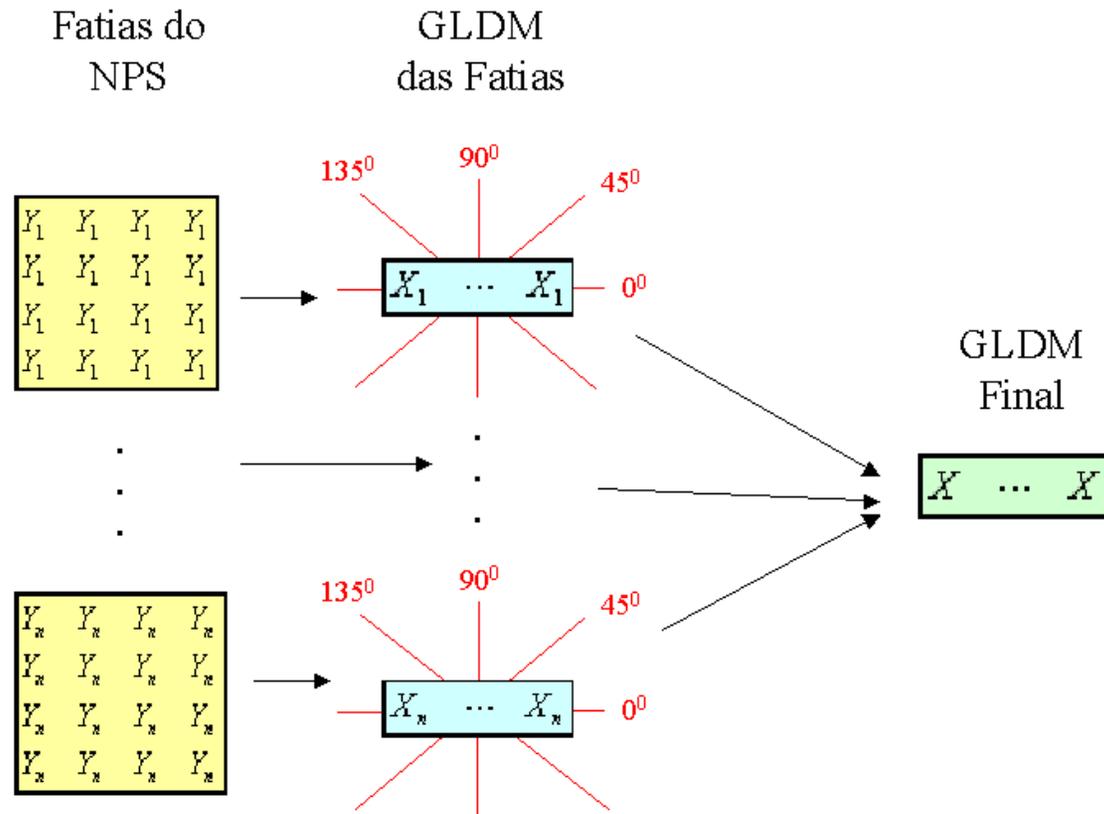
## Coeficiente de Atenuação (CA) GLDM – Medidas

- Contraste (CON) – Mede a quantidade de variação local da imagem
- Homogeneidade (IDM) – Mede a homogeneidade da imagem
- Segundo Momento Angular (ASM) – Mede a uniformidade da imagem
- Entropia (ENT) – Mede a quantidade de variação local da imagem
- Variância (VAR) – Mede a dispersão em relação à média
- Correlação (COR) – Mede a dependência linear de um nível de cinza em relação aos seus vizinhos

# Coeficiente de Atenuação (CA) GLDM em 3D – Métodos

## ● Método I

- Soma dos histogramas de cada fatia da TC



## ● Método II – Semelhante ao do SGLDM

# Coeficiente de Atenuação (CA)

## GLRLM em 2D – Definições e Exemplo

- *Gray Level Run Length Method*
- Estatística de nível de cinza de alta ordem
- Descreve um conjunto de voxels de intensidade constante localizado em uma direção
- Normalmente os ângulos são quantizados em  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  e simétricos
- Problemas : **quantização?**
- $B(a,r)$  é o número de primitivas (grupos de repetições) em todas as direções tendo comprimento  $r$ ,  $a$  é o nível de cinza,  $m$  e  $n$  são as dimensões da imagem, e  $L$  é o número de níveis de cinza da imagem

1	1	2	2	2
1	1	2	2	2
1	3	3	3	3
3	3	4	4	4
3	3	4	4	4

Image

		Run length			
		1	2	3	4
Grey level	1	1	2	0	0
	2	0	0	2	0
	3	0	2	0	1
	4	0	0	2	0

$M(i,j) | 0^\circ$

		Run length			
		1	2	3	4
Grey level	1	0	1	1	0
	2	0	3	0	0
	3	3	1	1	0
	4	0	3	0	0

$M(i,j) | 90^\circ$

# Coeficiente de Atenuação (CA)

## GLRLM – Medidas

- Ênfase em primitivas longas (LRE)

Enfatiza as primitivas longas de nível de cinza da imagem. Valor alto significa grandes quantidades de primitivas longas com a mesma densidade

$$LRE = \frac{1}{K} \sum_{a=1}^L \sum_{r=1}^{N_r} B(a, r) r^2$$

- Ênfase em primitivas curtas (SER)

Enfatiza as primitivas curtas de nível de cinza da imagem. Valor alto significa grandes quantidades de primitivas curtas com a mesma densidade

$$SRE = \frac{1}{K} \sum_{a=1}^L \sum_{r=1}^{N_r} \frac{B(a, r)}{r^2}$$

- Não uniformidade do nível de cinza (GLD)

Mede a não uniformidade das densidades. Longas primitivas contribuem mais para a função

$$GLD = \frac{1}{K} \sum_{a=1}^L \left[ \sum_{r=1}^{N_r} B(a, r) r^2 \right]^2$$

# Coeficiente de Atenuação (CA)

## GLRLM – Medidas

- Não uniformidade do comprimento (RLD)

Mede a não uniformidade das primitivas. Se as primitivas são igualmente distribuídas em relação ao comprimento, a função terá valores baixos.

$$RLD = \frac{1}{K} \sum_{a=1}^{N_r} \left[ \sum_{r=1}^L B(a, r) r^2 \right]^2$$

- Percentagem de primitivas dentro da imagem (RP)

$$RP = \frac{K}{mn}$$

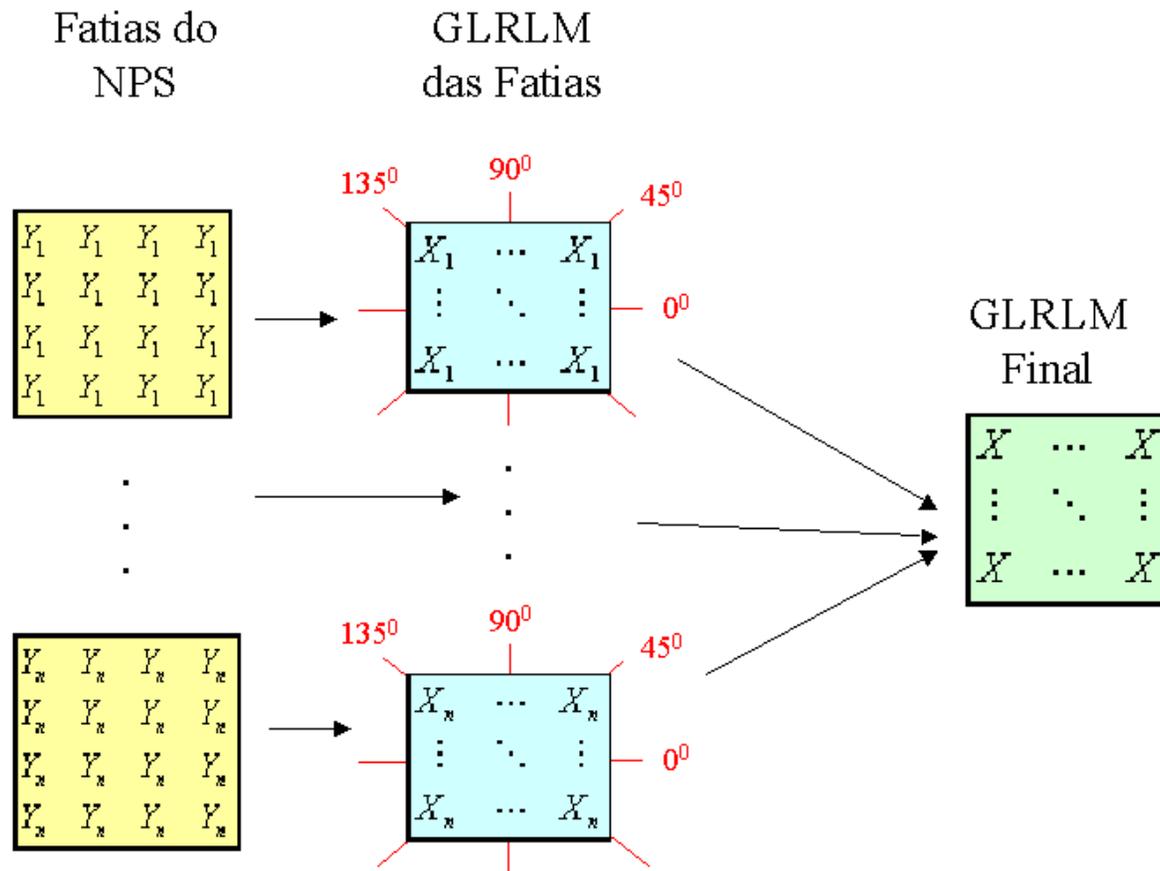
K é o número de primitivas

$$K = \sum_{a=1}^L \sum_{r=1}^{N_r} B(a, r)$$

# Coeficiente de Atenuação (CA) GLRLM em 3D – Método

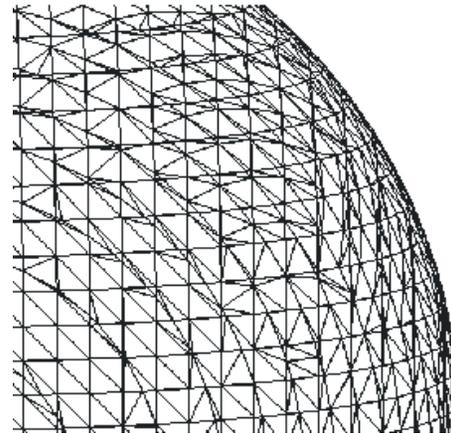
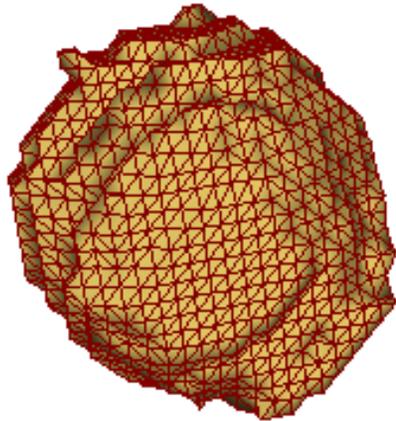
## ● Método I

- Soma das matrizes de cada fatia da TC

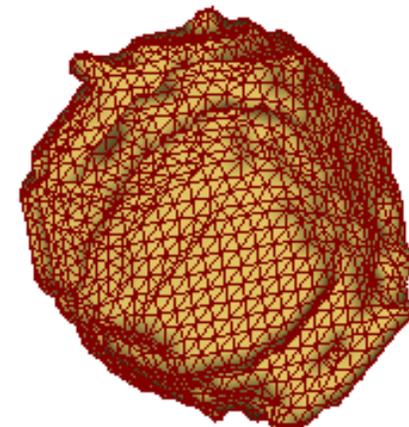
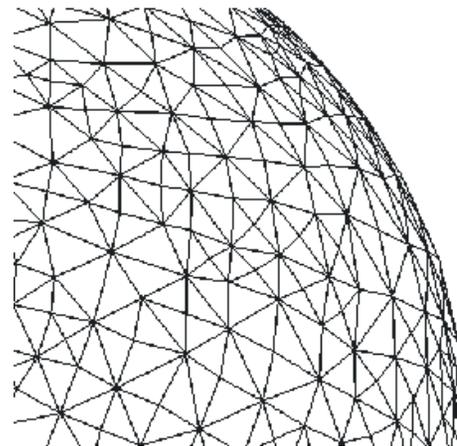
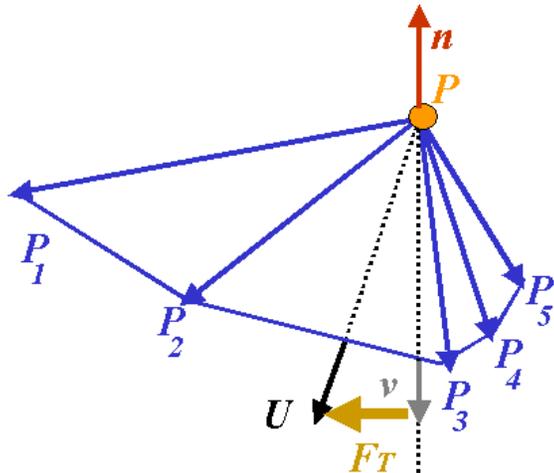


# Forma Reconstrução 3D

## ● Marching Cubes



## ● Suavização – método tangencial laplaciano



# Forma Medidas

- Taxa de Esfericidade (ESF)

Mede o comportamento do NPS em relação à esfera. Se a forma for próxima a da esfera o valor será próximo de 1.

$$ESF = \frac{6\sqrt{\pi}V}{A^{3/2}} \quad , V \text{ é o volume e } A \text{ é a área da superfície.}$$

- Taxa de Convexidade (CON)

Mede o grau de convexidade do NPS. Quanto mais convexo for nódulo, mais próximo de 1 será o seu valor.

$$CON = \frac{A(B)}{A(H_B)}$$

$A(B)$  é área da superfície do NPS e  $A(H_B)$  é área da superfície do seu fecho convexo

# Forma Medidas

## ● Taxa de Curvatura no NPS

curvaturas principais

$$k_{\min, \max} = H \mp \sqrt{H^2 - K}$$

$K$  e  $H$  são as curvaturas gaussianas e média, respectivamente

- Intrínseca – gaussiana (TCI) : conta o número de regiões com ondulação e saliência na superfície com a forma de metade de uma esfera

$$TCI = \frac{1}{4\pi} \iint |k_{\max} k_{\min}| dA$$

$$TCI = \frac{1}{4\pi} \sum_{\text{faces}} K^+ A_f$$

onde  $K^+$  é a curvatura gaussiana  $\geq 0$  e  $A_f$  é a área da face.

- Extrínseca – média (TCE) : conta o número de regiões com sulcos e fendas na superfície com a forma de metade de um cilindro

$$TCE = \frac{1}{4\pi} \iint |k_{\max}| (|k_{\max}| - |k_{\min}|) dA$$

$$TCE = \frac{1}{4\pi} \sum_{\text{faces}} |k_{\max}| |k_{\max} - k_{\min}| A_f$$

# Forma Medidas

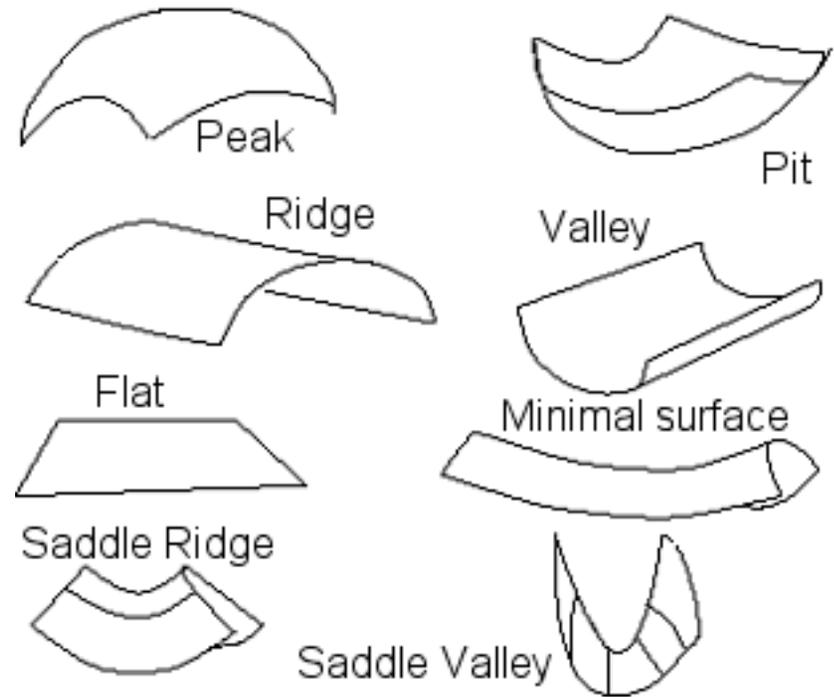
## ● Quantidade de cada Tipo de Superfície no NPS

K			
H	+	0	-
-	Peak	Ridge	Saddle ridge
0	none	Flat	Minimal surface
+	Pit	Valley	Saddle valley

$$Q_i = \sum_{i=1}^4 TS_i$$

Na prática só 4 tipos de superfícies

- Quantidade do tipo peak (QPK)
- Quantidade do tipo pit (QPI)
- Quantidade tipo saddle ridge (QSR)
- Quantidade tipo saddle valley (QSV)



# Forma Medidas

- Taxa da Área de cada Tipo de Superfície

$$A_i = \frac{1}{AS} \sum_{i=1}^4 ATS_i$$

onde  $ATS_i$  é a área de cada tipo da superfície,  $AS$  é a área total da superfície

- Taxa de área do tipo peak (APK)
- Taxa de área do tipo pit (API)
- Taxa de área do tipo saddle ridge (ASR)
- Taxa de área do tipo saddle valley (ASV)

# Forma Medidas

- Média de Curvedness de cada Tipo de Superfície no NPS

Curvedness – Mede o grau de curvatura

$$c = \sqrt{\frac{k_{\min}^2 + k_{\max}^2}{2}}$$

$$C_i = \frac{1}{N_i} \sum_{i=1}^4 c_i ,$$

onde  $c_i$  é a curvedness de cada tipo da superfície,  $N_i$  é o quantidade de curvedness de cada tipo de superfície

- Média de Curvedness do tipo peak (CPK)
- Média de Curvedness do tipo pit (CPI)
- Média de Curvedness do tipo saddle ridge (CSR)
- Média de Curvedness do tipo saddle valley (CSV)

# Análise Discriminante Linear

- Técnica estatística multivariada usada para investigar, avaliar diferenças entre grupos e classificar entidades dentro de grupos
- Introduzido por Fisher em 1936
- Função discriminante linear

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \cdots + \beta_p X_{ip} = \beta_0 + \sum_{j=1}^p \beta_j X_{ij}$$

onde  $i = 1, 2, \dots, n$  representa as observações;  $j=1, 2, \dots, p$  representa as variáveis independentes;  $Y_i$  é o escore discriminatório (variável dependente)

- Vários grupos ou dois grupos (ou Binária ou Duo Grupal)
- Maximiza a separação entre os grupos
- Classificação supervisionada

# Análise Discriminante Linear

- Função discriminante linear para 2 grupos

$$Y = aX \quad a = S^{-1} \left( \bar{X}_1 - \bar{X}_2 \right) \quad Y = \left( \bar{X}_1 - \bar{X}_2 \right)^T S^{-1} X$$

$$\bar{Y}_1 = \left( \bar{X}_1 - \bar{X}_2 \right)^T S^{-1} X_1 \quad \bar{Y}_2 = \left( \bar{X}_1 - \bar{X}_2 \right)^T S^{-1} X_2 \quad M = \frac{1}{2} \left( \bar{Y}_1 + \bar{Y}_2 \right)$$

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

$$S_i = \frac{1}{n_i - 1} \sum_{h=1}^{n_i} \left( X_{hi} - \bar{X}_i \right) \left( X_{hi} - \bar{X}_i \right)^T$$

Classificação :

Grupo 1

$$\bar{Y}_1 \geq M$$

Grupo 2

$$\bar{Y}_2 < M$$

# Testes e Resultados

## Considerações

- SGLDM, GLDM e GLRLM
  - Distância – 1,2 e 3, pois interessa os pequenos detalhes
  - Quantização – 8, 16, 32 e 64 níveis de cinza
  - Método I e II – todas as distâncias e todas as quantizações
- Validação do treinamento – Resubstituição e Jackknife (deixa um de fora)
- Dois testes
  - Individualmente – Histograma, SGLDM, GLDM, GLRLM e Forma
  - Forma e conjunto de medidas de textura (CA) mais estável (resubstituição e Jackknife).
- Treinamento e classificação – SPSS ( *Statistical Package for the Social Sciences*)
- SPSS eliminou automaticamente variáveis baseado na tolerância de 0.001 (significância das variáveis).

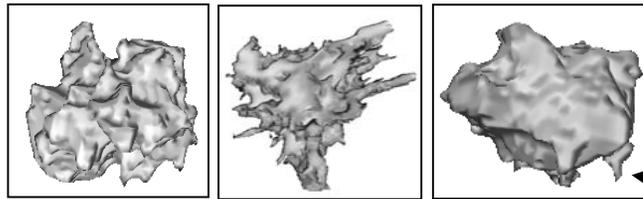
# Testes e Resultados

## Considerações

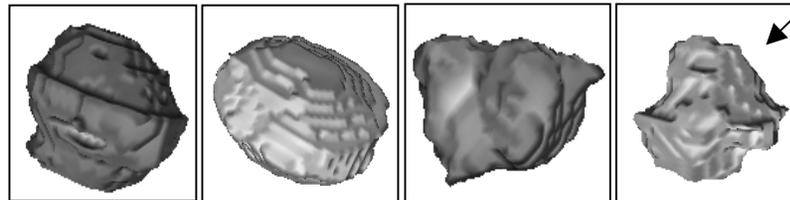
### Amostra

- 22 NPS (17 benignos e 5 malignos)
- NPS de diversos tamanhos, níveis de evolução e diagnósticos (tuberculoma, adenocarcinoma, etc.)

Maligno



Benigno



= #?

# Testes e Resultados

## Teste I

### ● Resubstituição – acertos

	<i>SGLDM - %</i>	<i>GLDM - %</i>	<i>GLRLM - %</i>	<i>Histograma - %</i>	<i>Forma - %</i>
Método I	86.4	68.2	90.9	-	-
Método II	77.3	72.7	-	-	-
-	-	-	-	72.7	100.0

### ● Jackknife – acertos

	<i>SGLDM - %</i>	<i>GLDM - %</i>	<i>GLRLM - %</i>	<i>Histograma - %</i>	<i>Forma - %</i>
Método I	86.4	50.0	54.5	-	-
Método II	36.4	36.4	-	-	-
-	-	-	-	50.0	86.4

# Testes e Resultados

## Teste II

### ● Resubstituição e Jackknife

**Resultados da Classificação<sup>b,c</sup>**

			Predicted Group Membership		Total
			benignos	malignos	
Resubstituição	Count	benignos	17	0	17
		malignos	2	3	5
	%	benignos	100.0	.0	100.0
		malignos	40.0	60.0	100.0
Jackknife	Count	benignos	11	6	17
		malignos	2	3	5
	%	benignos	64.7	35.3	100.0
		malignos	40.0	60.0	100.0

b. 90.9% dos casos com resubstituição classificados corretamente.

c. 63.6% dos casos com Jackknife classificados corretamente.



# Testes e Resultados Análise

- SGLDM, GLDM e GLRLM
  - O Método I teve o melhor desempenho
  - Variáveis com quantização de 8 e 16 níveis de cinza apareceram com mais frequência
  - SGLDM foi a técnica mais estável
- Forma
  - Teve um ótimo desempenho
- Os resultados individualmente de SGLDM e Forma foram melhor que o conjunto das duas



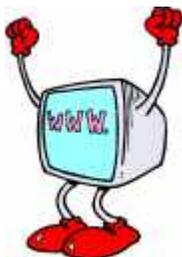
## Conclusão

- O tamanho da amostra prejudicou a análise. É necessário uma amostra bem maior que o número de variáveis, para um diagnóstico mais confiável
- A seleção de variáveis é fundamental para uma boa análise (ex.: stepwise, forward, etc.)
- Aplicar outra técnica estatística de classificação (ex.: regressão logística)

# Agradecimentos



**Tecgraf**  
PUC-RIO



<http://www.visgraf.impa.br/Projects/vismed/>