Aplicações de Conceitos de Teoria de Informação em Processamento de Imagens Digitais

Pesquisadores

Gilson Antonio Giraldi (LNCC - Petrópolis)
gilson@lncc.br

Paulo Sérgio Silva Rodrigues (FEI - SP)
psergio@fei.edu.br

Outlines

- Introduction
- Entropy and Segmentation
- Nonextensive entropy segmentation
- PME and Image Processing
- Nonextensive entropy for CAD Systems and Anomaly Network Detection
- Final Comments

INCT - MACC

• Medicina Assistida por Computação Científica

• 128 pesquisadores

•23 Laboratórios

• 33 instituições nacionais (localizadas em 11 estados da federação)

• 10 instituições do exterior (Espanha, Argentina, Canadá, Suiça, etc)

Histogram and Probability

Original Image

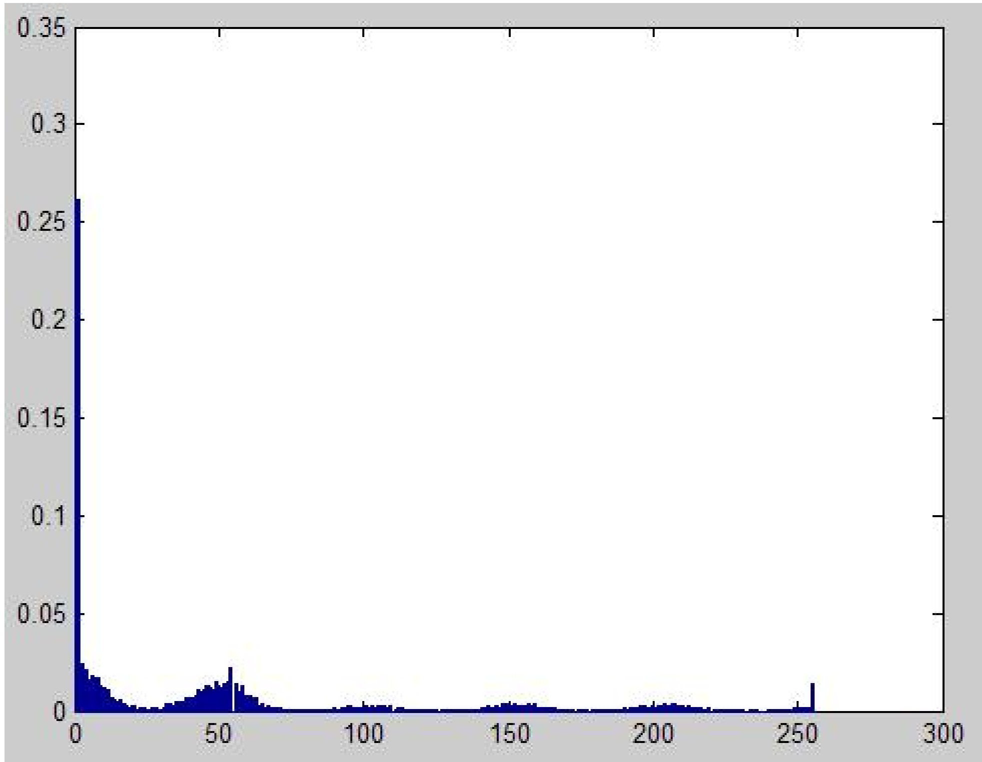


Image Histogram

Entropy and Segmentation [T. Pun (1981)]

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

$$H^{A+B}(t) = H^{A} + H^{B}$$

$$t^{opt} = \arg\max\left[H^{A+B}(t)\right]$$

Segmentation Based on NonExtensive Entropy Concepts [Albuquerque at. al. (1984)]

$$S_q = \frac{1 - \sum_{i=1}^{n} p_i^q}{q - 1} \quad \xrightarrow{q \to 1} \quad H = -\sum_{i=1}^{n} p_i \log p_i$$

$$S_q^{A+B} = S_q^A + S_q^B + (1-q) \cdot S_q^A \cdot S_q^B$$

$$t_q^{opt} = \arg\max\left[S_q^{A+B}(t)\right]$$

PME for Histogram Transformation

Constraints

$$\sum_{i=1}^{W} p_i = 1,$$

$$\frac{\sum_{i=1}^{W} e_i p_i^q}{\sum_{i=1}^{W} p_i^q} = U_q,$$

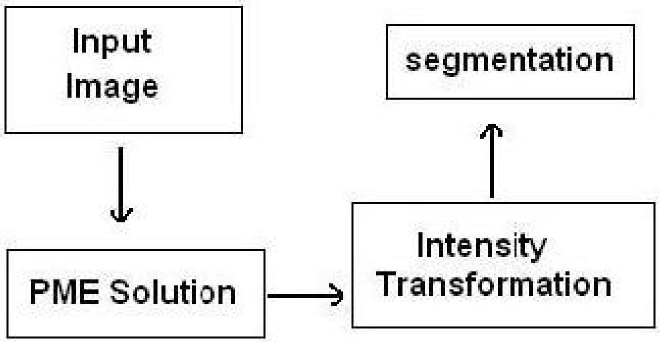
$$p_j = \frac{\left[1 - \frac{(q-1)}{k}\beta\left(\frac{e_j - U_q}{\sum_{i=1}^{W} p_i^q}\right)\right]^{\frac{1}{1-q}}}{\sum_{m=1}^{W}\left[1 - \frac{(q-1)}{k}\beta\left(\frac{e_m - U_q}{\sum_{i=1}^{W} p_i^q}\right)\right]^{\frac{1}{1-q}}},$$

$$\beta = -k\frac{\left(\frac{(1-q)\left(\sum_{i=1}^{W} p_i^q \ln p_i\right) - \left(1 - \sum_{i=1}^{W} p_i^q\right)}{(q-1)^2}\right)}{\left(\frac{\left(\sum_{i=1}^{W} p_i^q\right)\left(\sum_{i=1}^{W} e_i p_i^q \ln p_i\right) - \left(\sum_{i=1}^{W} e_i p_i^q\right)\left(\sum_{i=1}^{W} p_i^q \ln p_i\right)}{\left(\sum_{i=1}^{W} p_i^q\right)^2}\right)}.$$

Image Enhacement

If the right-hand side works as a contraction map F ($\|F(x) - F(y)\| \leq \alpha \|x - y\|$, with $\alpha \in [0, 1)$) then, we can obtain a solution through a recursive procedure:

$$\begin{aligned}
p_1^{n+1} &= F_1\left(p_1^n, p_2^n, ..., p_W^n\right), \\
p_2^{n+1} &= F_2\left(p_1^n, p_2^n, ..., p_W^n\right), \\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots \\
p_{W-1}^{n+1} &= F_{W-1}\left(p_1^n, p_2^n, ..., p_W^n\right), \\
p_W^{n+1} &= F_W\left(p_1^n, p_2^n, ..., p_W^n\right).
\end{aligned}$$



Application

Original Image

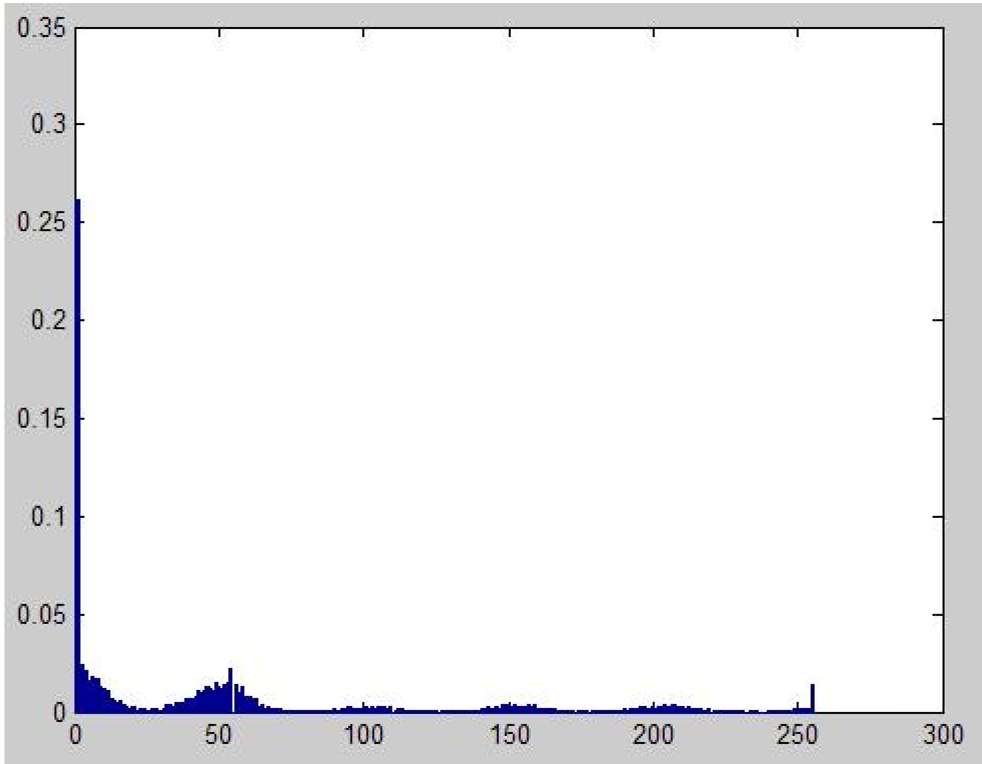
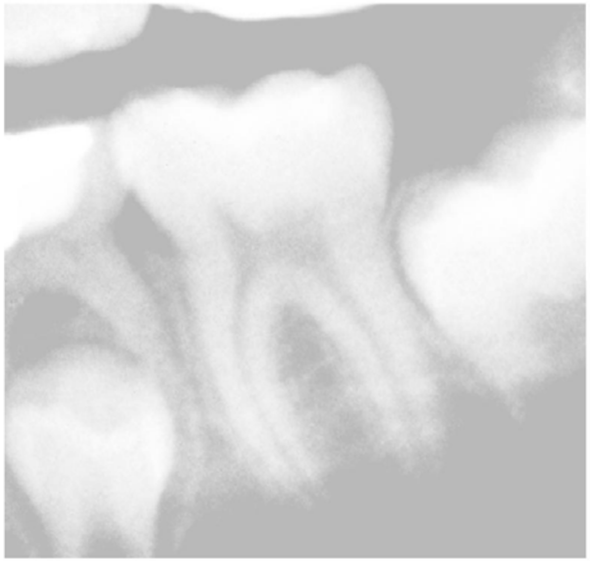
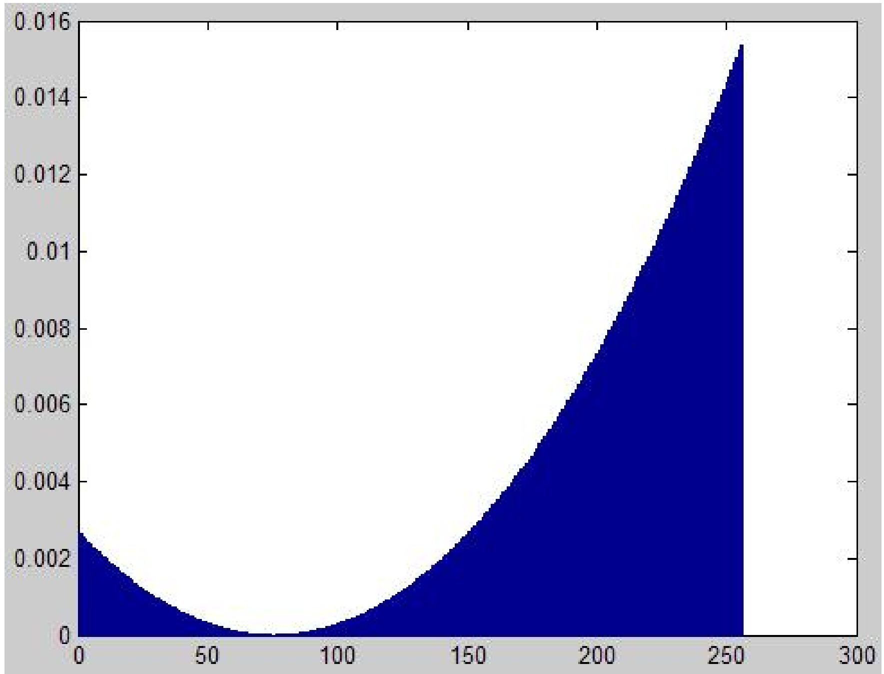


Image Histogram

Application: Results



Output Image



PME solution

Segmentation Results

Non-extensive segmentation: q=0.5

T=115

T=224

Non-Extensive Entropy for CAD Systems for Breast Cancer Images

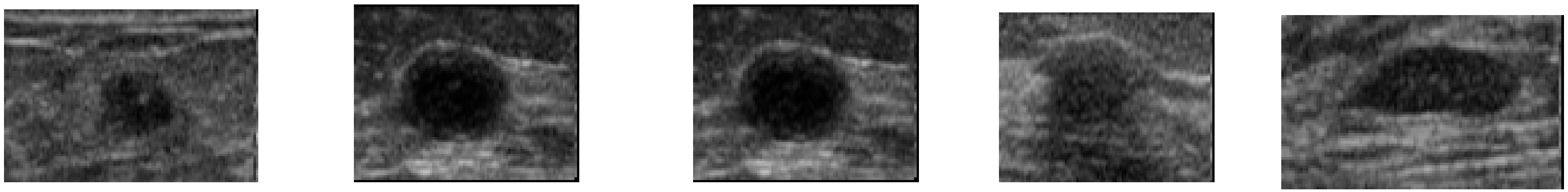
Department of Electrical Engineering,
FEI, Sao Paulo – Brazil

National Laboratory for Scientific Computing,
Rio de Janeiro - Brazil

Motivation

- Breast cancer
- Prevention and early detection
- Improvement in technology
- CAD systems
- 2D and 3D scanners
- The lesion extraction from US image
- Segmentation in the early steps

US DAC



BGS Entropy

$$S = -\sum p_i \ln p_i$$

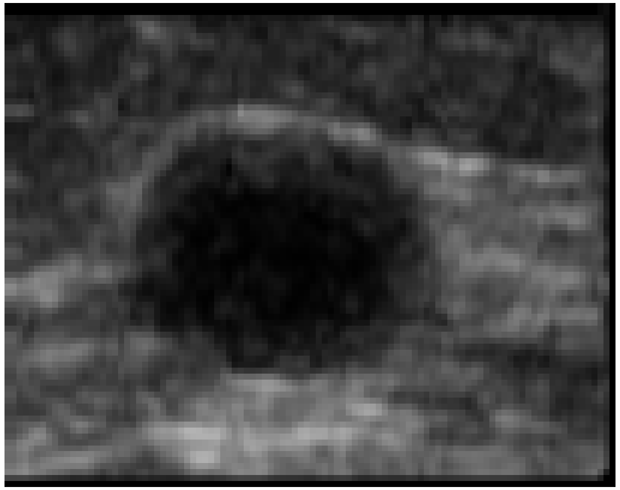
$$S(P * Q) = S(P) + S(Q)$$

Non-extensive Entropy

M. P. Albuquerque, M. P. Albuquerque, I. A. Esquef and A.R.G. Mello, *Image Thresholding using Tsallis Entropy*. Pattern Recognition Letters, 25:1059-1065, 2004

$$S_q = \frac{1 - \sum_i^W p_i^q}{1 - q}$$

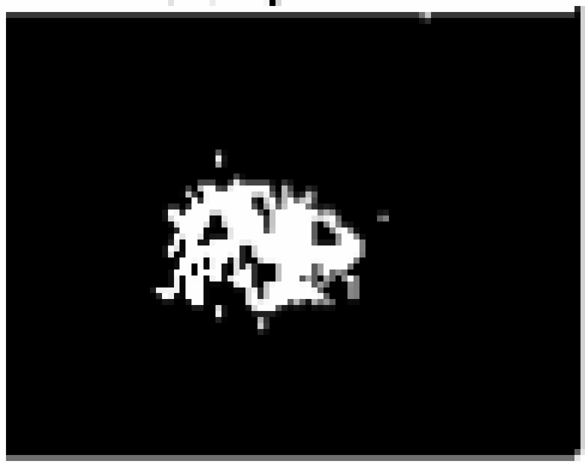
$$S(P*Q) = S(P) + S(Q) + (1-q)S(P)*S(Q)$$



(a) original US image

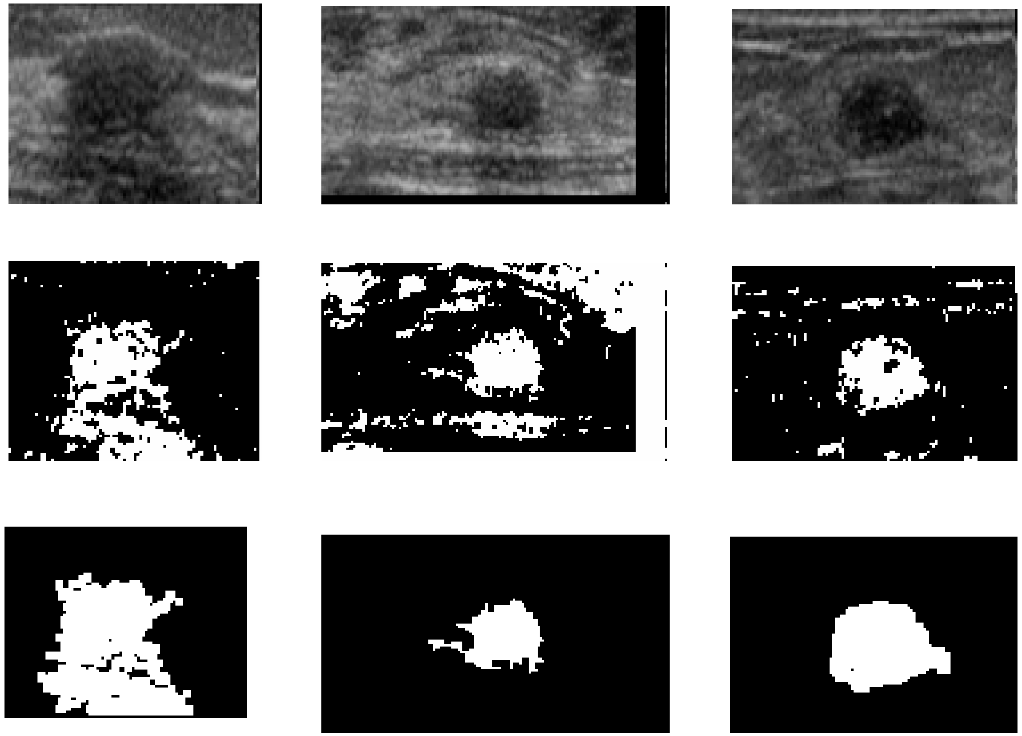
(b) q = 1.0

(c) q = 6.0

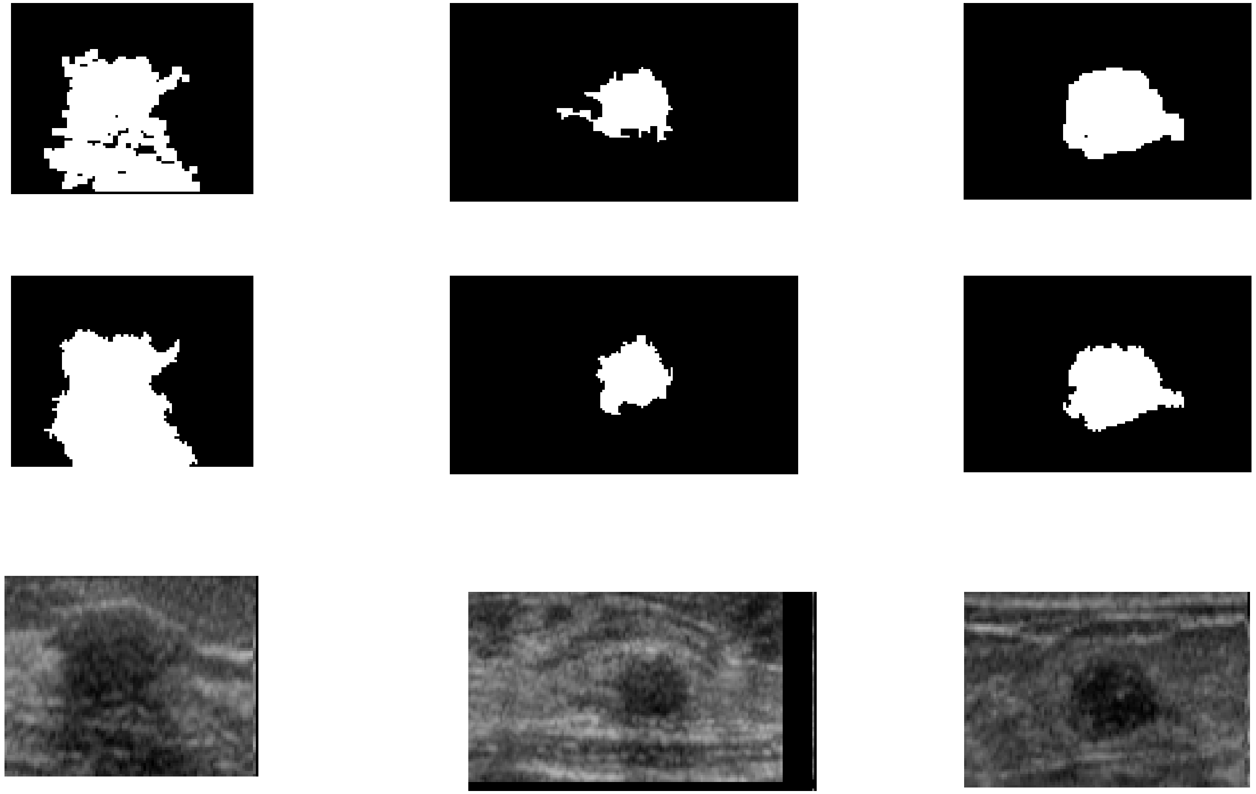


(d) q = 10.0

Morphological chain approach

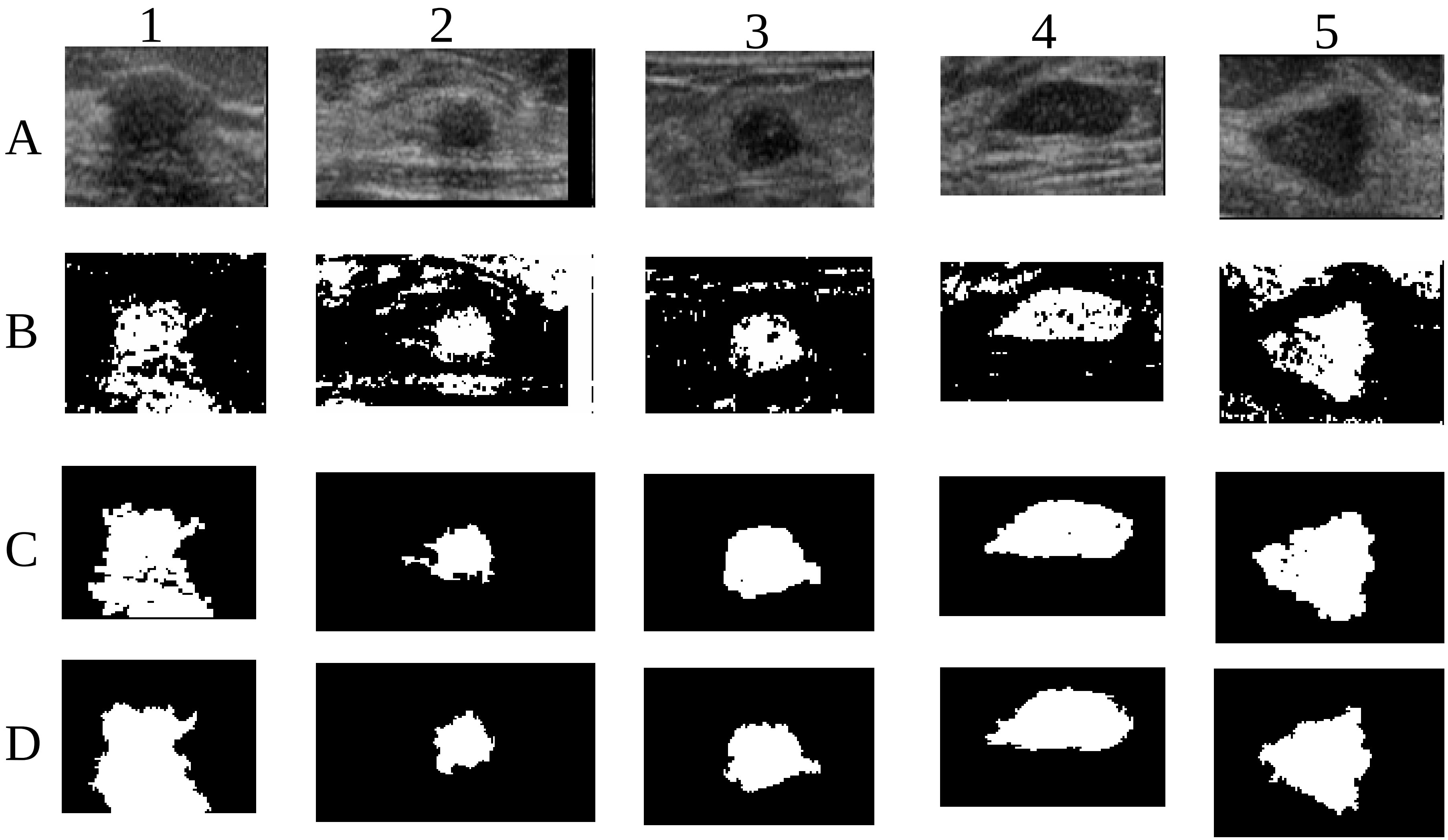


Level Set evolution



The *First Three* Steps of the Proposed Methodology

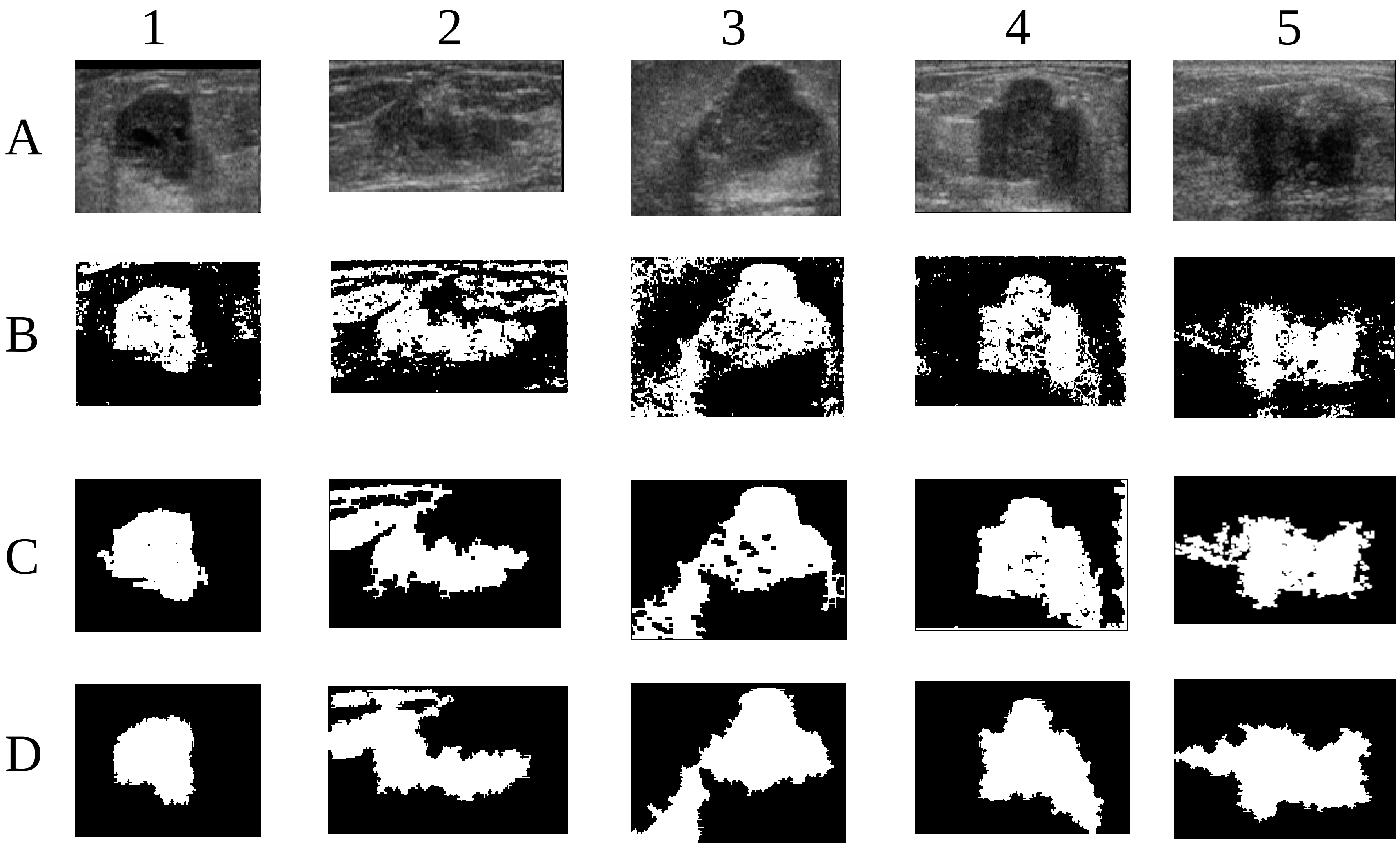
Some Examples for 5 benign Lesions



A = Original   B = Segmentação   C = Morphology   D = Level Set

The *First Three* Steps of the Proposed Methodology

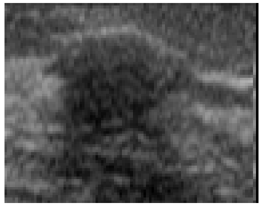
Some Examples for 5 Malignant Lesions

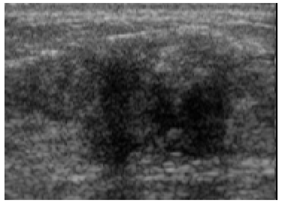


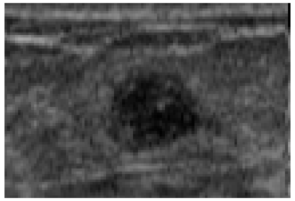
A = Original B = Segmentação C = Morphology D = Level Set

feature extraction

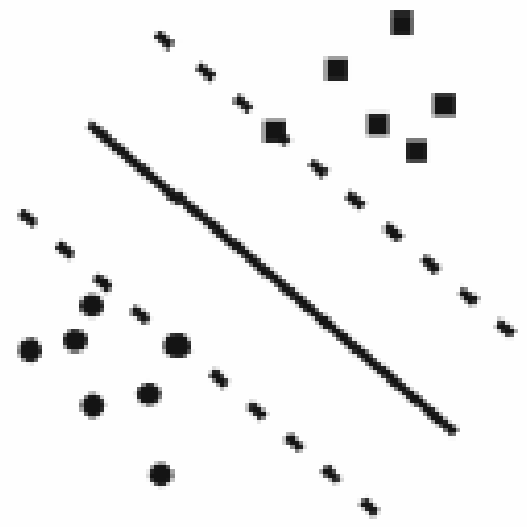
- Area
- Circularity
- Homegeneity
- Protuberance
- Acoustic Shadow







Support Vector Machine Classification



Separating Hyperplane

the dashed line identify the margin

$$f(x) = sgn\left(\sum_{i=1}^{N} \bar{\alpha}_i y_i K(x_i, y_i) + \bar{b}\right).$$

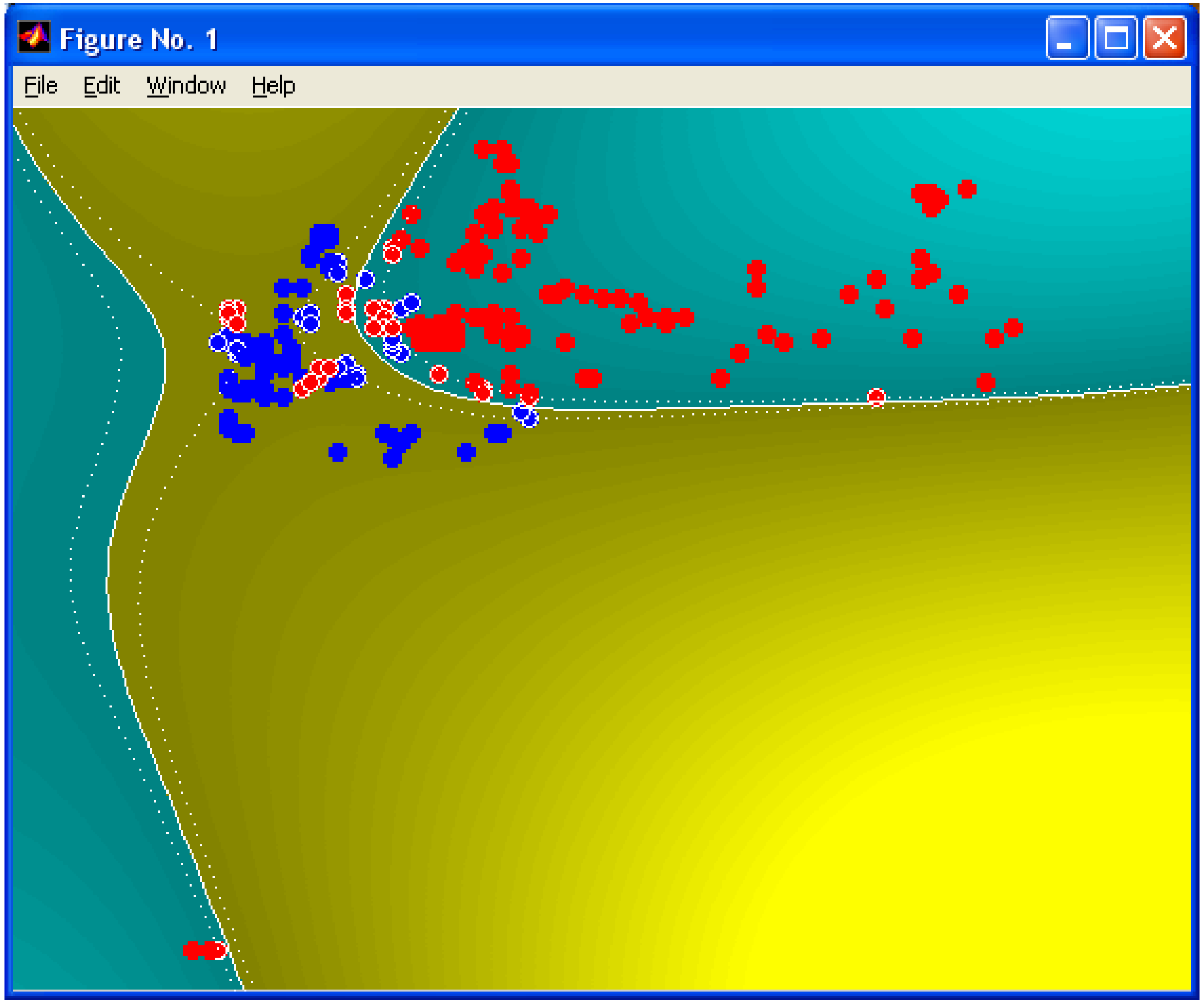
The *Fiveth* Step of the Proposed Methodology

The SVM classification

When we use a B-Spline kernel the results is the following:

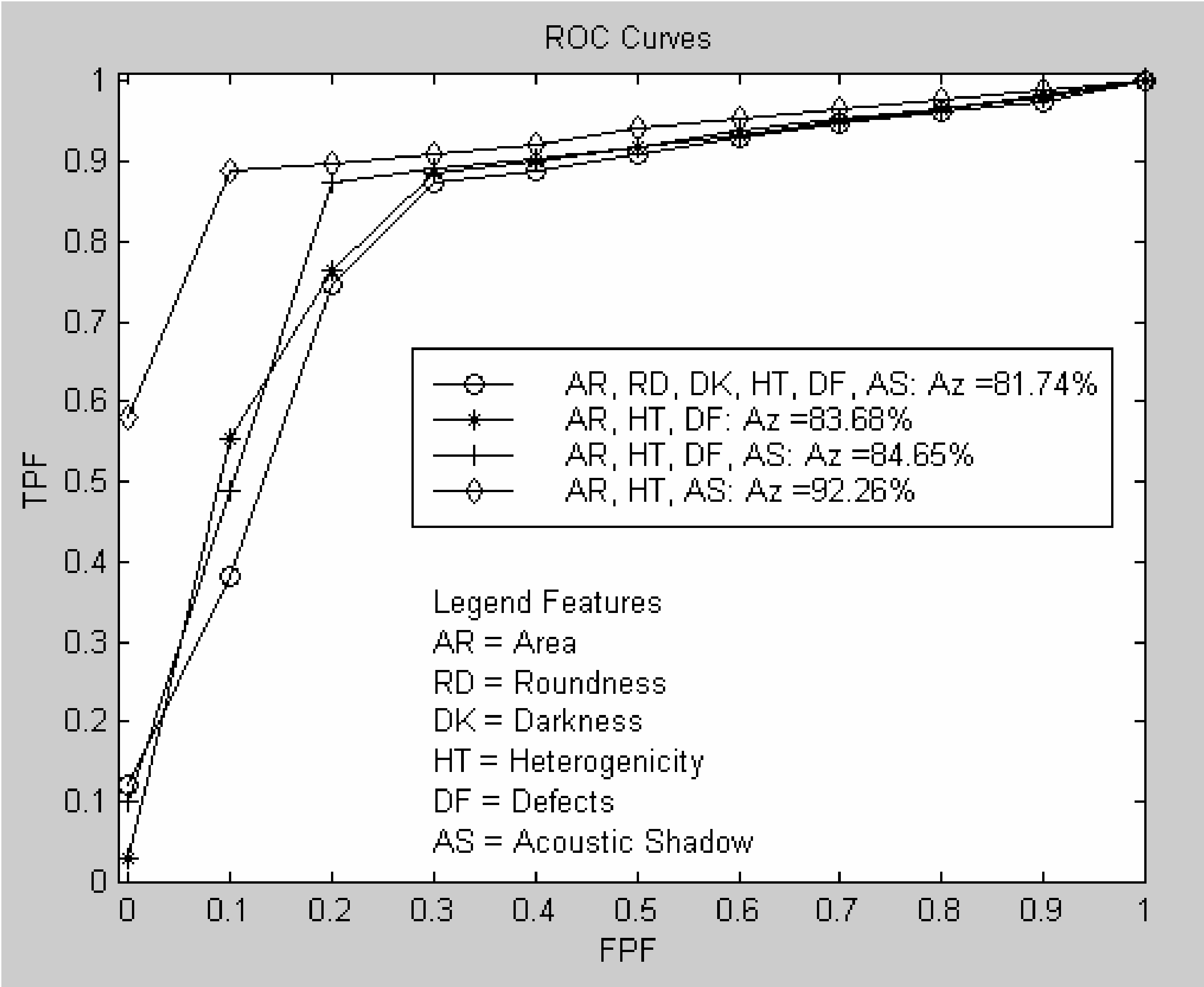
Test only for
Area and
Acoustic Shadow
Features

Conclusion: since
The data are not
Linear, the B-Spline
Kernel generates
A better separation



ROC Analysis

This tests were carried out for several features combinations



Analysis

O Trabalho de 2008, de Thomaz, Gilson e Rodrigues confirmou, utilizando Análise Discriminante Linear que as características mais discriminantes são **área, sombra acústica e heterogeneidade**

• A base analisada foi normalizada com relação à aquisição, tamanho, iluminação;

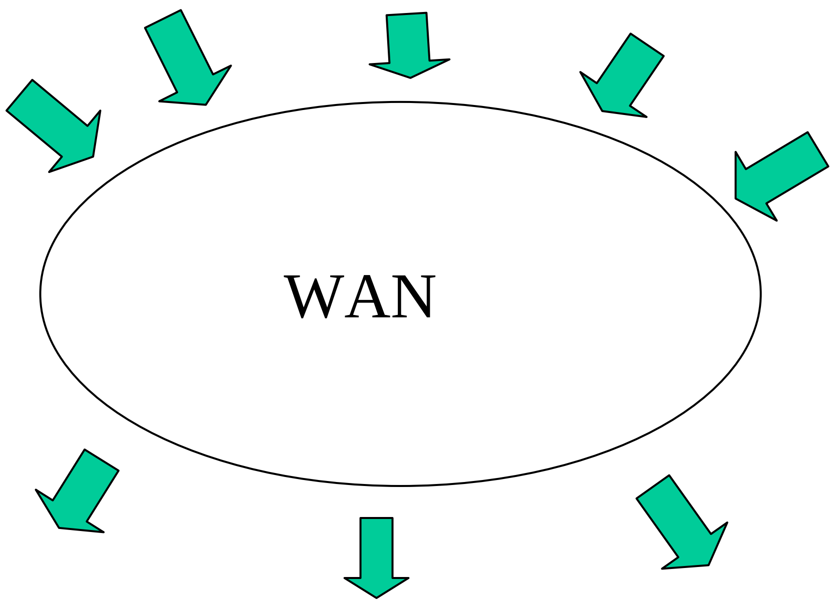
• possui 250 casos confirmados por biopsia, contendo 150 casos malignos e 100 benignos;

• Tem sido exaustivamente utilizada por pesquisadores do mundo todo para comparar e aperfeiçoar os sistemas DACs.

• No Brasil, existem sérias dificuldades de se encontrar bases de estudo

Detecção de Anomalias em Redes

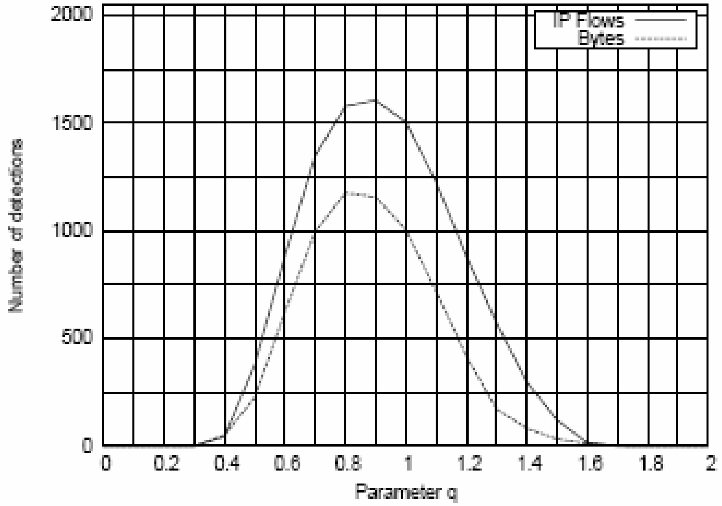
CC – Concentrated origin and concentrated destination;
CD – Concentrated origin and dispersed destination;
DC – Dispersed origin and concentrated destination;
DD – Dispersed origin and dispersed destination.



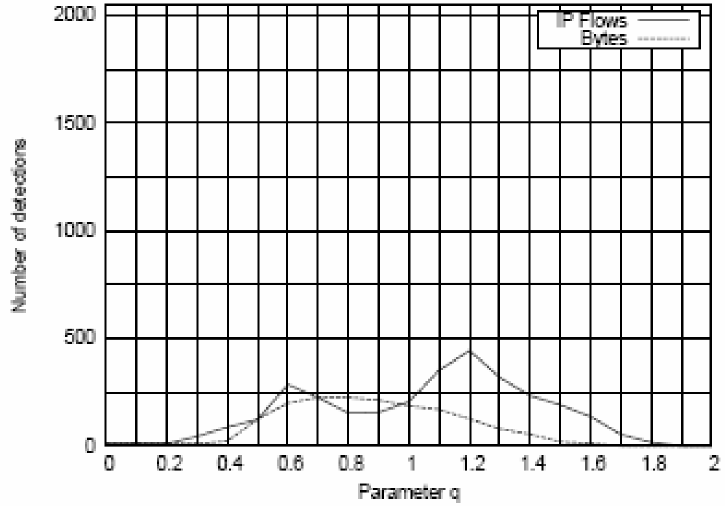
Detecção de Anomalias em Redes

Table 1: Performance comparison considering the number of detected anomalies.

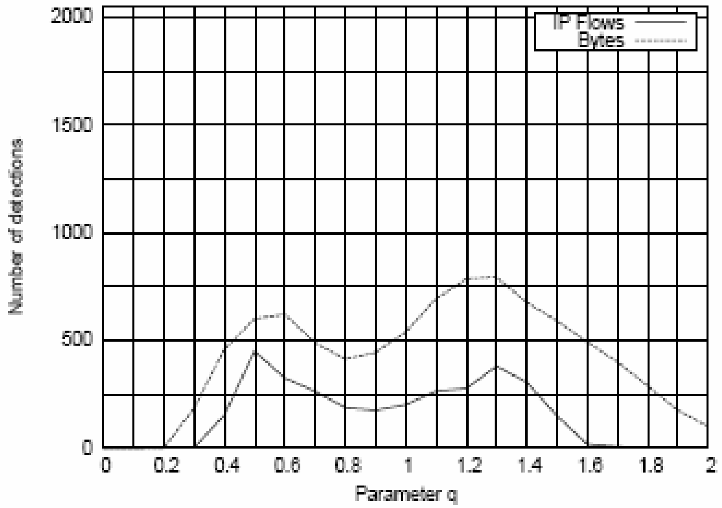
Dataset	Metric	Traffic Pattern	Number of detected anomalies H_S	$H_{q_{\text{optimal}}}$	Improvement (%)
Abilene	IP Flows	CC	1505	1606	6%
	IP Flows	CD	206	444	115%
	IP Flows	DC	201	450	124%
Abilene	Bytes	CC	1005	1177	17%
	Bytes	CD	186	226	22%
	Bytes	DC	539	793	47%
Géant	IP Flows	CC	6970	9964	43%
	IP Flows	CD	1838	5364	192%
	IP Flows	DC	1090	1582	45%



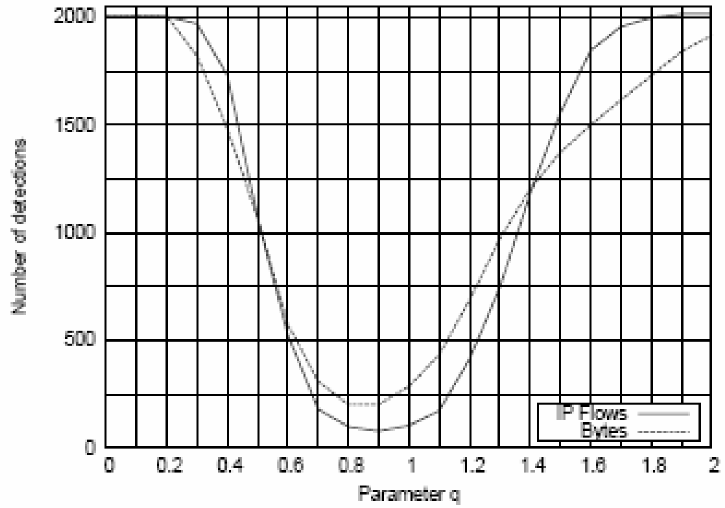
(a) CC Traffic



(b) CD Traffic

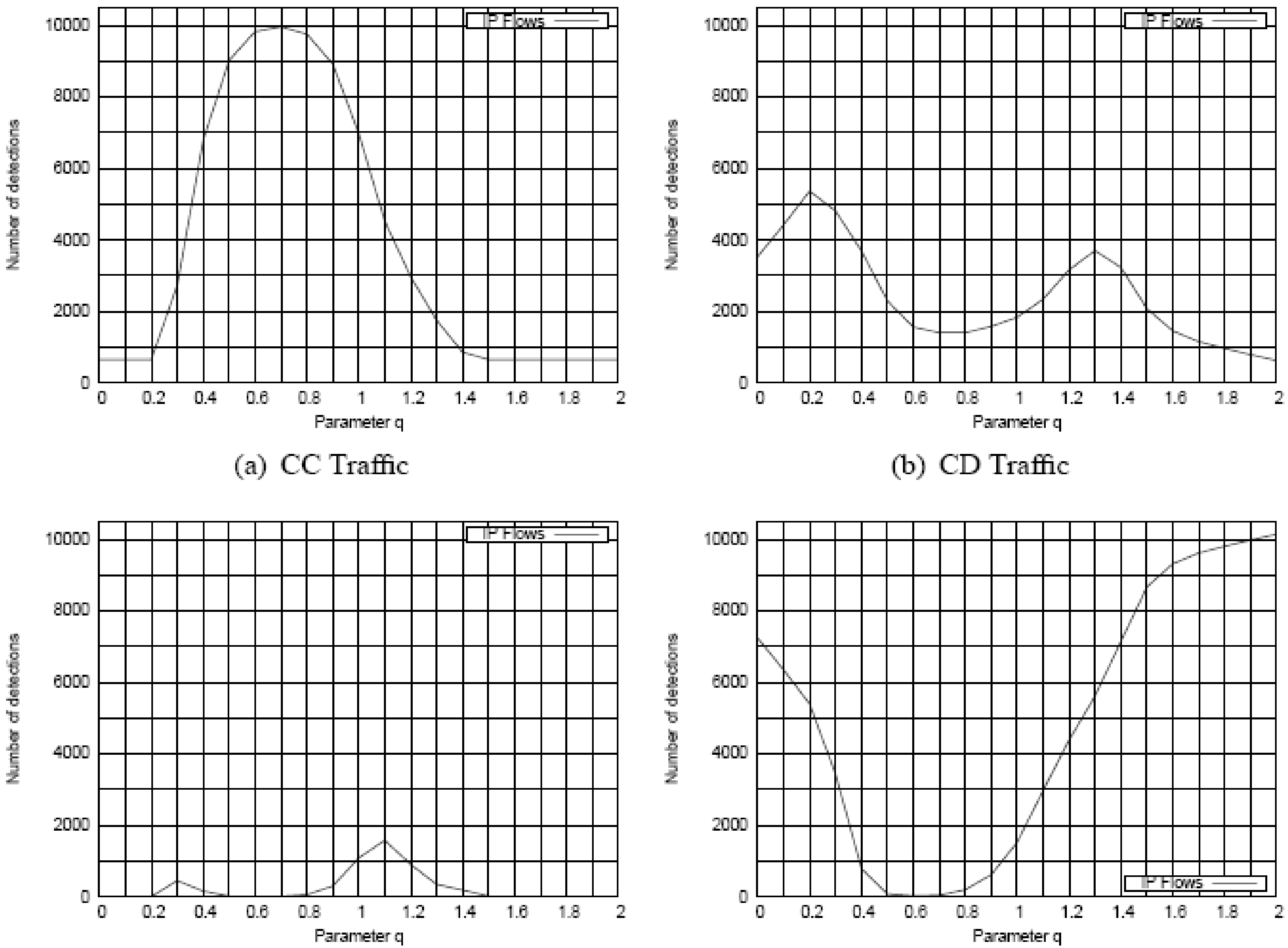


(c) DC Traffic



(d) DD Traffic

Figure 4: Number of detections of each traffic pattern in the Abilene dataset.



(a) CC Traffic

(b) CD Traffic

(c) DC Traffic

(d) DD Traffic

Figure 5: Number of detections of each traffic pattern in the Géant dataset.

Final Comments

- Interaction with other institutions and INCT

- Other topics of interest:

  - Image Registration
  - Pattern Recognition and Classification