New features for classification of cancerous masses in mammograms based on morphological dilation

Konrad Bojar
kbojar@ippt.gov.pl

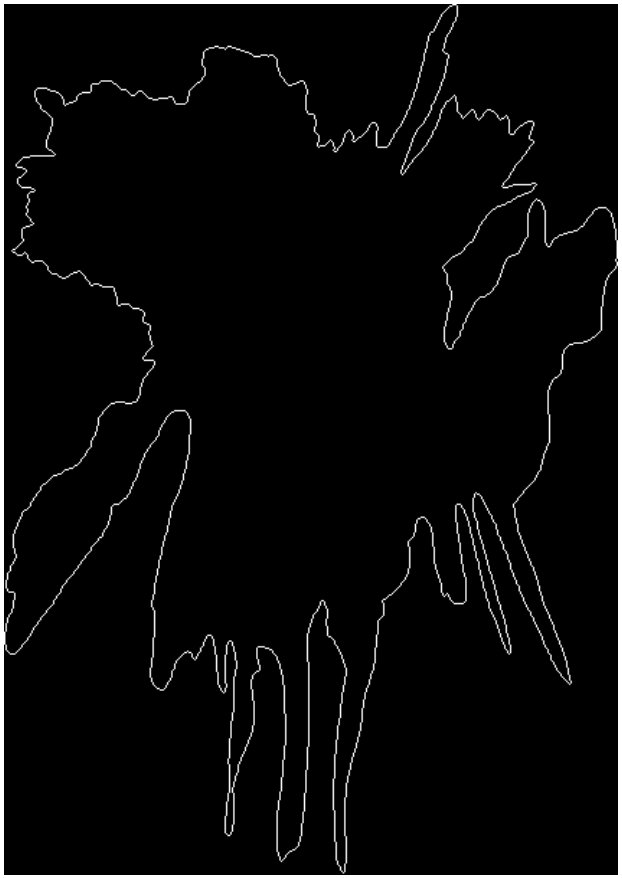
Mariusz Nieniewski
mnieniew@ippt.gov.pl

Institute of Fundamental Technological Research
Polish Academy of Sciences

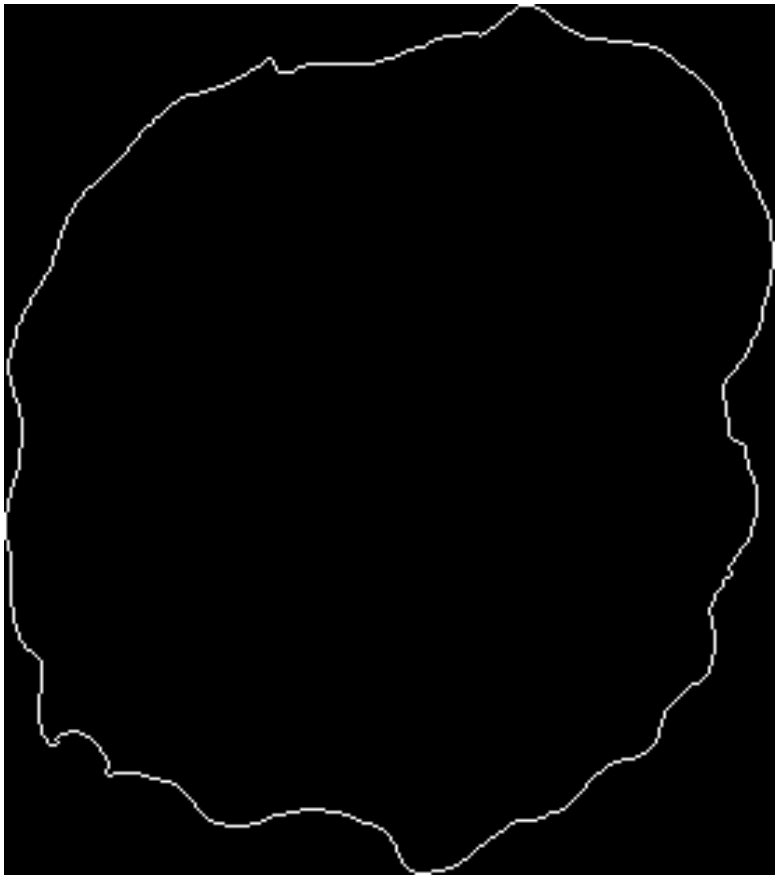
Presentation outline

- Typical contours of masses in mammograms
- Dilations of such contours: the deviation function
- Proposed spicularity/malignancy features
- Existing spicularity/malignancy features
- Conducted experiments
- Conclusions

Benign contours

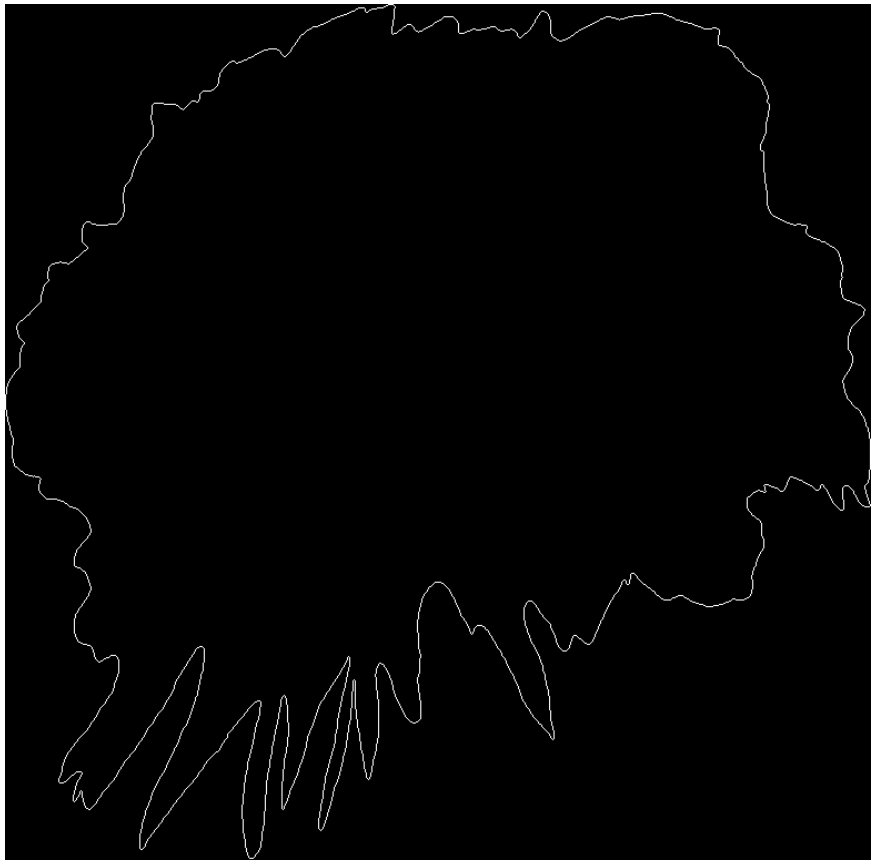


MDB127LM



MDB069LL

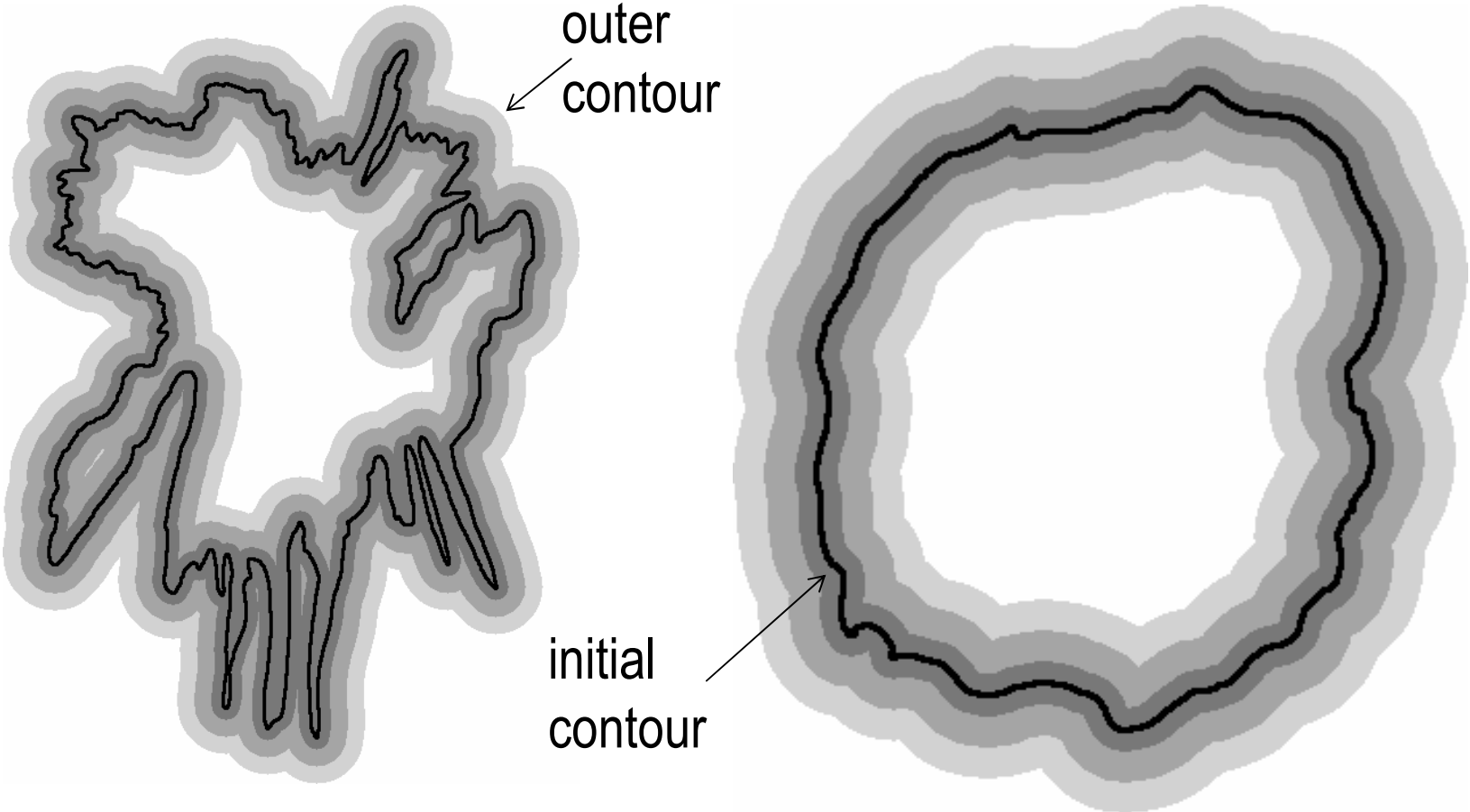
Malignant contours



MDB184RL

MDB120LL

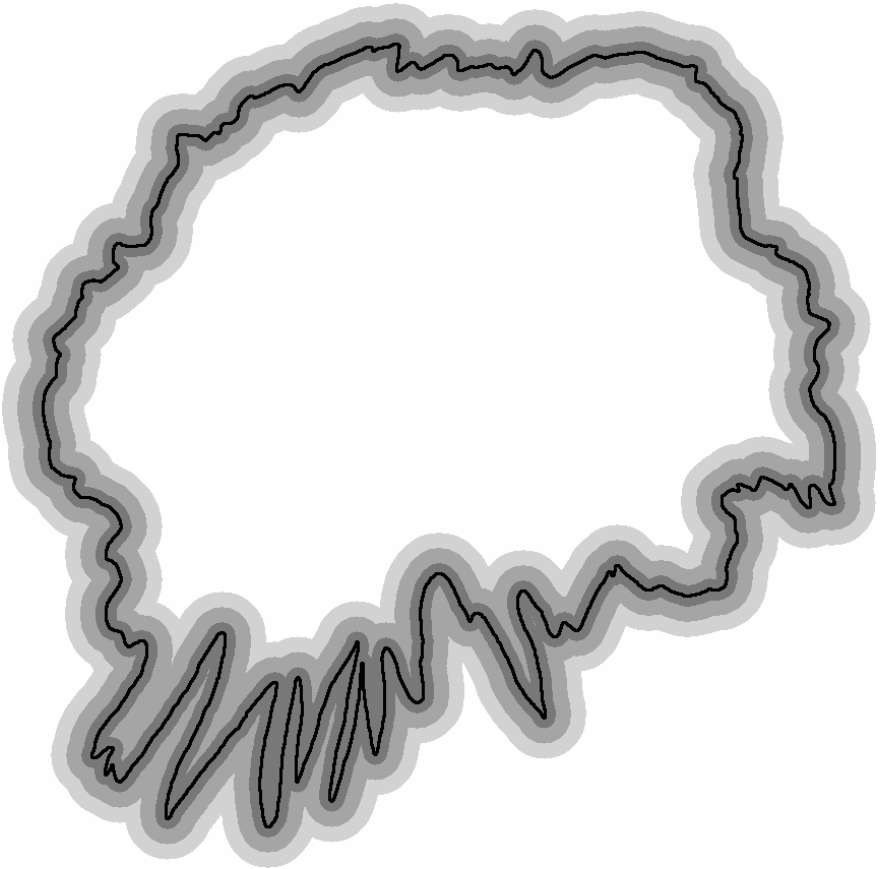
Dilations of benign contours



MDB127LM

MDB069LL

Dilations of malignant contours



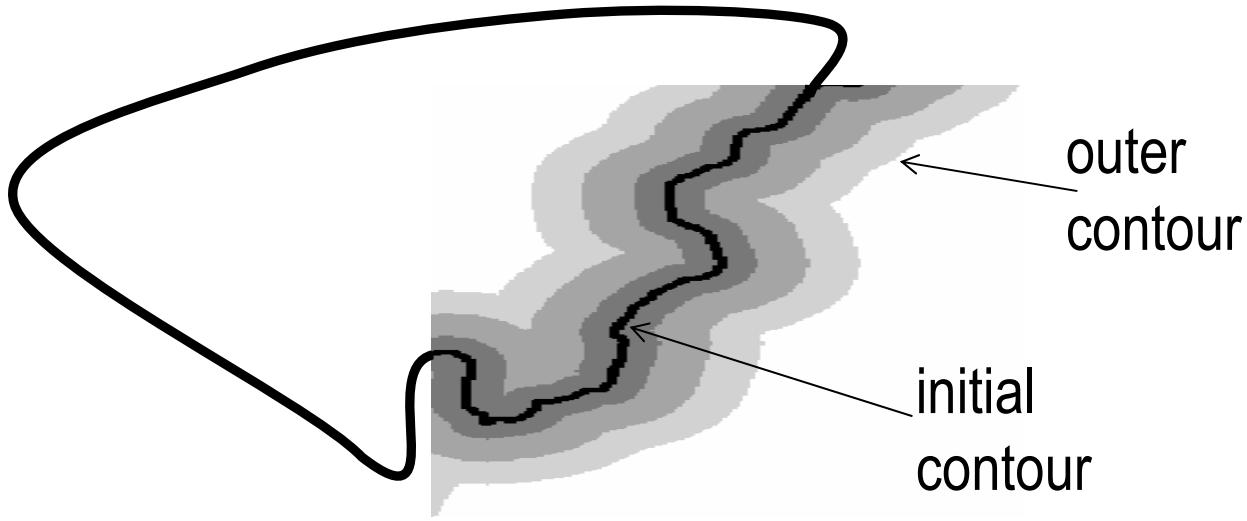
MDB184RL

MDB120LL

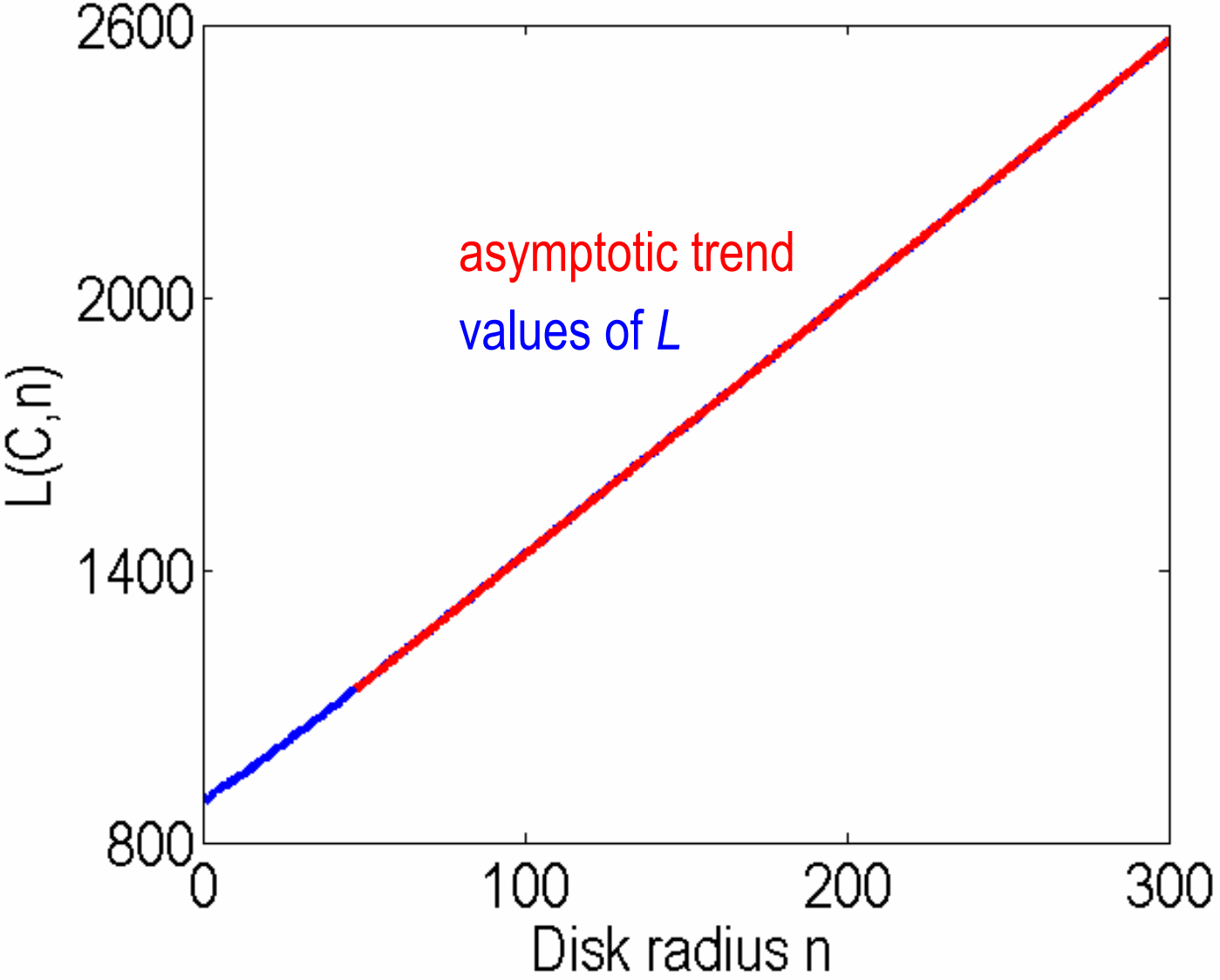
Characteristic function *L*

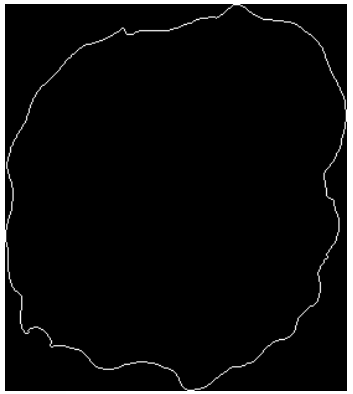
- Characterizes complexity of a given contour
- Reflects smoothing properties of dilation

$$L(C,n)=\begin{cases}\text{No. of pixels on the } \textbf{initial} \text{ contour } C & n=0\\ \text{No. of pixels on the } n\text{-th } \textbf{outer} \text{ contour } C & n>0\end{cases}$$



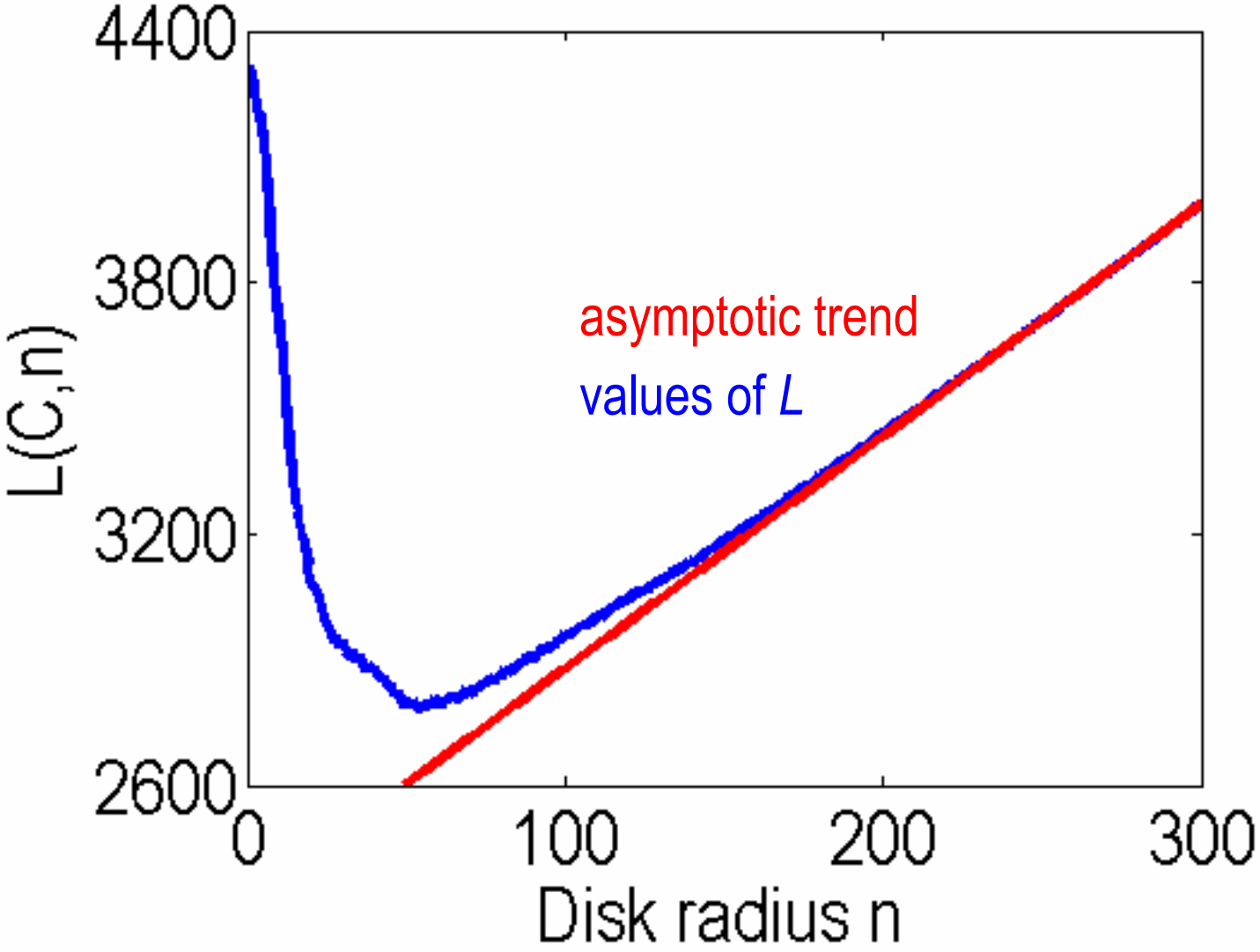
Characteristic function *L*

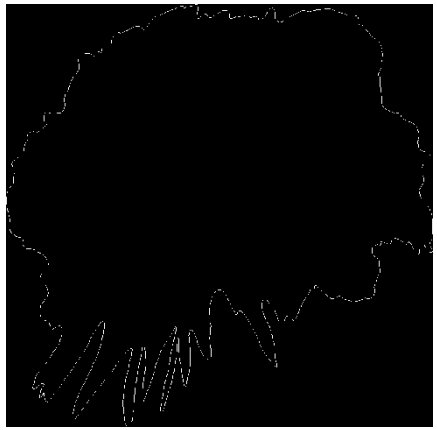




MDB069LL

Characteristic function *L*





MDB184RL

Properties of L

- Almost linear for convex contours
- Substantial initial fall-off for spicular contours
- $L(C,n) \sim a_C n + b_C$ for large n's, a_C, b_C – some consts
- $L(\alpha C,n) = \alpha L(C, n/\alpha)$ (up to noise)
- L is rotation- and shift- invariant (up to noise)
- **Most important part of L: small n's**

Deviation function *Dev*

- $Dev(C,n)=L(C,n) - a_C n - b_C$
- Estimate a_C and b_C along with computations of L

  ⇨Use least squares fitting and last 50 values of L

  ⇨Stop criterion: average square deviation falls below $10^{-4}\ L(C,0)$

Properties of *Dev*

- Close to zero for contours of low curvature
- $Dev(\alpha C,n)=\alpha Dev(C,n/\alpha)$ (up to noise)
- *Dev* is rotation- and shift- invariant (up to noise)
- Converges to zero for all contours

  ⇨First denoise *Dev* function

  ⇨Define cut-off value N_C to stop computation

Cut-off value N_C and smoothing

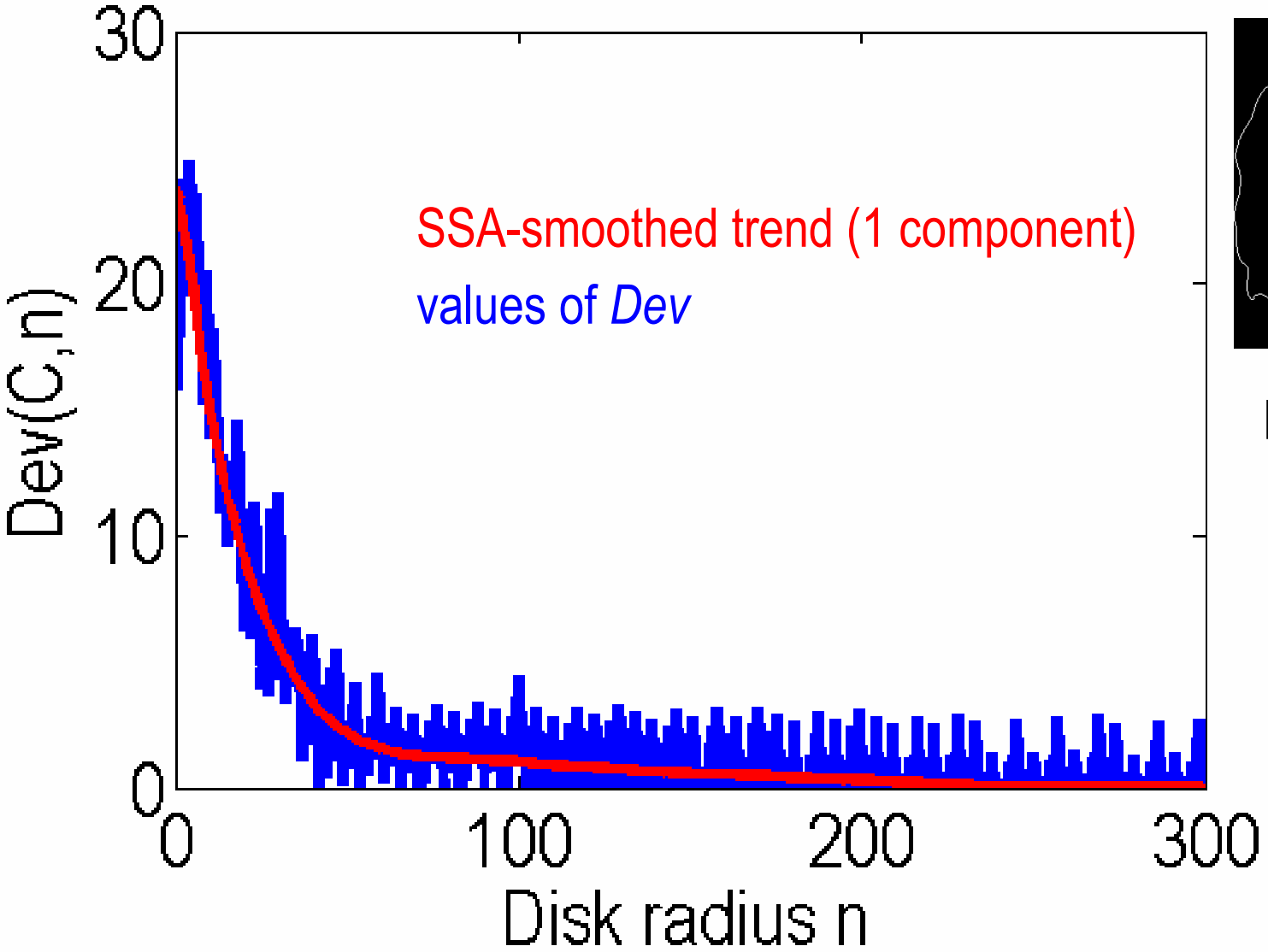
- Smoothing using first SSA (Singular Spectrum Analysis)-based components

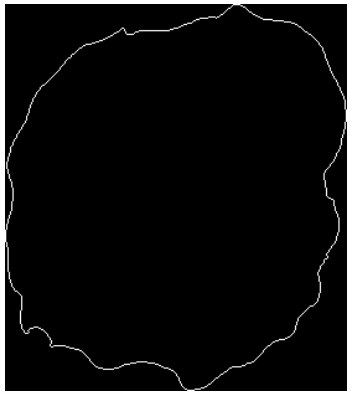
  ⇨until average oscillation (absolute value of derivative) of smoothed *Dev* falls below 0.4

- N_C : smoothed *Dev* falls below 0.1 of its value at 0

  ⇨Scaling property: $N_{\alpha C} = \alpha N_C$

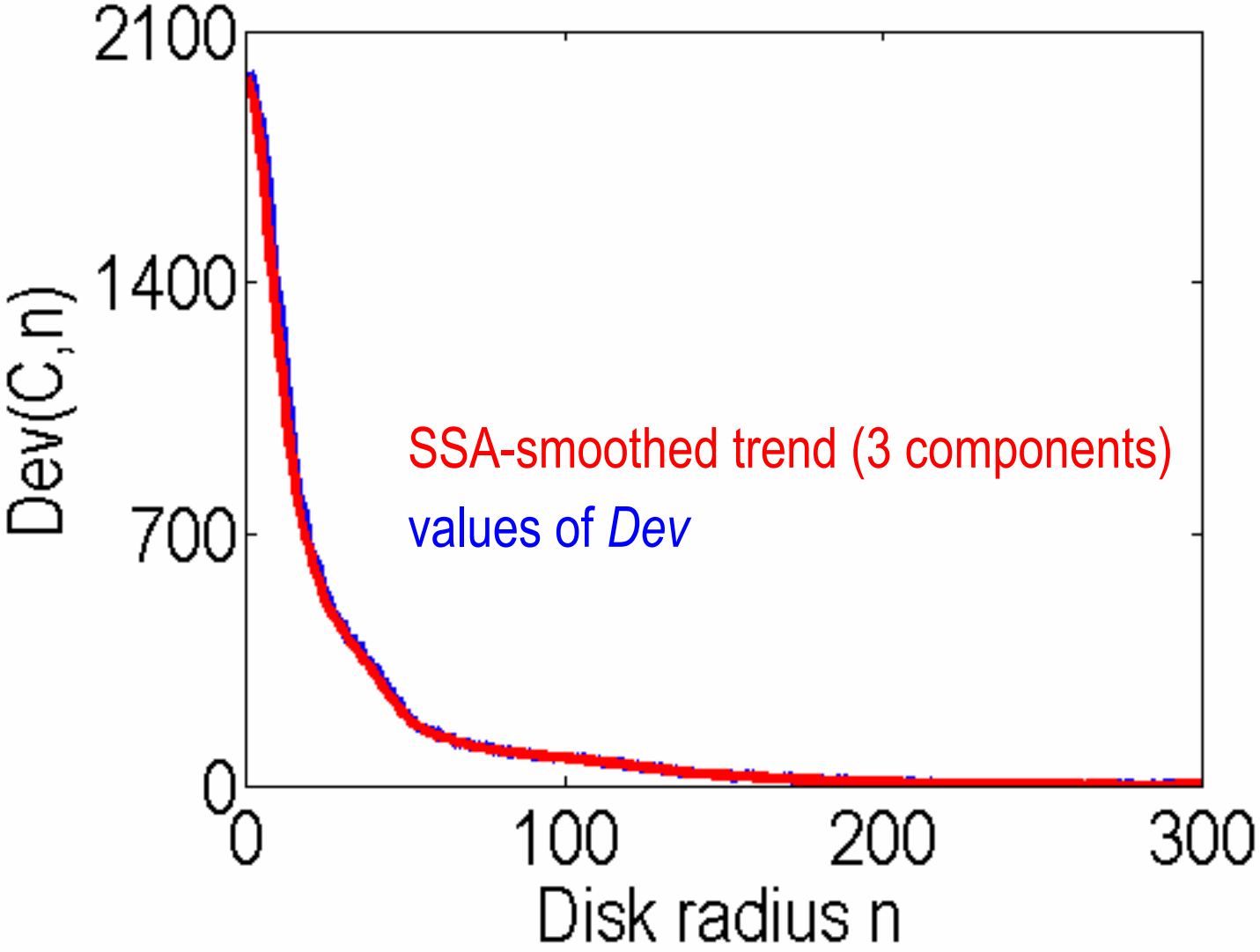
Deviation function *Dev*

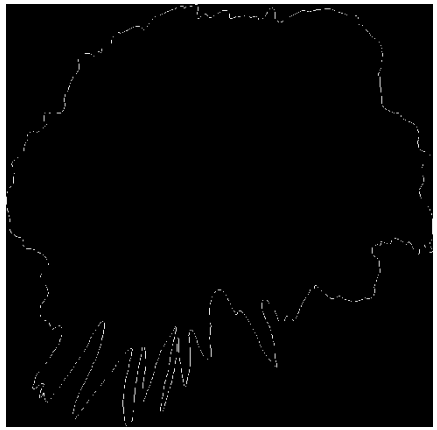




MDB069LL

Deviation function *Dev*





MDB184RL

Proposed features

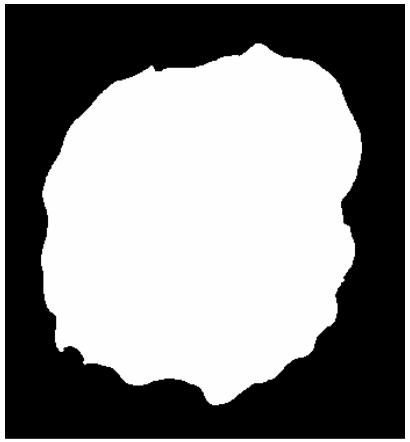
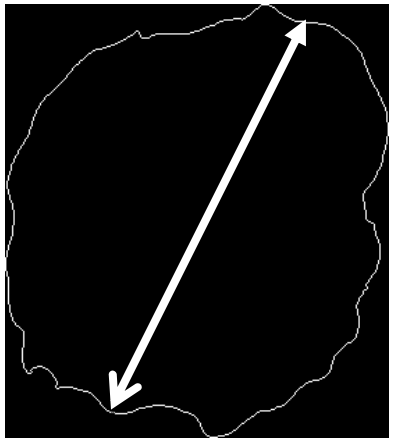
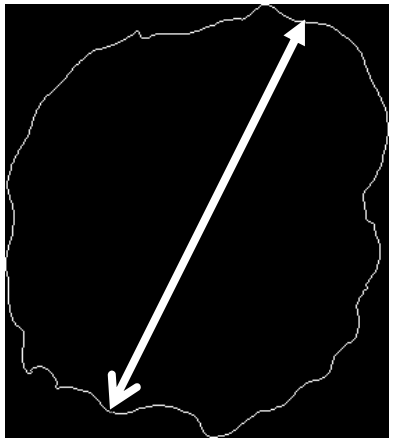
$$D = \frac{Dev(C,0)}{N_C}$$ Average deviation

$$A = \frac{\sum_{n=0}^{N_C} Dev(C,n)}{N_C Dev(C,0)}$$ Relative total deviation

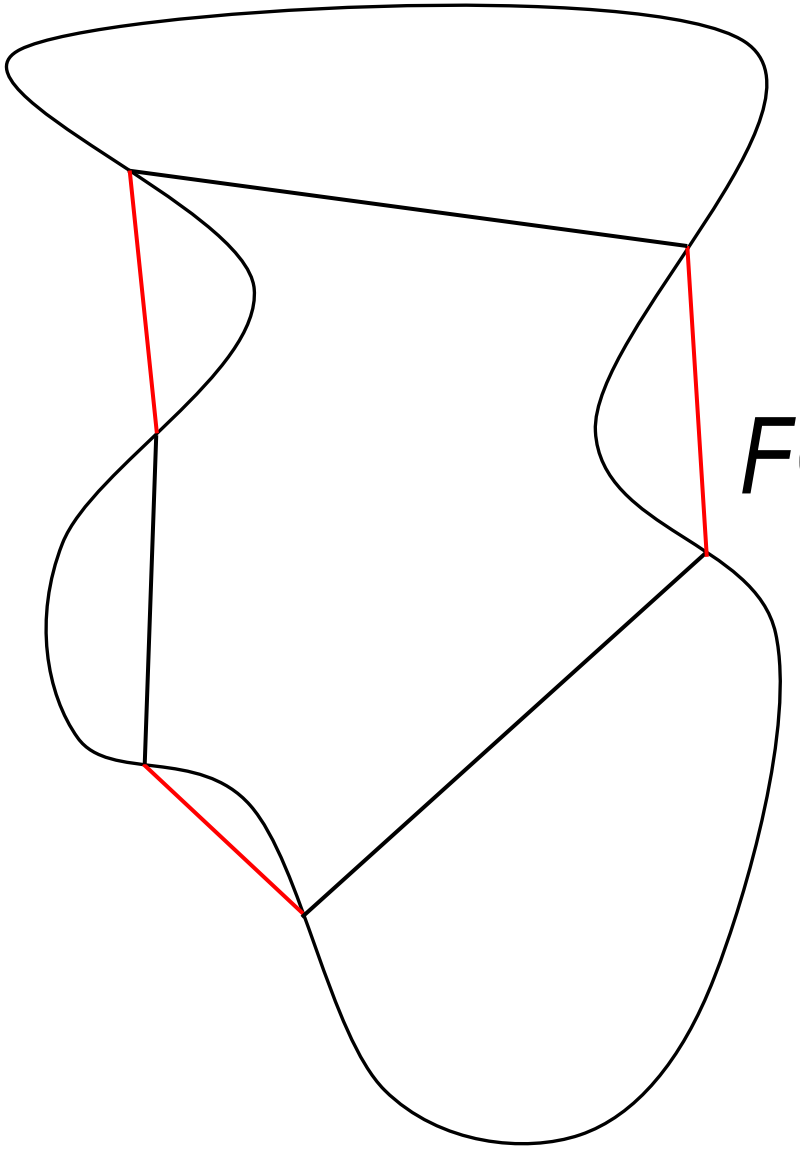
$$S = \sum_{n=0}^{N_C} \frac{Dev(C,n)}{n^2}$$ Small deviation emphasis

Scale, rotation and shift invariant!

Existing features: Compactness

$$CO = 1 - \frac{4\pi * \text{  }}{\text{  } * \text{  }}$$

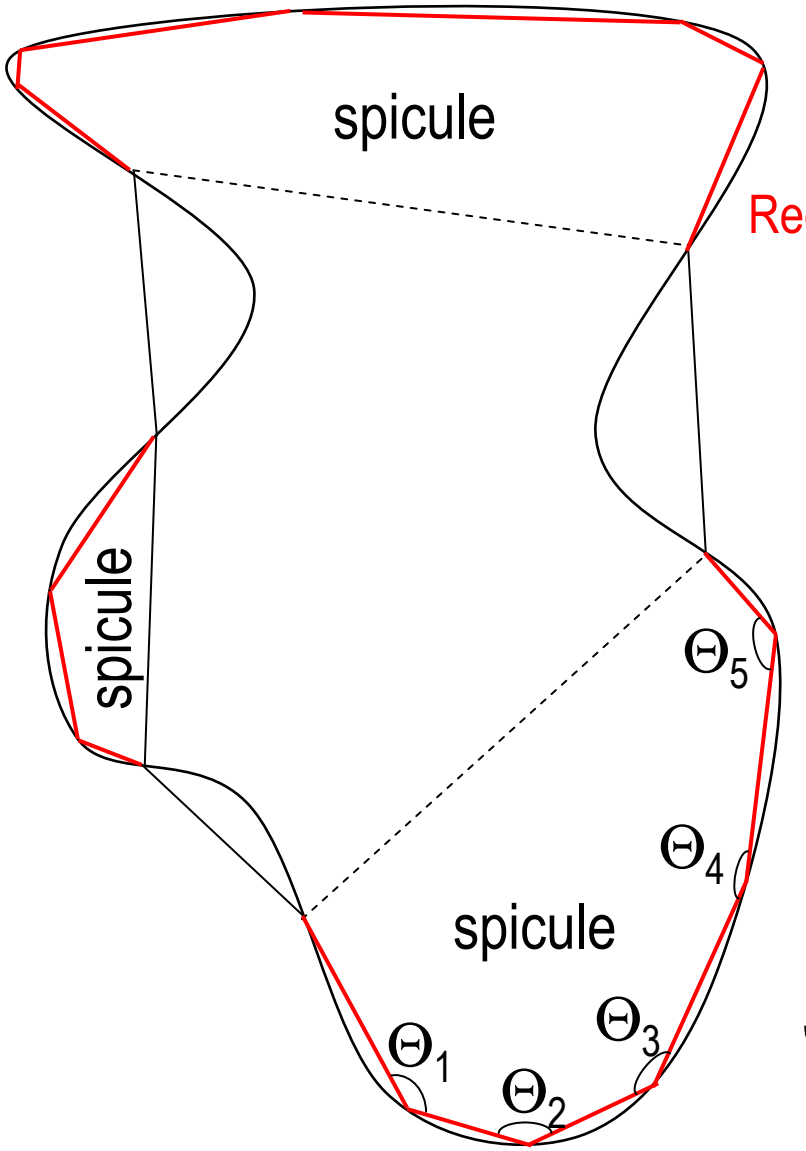
Existing features: Fractional Concavity



Segments join inflection points

$$FC = \frac{\text{length of red segments}}{\text{len. of red segm.} + \text{len. of black segm.}}$$

Existing features: Spicularity Index



Black segments join inflection points

Red segments approximate spicules (convex parts)

Length of *i*-th spicule:

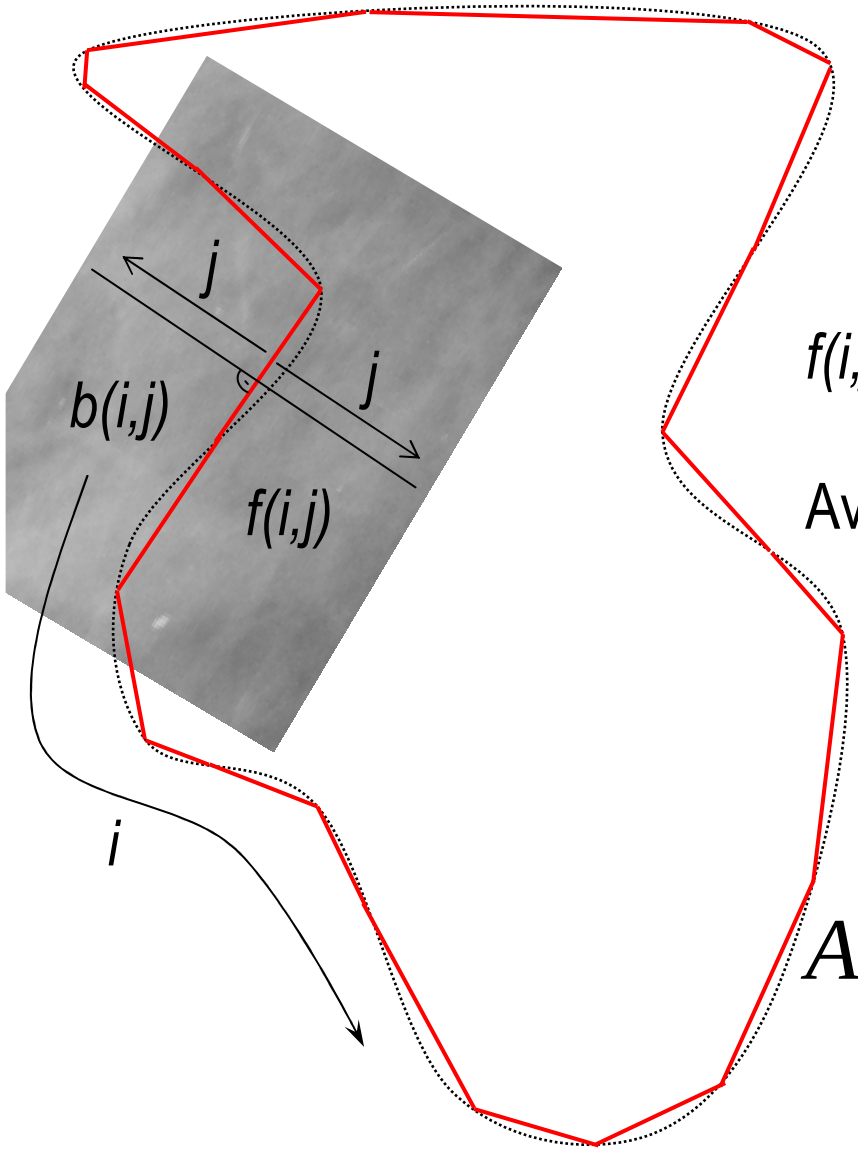
S_i = length of its red segments

Narrowness of *i*-th spicule:

$\theta_i = \text{mean}(\{\Theta_i : \Theta_i \leq \text{mean}(\Theta_i)\})$

$$SI = \frac{\sum_{i \in \text{spicules}} (1+\cos\theta_i) S_i}{\sum_{i \in \text{spicules}} S_i}$$

Existing features: Acutance



Red segments approximate the black contour

b(i,j) – outer grayvalues

f(i,j) – inner grayvalues

Average strength of normal gradient:

$$d(i) = \sum_j \frac{|\, f(i,j) - b(i,j) \,|}{2j}$$

$$AC = \frac{1}{\max d} \sqrt{\frac{1}{\text{red segments length}} \sum_i d^2(i)}$$

Experiments

- All 86 statistically independent masses (MIAS)
- 86 = 50 benign + 36 malignant
- 86 = 69 non-spiculated + 19 spiculated
- 23 feature sets including *D, A, S, CO, FC, SI, AC*
- Classifier: *Fisher Linear Discriminant*
- Classification: *malignant vs. benign,*
  spiculated vs. non-spiculated
- Evaluation: *ROC* and $\boldsymbol{A_z}$

Results in terms of $A_z \pm \sigma_{A_z}$

Feature set	Malignant vs. Benign	Spiculated vs. Non-spiculated
D	0.660 ± 0.058	0.709 ± 0.068
A	0.541 ± 0.061	0.650 ± 0.072
S	0.682 ± 0.056	0.724 ± 0.073
D, A	0.599 ± 0.060	0.572 ± 0.074
D, S	0.644 ± 0.059	0.720 ± 0.072
A, S	0.570 ± 0.061	0.649 ± 0.072
D, A, S	0.597 ± 0.060	0.624 ± 0.073
CO	0.540 ± 0.061	0.573 ± 0.067
FC	0.612 ± 0.060	0.521 ± 0.071
SI	0.676 ± 0.057	0.632 ± 0.069
AC	0.611 ± 0.061	0.539 ± 0.077
CO, FC	0.594 ± 0.060	0.517 ± 0.068
CO, SI	0.588 ± 0.062	0.577 ± 0.065
CO, AC	0.522 ± 0.061	0.507 ± 0.076
FC, SI	0.516 ± 0.062	0.548 ± 0.071
FC, AC	0.610 ± 0.060	0.531 ± 0.073
SI, AC	0.514 ± 0.062	0.542 ± 0.077
CO, FC, SI	0.513 ± 0.061	0.522 ± 0.069
CO, FC, AC	0.581 ± 0.061	0.509 ± 0.070
CO, SI, AC	0.518 ± 0.061	0.526 ± 0.074
FC, SI, AC	0.507 ± 0.062	0.520 ± 0.074
CO, FC, SI, AC	0.597 ± 0.060	0.563 ± 0.061
all seven features	0.575 ± 0.601	0.605 ± 0.062

Conclusions

- Conceptual simplicity of the features proposed
- Coding simplicity of the features proposed
- High discriminatory power
- Translation, rotation, and scale invariance

The end