Image segmentation applied to cytology

Niels VAN VLIET <niels@lrde.epita.fr>

LRDE seminar, May 14, 2003

Table of contents

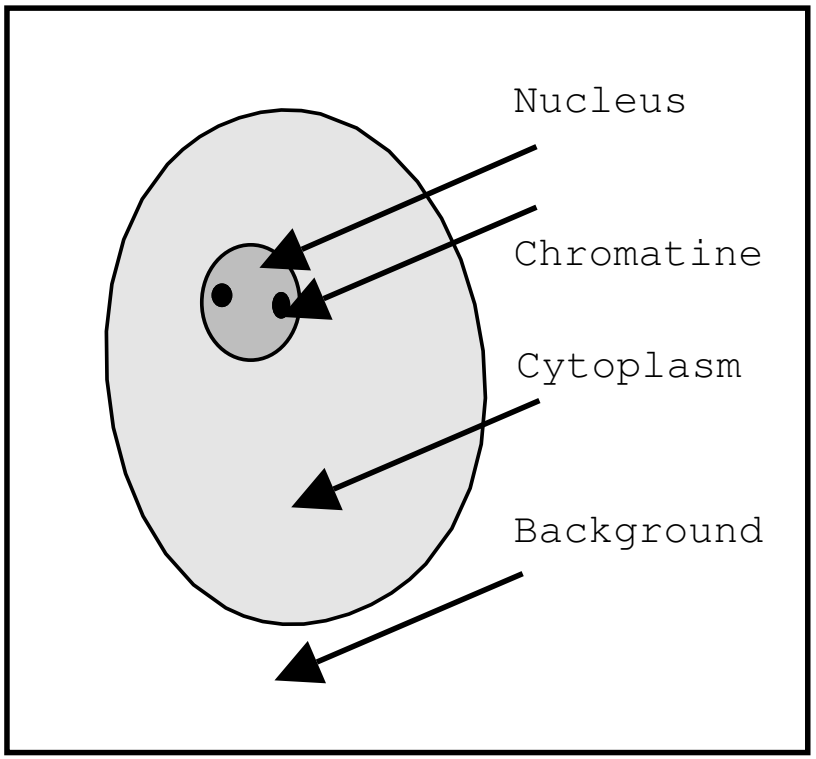
Introduction

The Automatic Detection of Healthy Or Cancerous Cells (**AD-HOC**) is divided into two parts:

- Extraction of the data from the image
- Analysis of the data (future work)

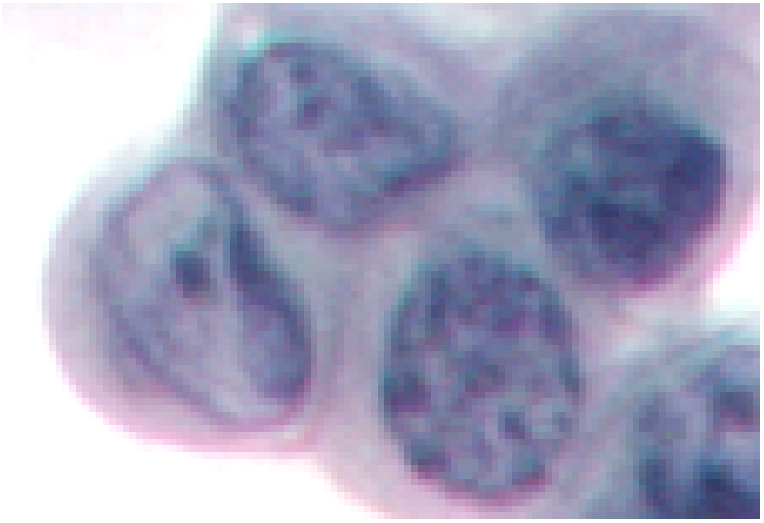
A few definitions

What is a (healthy) cell ?



- Nucleus ' diameter $\approx 10\mu$
- The nucleus 's boundary is **regular**
- The nucleus is **round** ($\neq$ oval!)
- Nucleus darker than the cytoplasm
- Cytoplasm darker than the background
- Cytoplasm **much bigger** than the nucleus

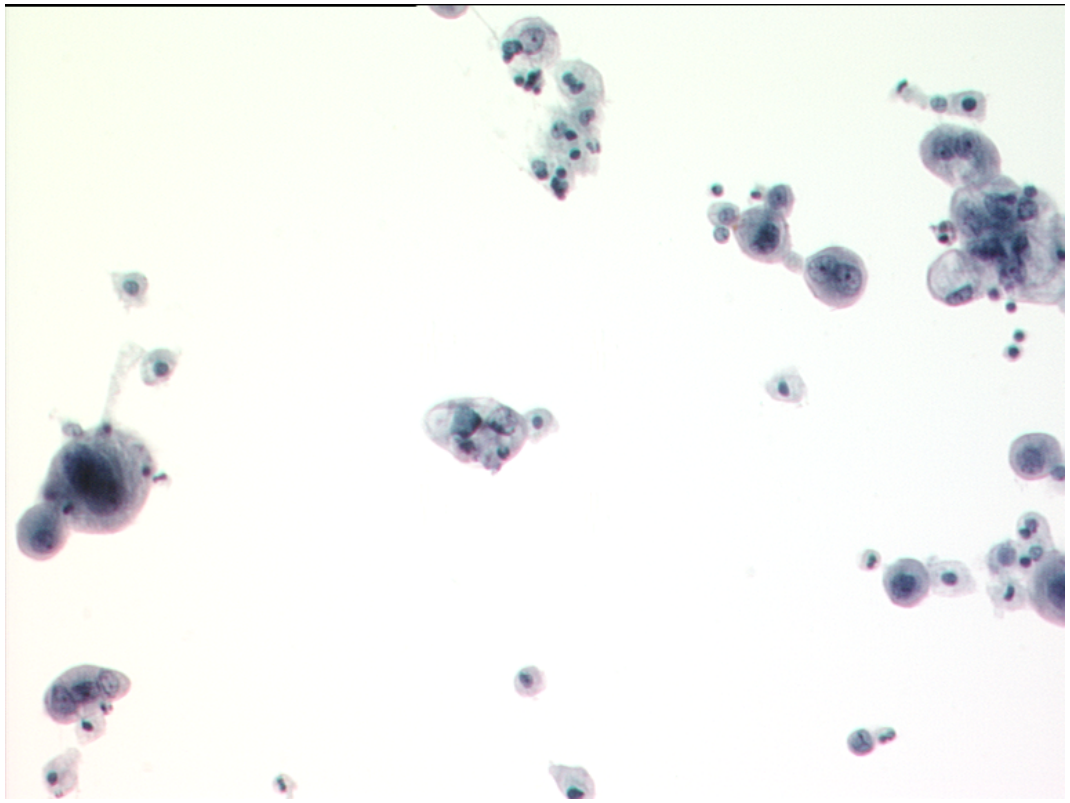
What is a cancerous cell ?



- $sizeof(nucleus)/sizeof(cytoplasm)$ is big
- Nucleus ' diameter $> 13\mu$
- **Dark** nucleus
- **Irregular** shape

A spot

A 2 cm spot magnified 400x:



Origin of the images

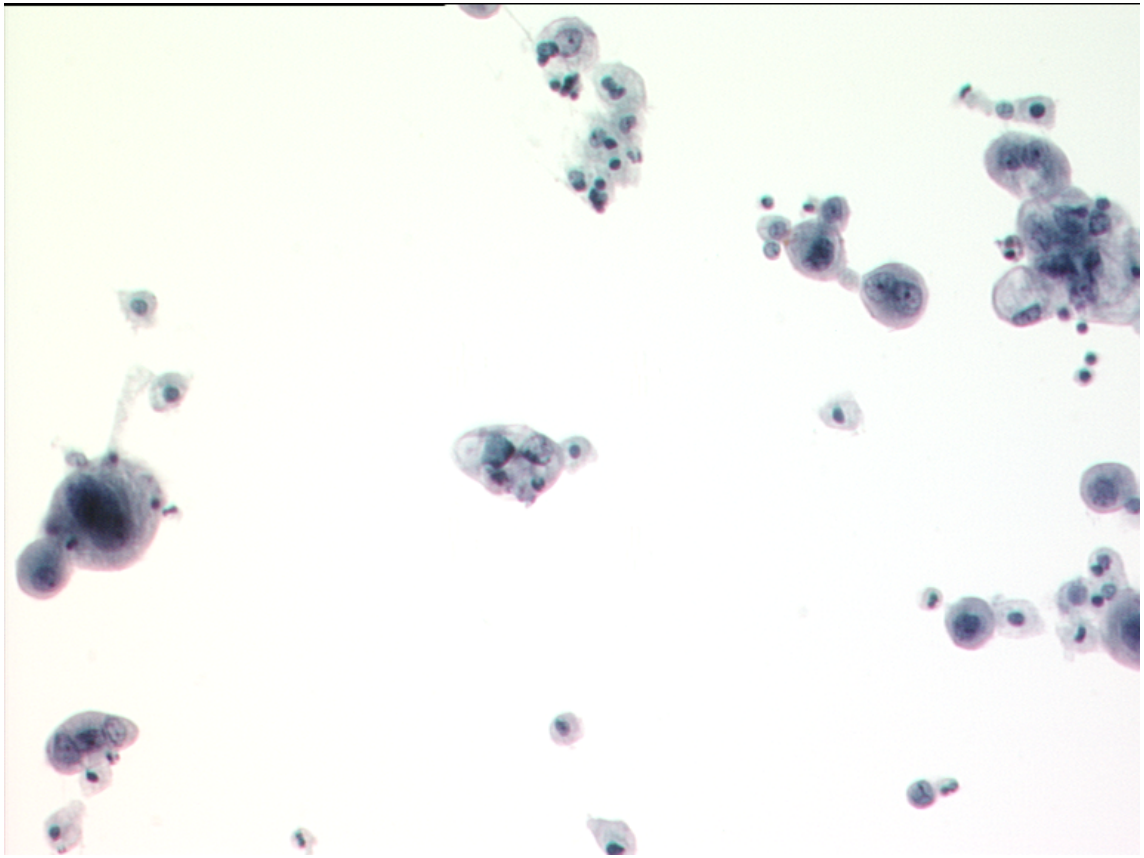
How to create a **spot** ?

- Fine needle aspiration
- Chemical destruction of useless objects
- Separation of the cells in a bath
- Centrifugation
- Extraction of the cells sticked on the sides by a centrifuge

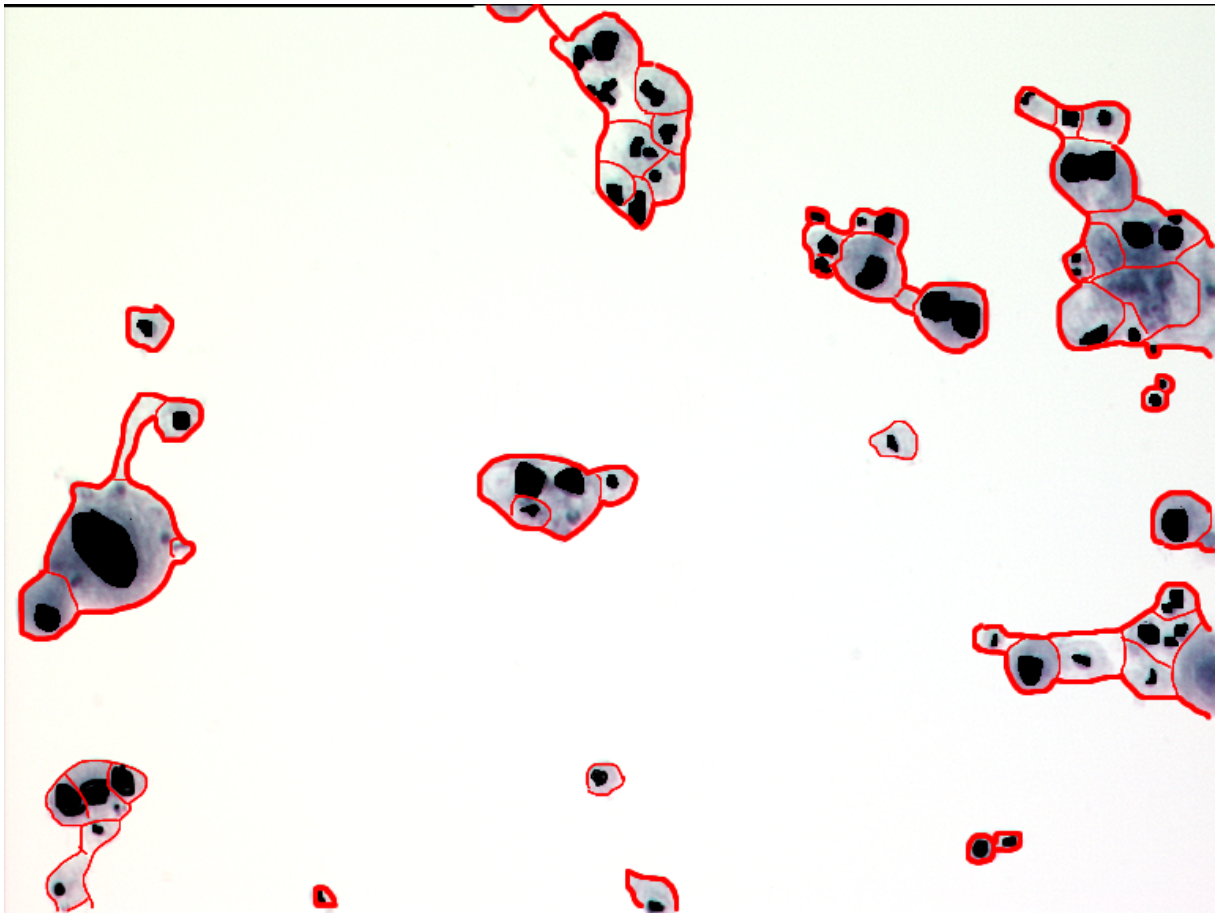
Problems

- Problems of the screening:
  - Slow
  - Harmful
  - Subjective
- Solution: **Automation**
  - Segmentation
  - Decision

Input

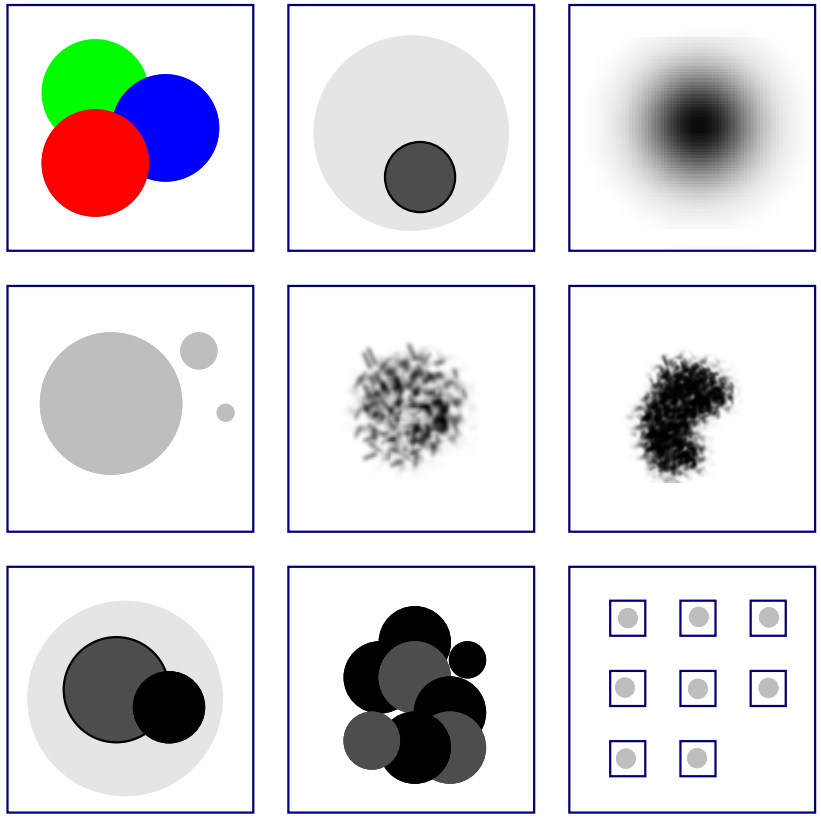


Output of the segmentation

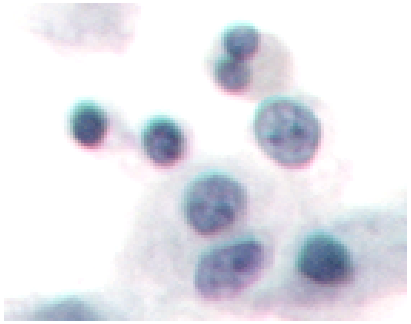


Problems encountered

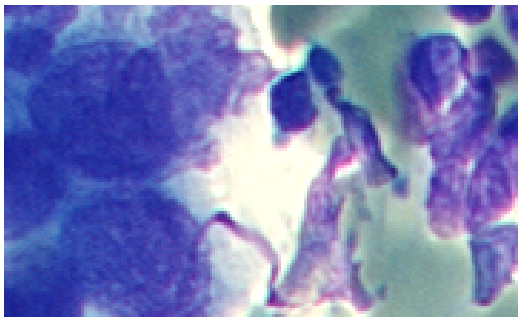
9 problems are going to be presented:



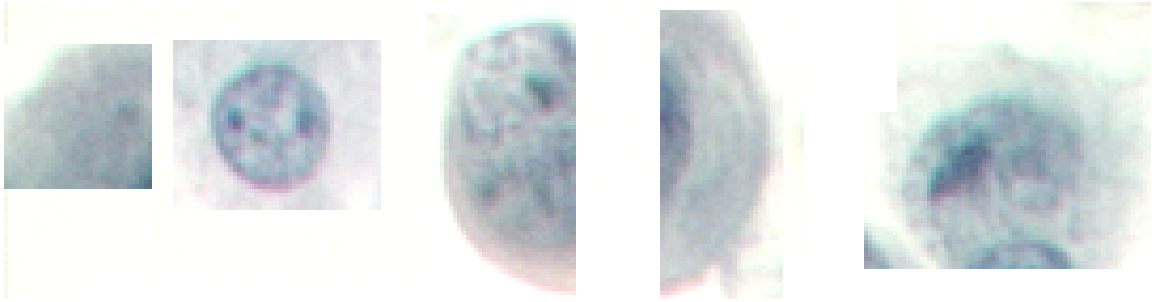
1. No color:



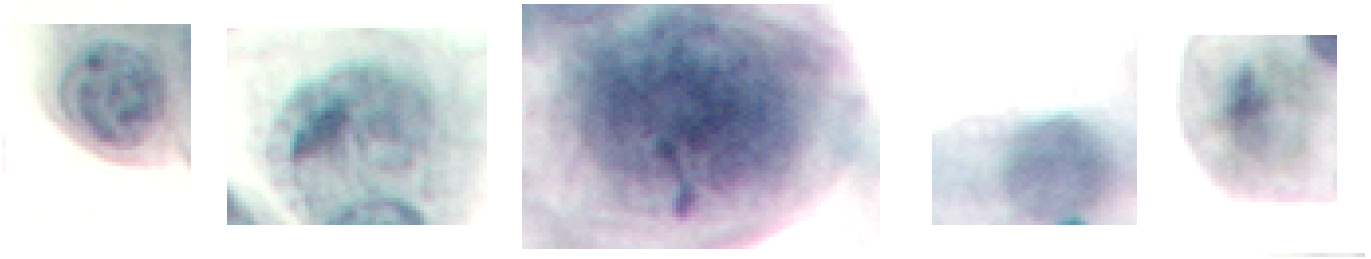
Normal case



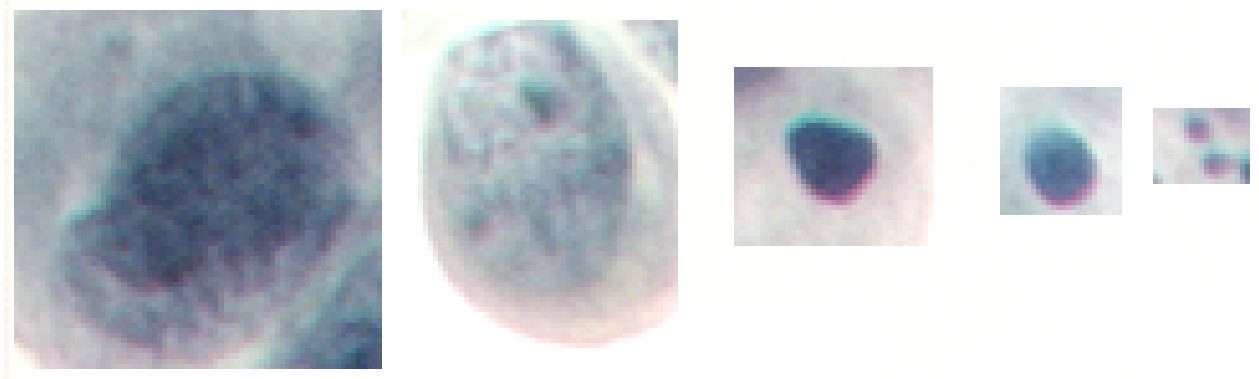
2. Problems of contrast:



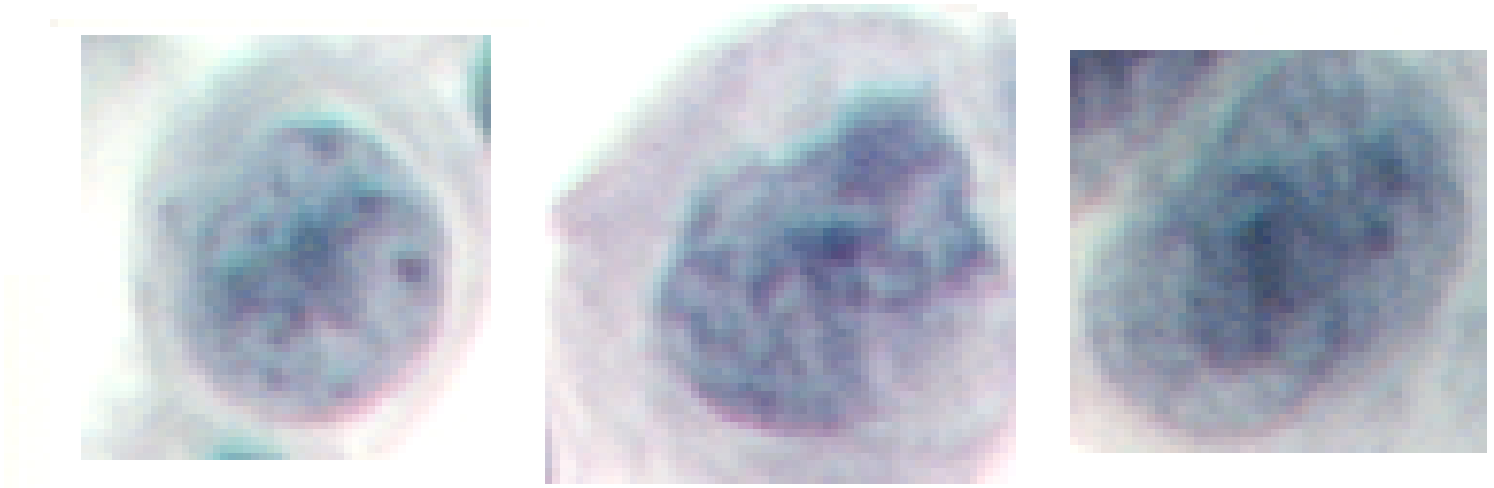
3. Fuzzy cells:



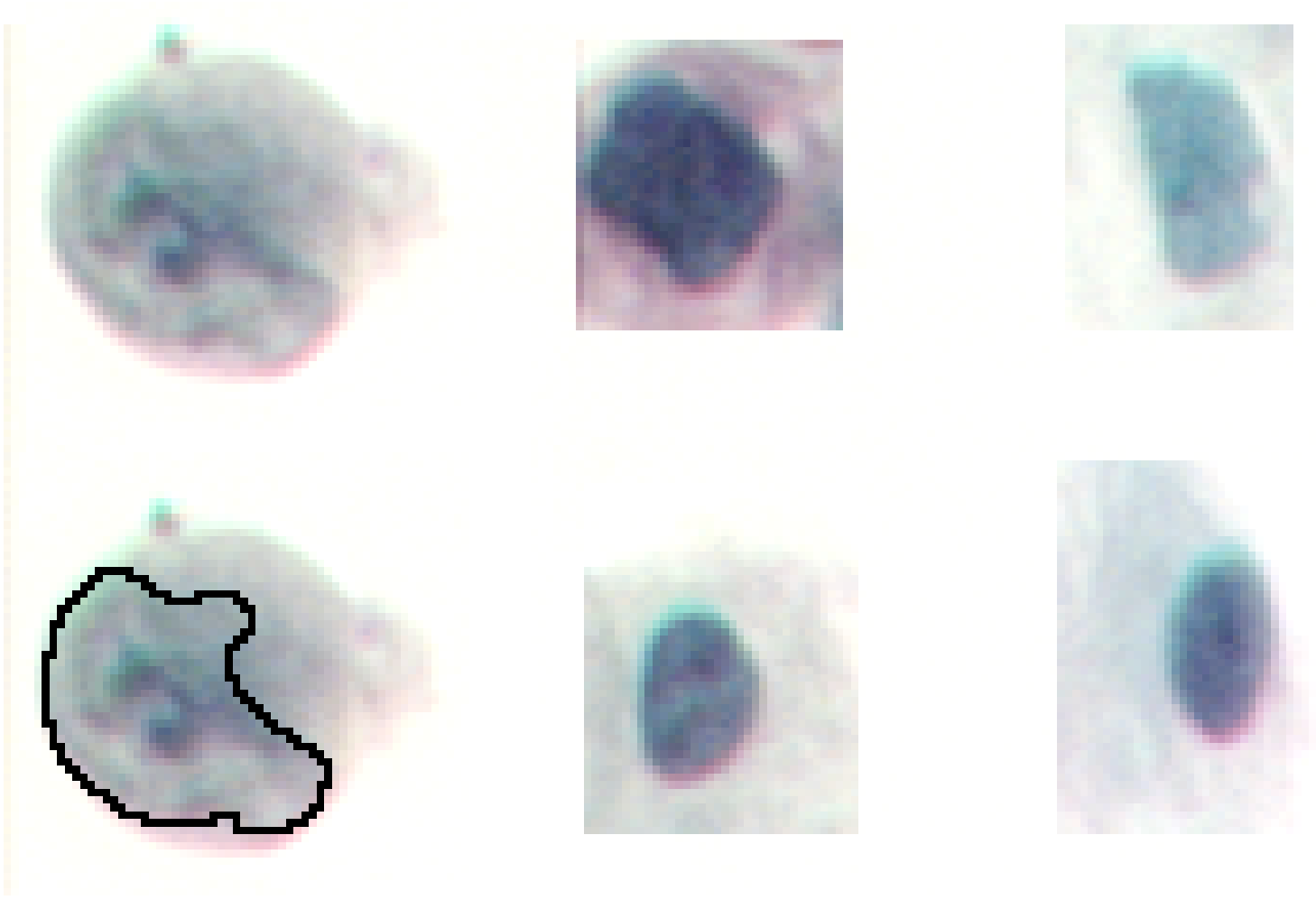
4. Different sizes:



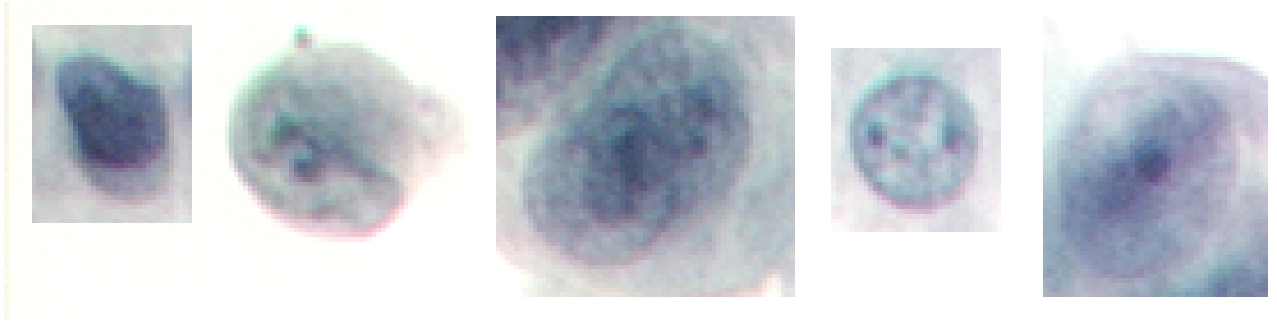
5. Heterogeneous surfaces:



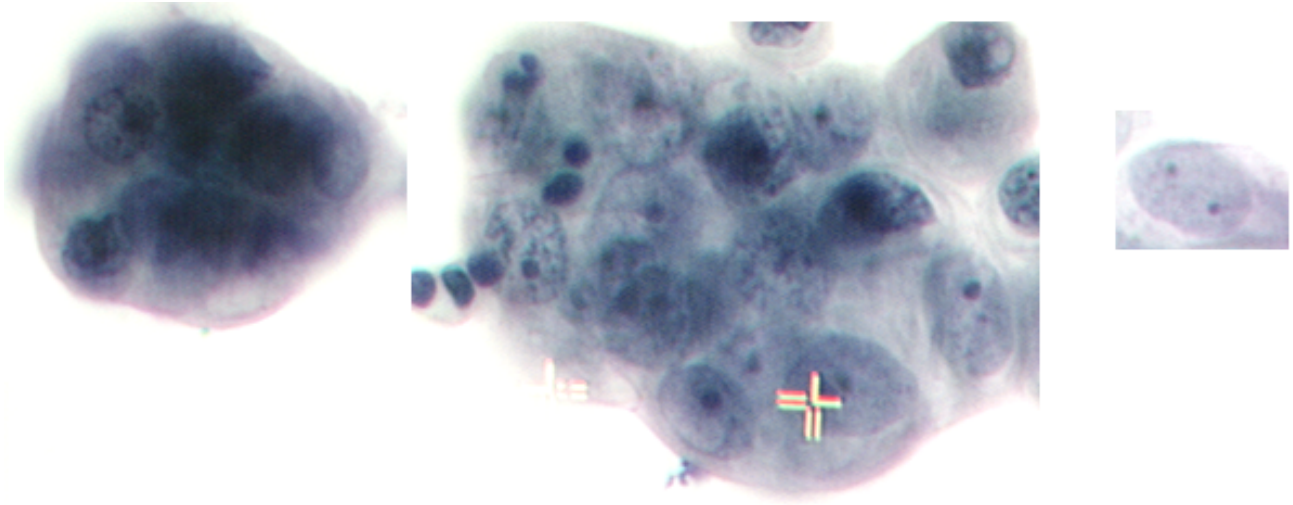
6. Heterogeneous shapes:



7. Multiple cells:

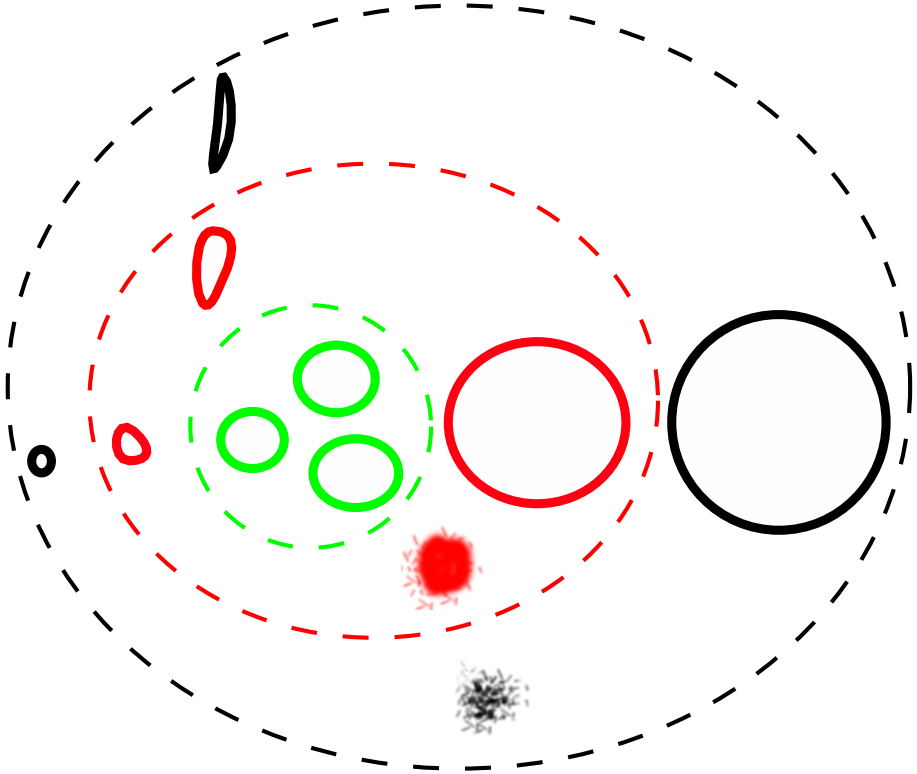


8. Heaps:



Finding something abnormal

We have to accept more than the normal (green) cells, but not to accept other objects (black).



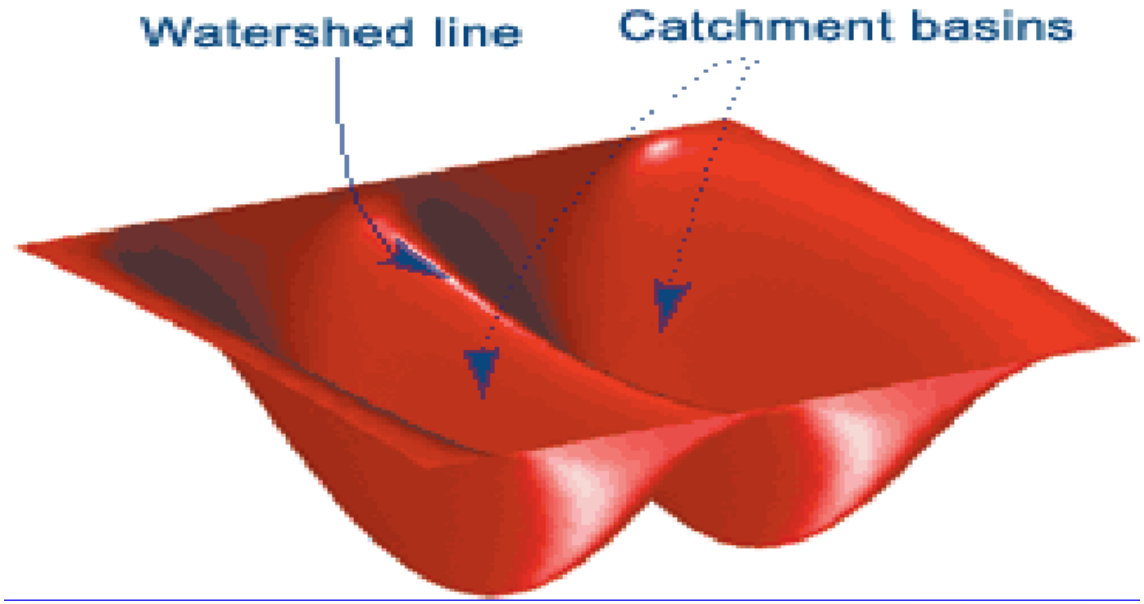
Segmentation

Extraction of

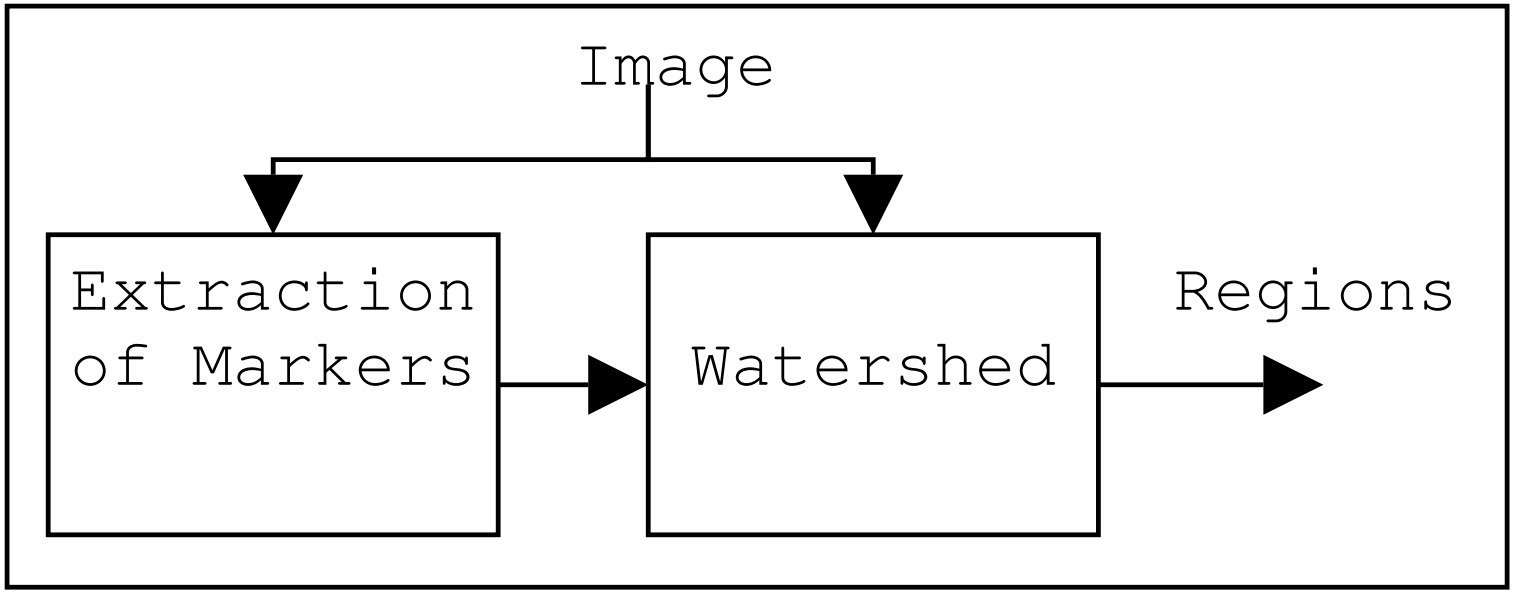
1. The background
2. The heaps
3. The position of the nuclei
4. The boundary of the nuclei

[1/4] Extraction of the background

Using Watershed[Lezoray 98]



Using Watershed

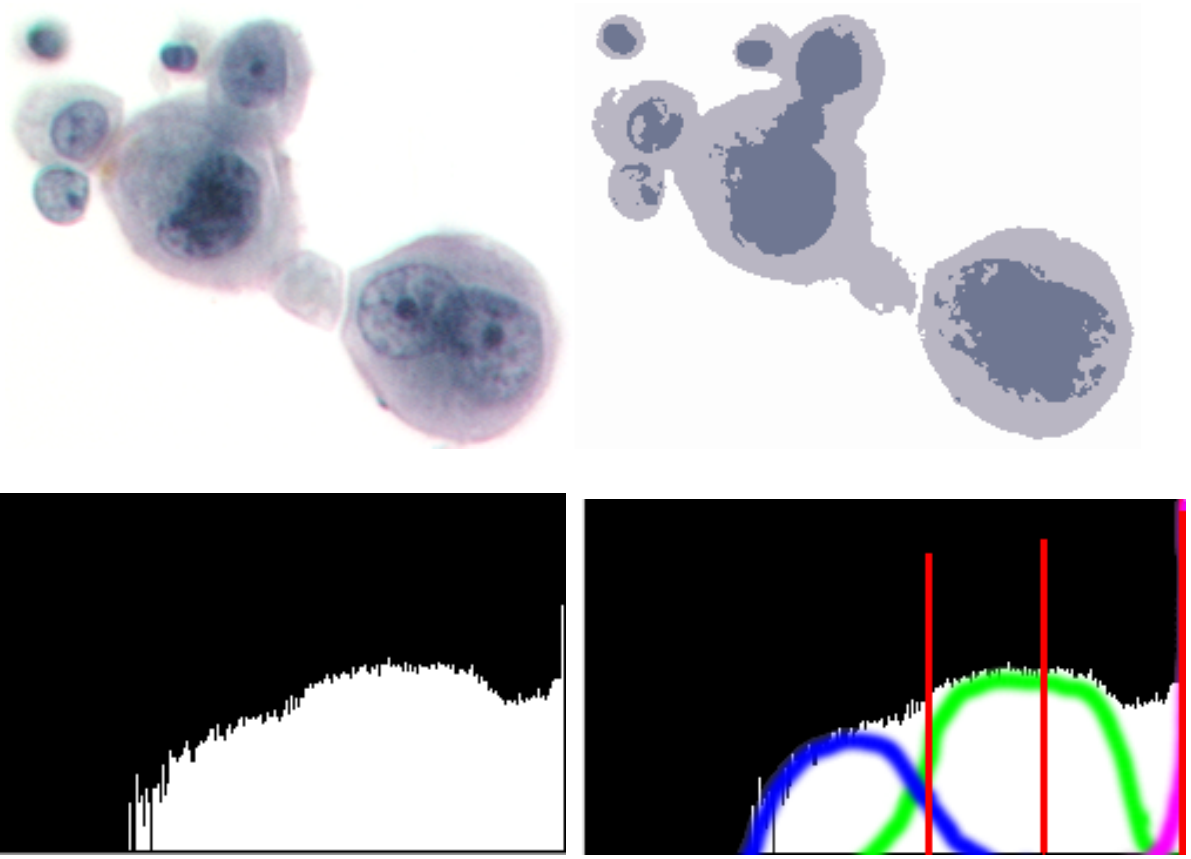


Do not work well:

- No color
- Fuzzy boundaries between cytoplasm and background

[1/4] Using thresholds

K-Mean

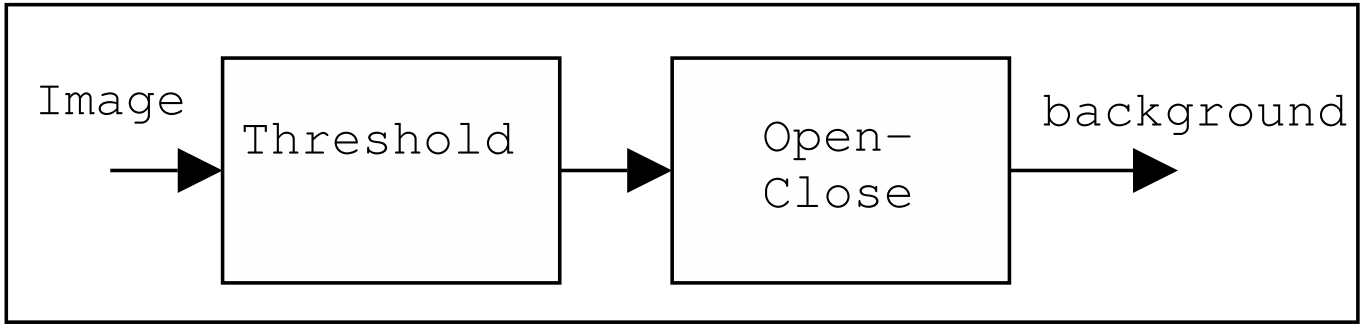


Threshold using the histogram

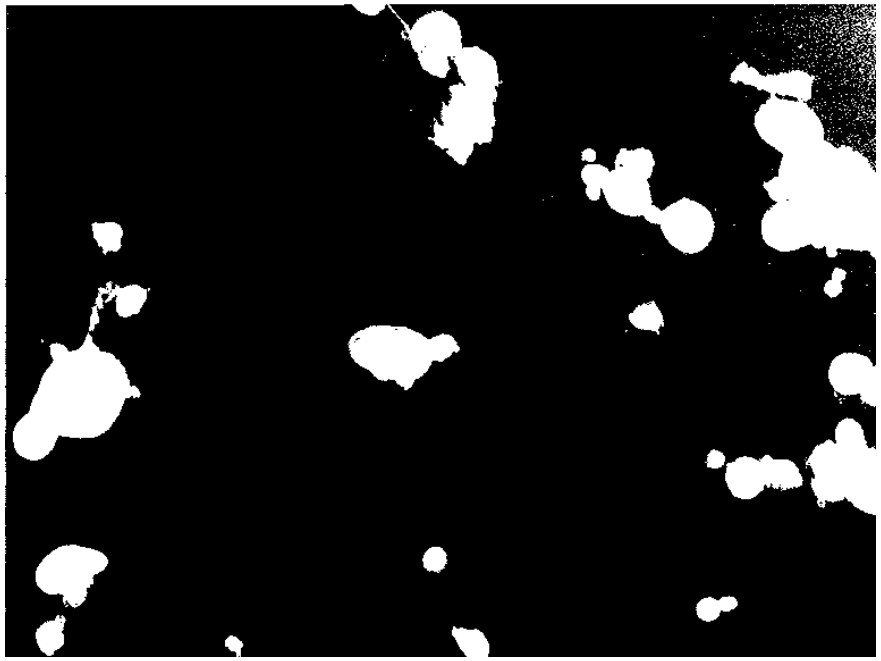
Problems:

- Dust on the light
  $\Rightarrow$ Dark points in the background $\Rightarrow$ Opening

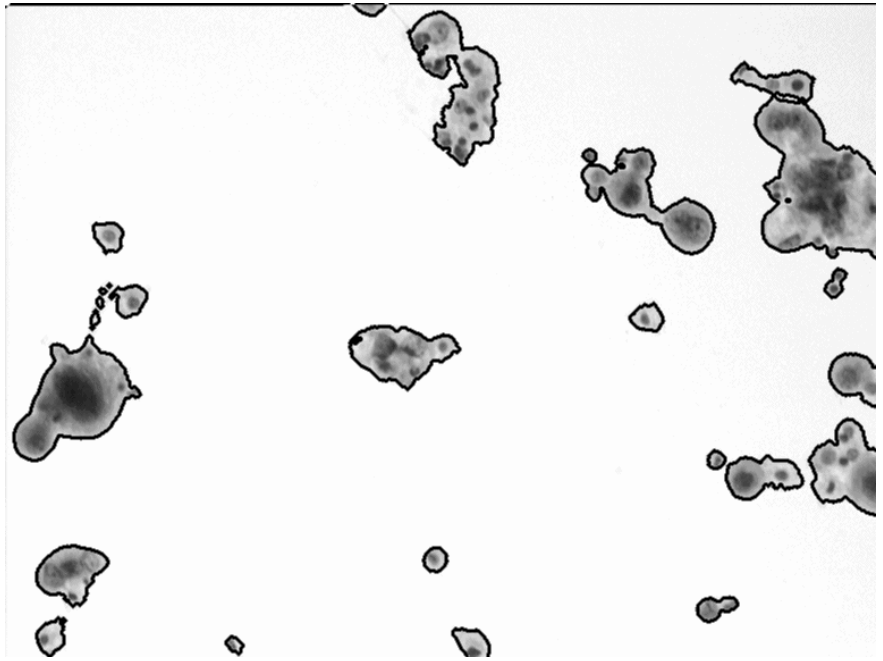
- Impurities and heterogeneous cells
  $\Rightarrow$ White points in the cells $\Rightarrow$ Closing



[1/4] Threshold and opening



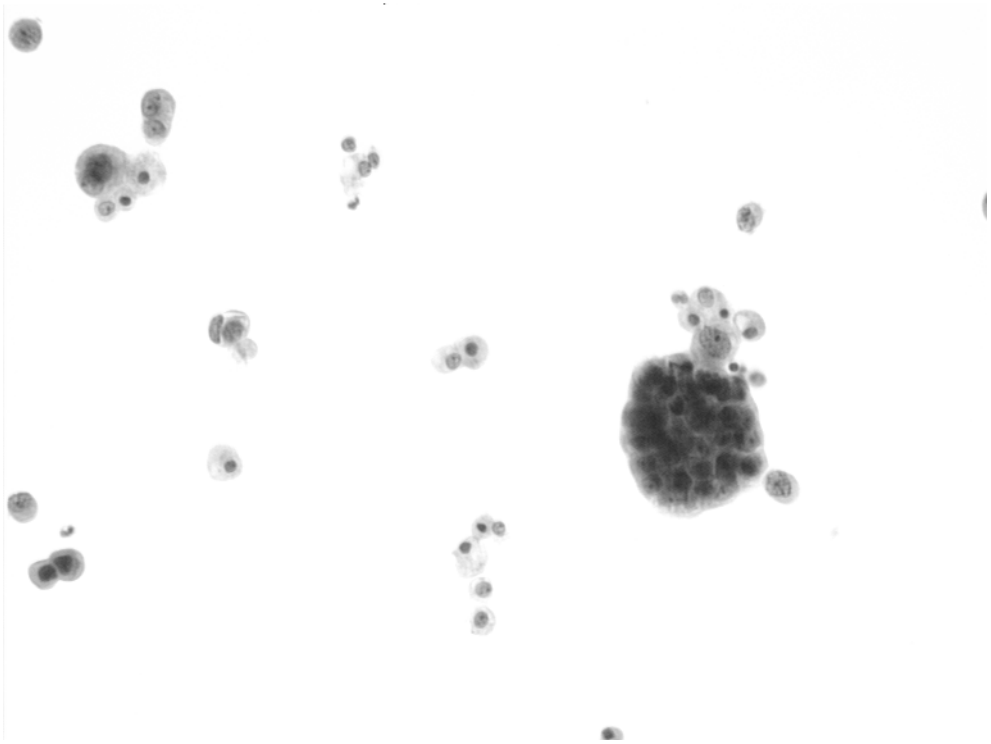
Threshold



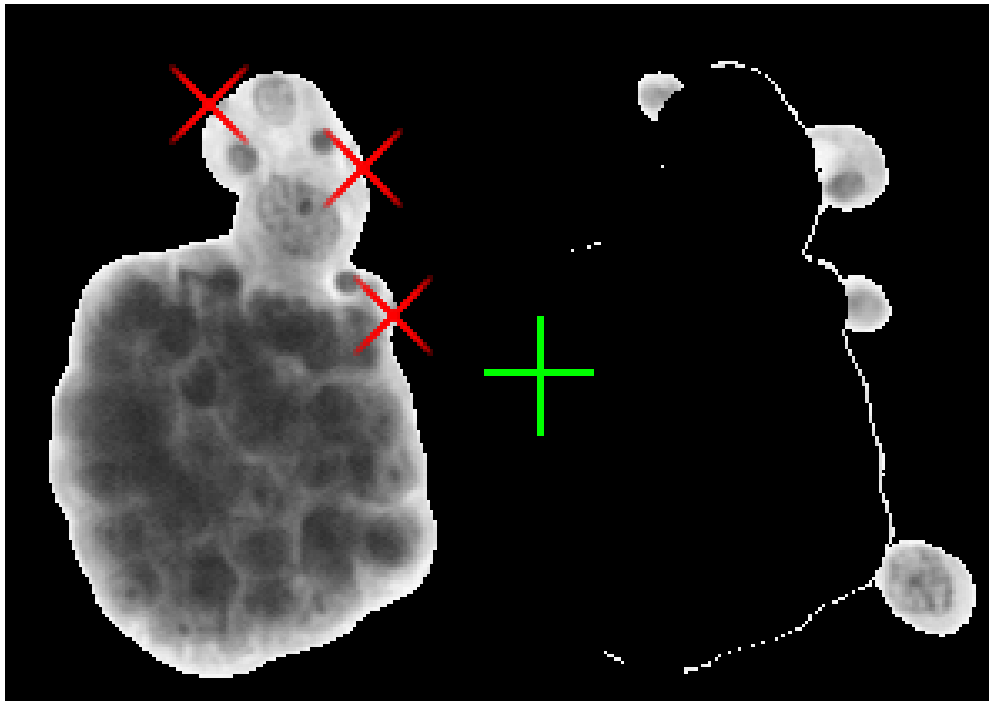
After the opening

[2/4] Extraction of the heaps

Separation of the **isolated** cells and the **heaps**



Heap 's boundary



$$Heap = Heap + isolated\ cells\ stick\ on\ the\ heap$$

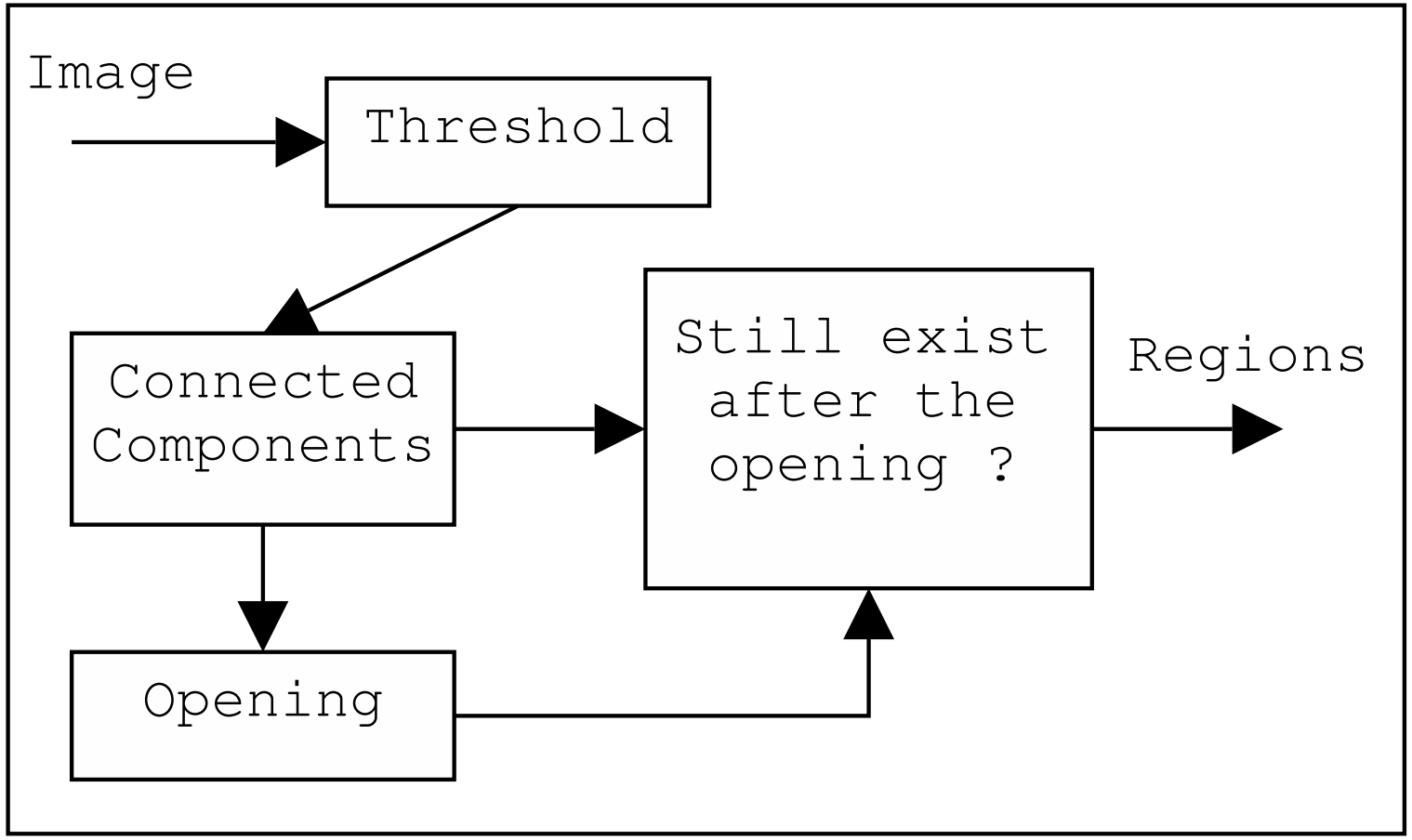
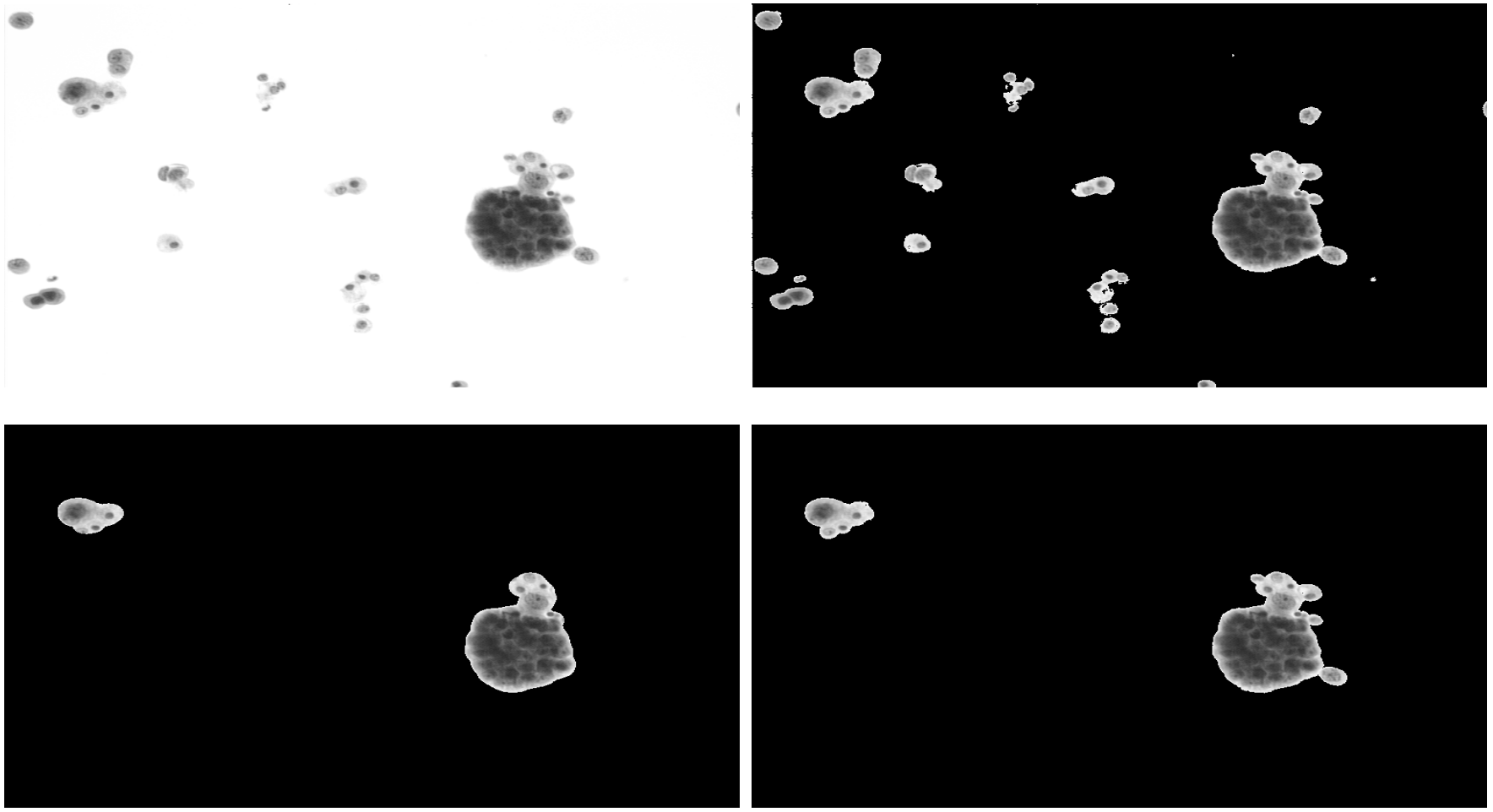
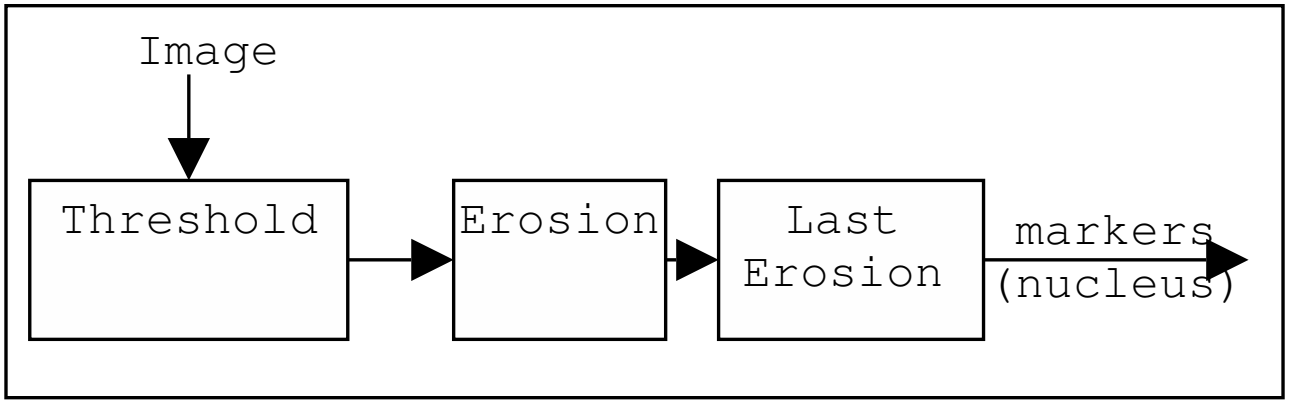


Image
Threshold
Connected Components
Still exist after the opening ?
Regions
Opening

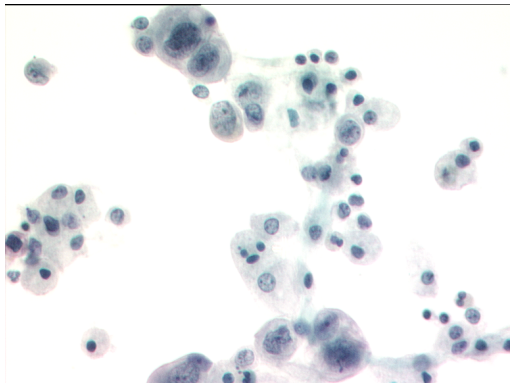
[2/4] Original/Threshold/Opening/Result

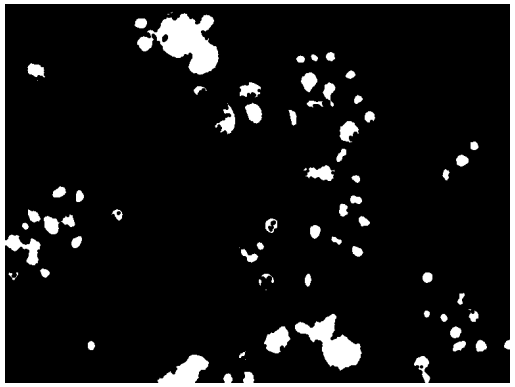


[3/4] Extraction of the nuclei ’s position

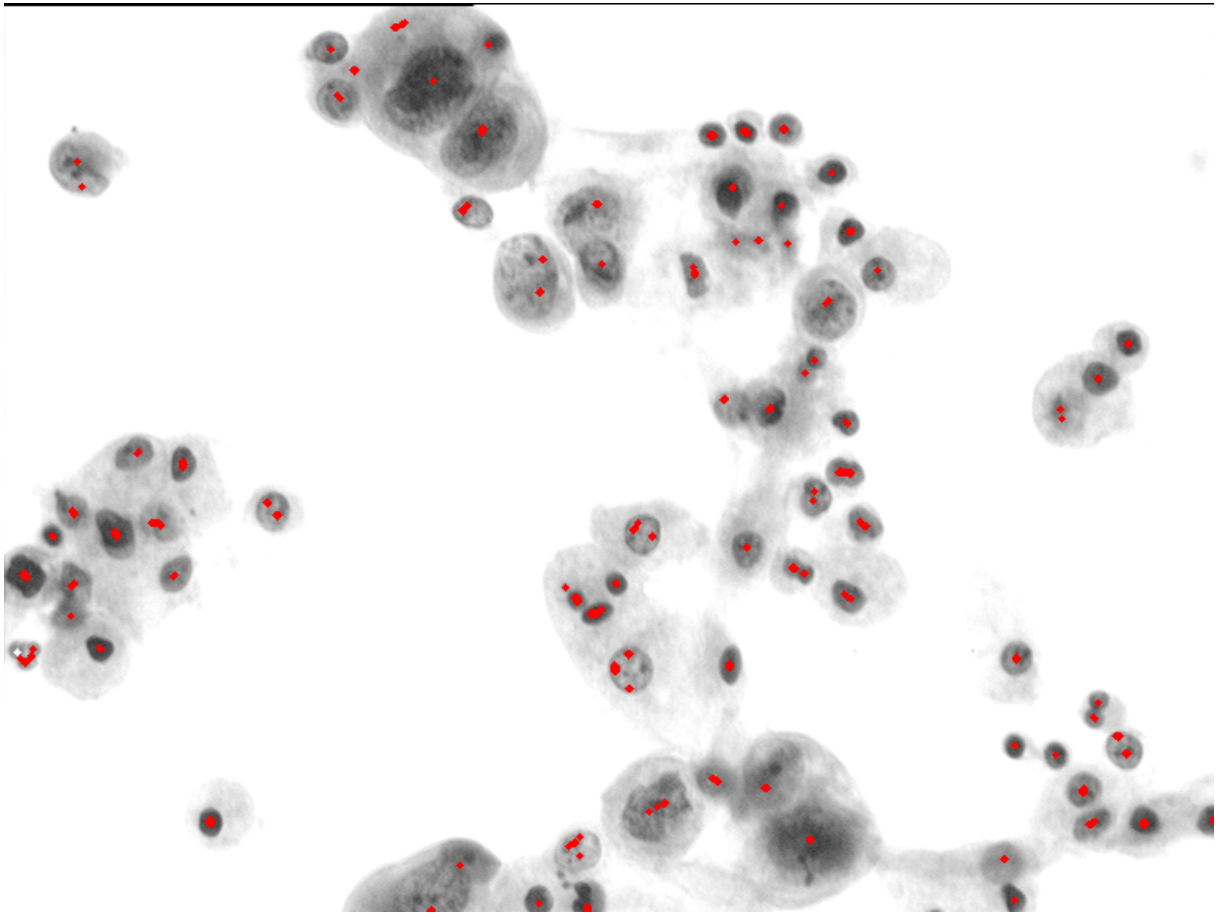


[3/4] Original/Threshold + Erosion/Last Erosion





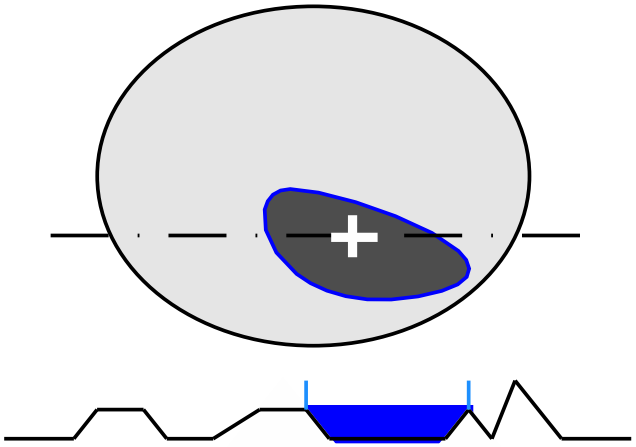
[3/4] Result



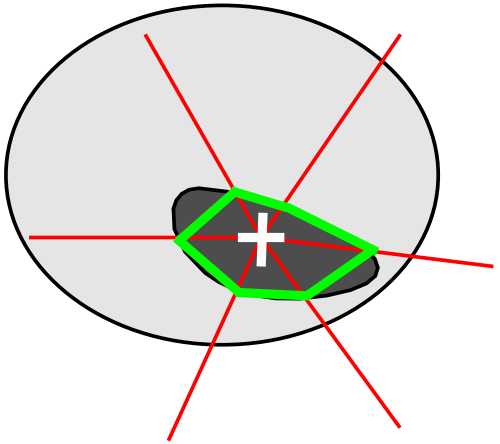
[4/4] Extraction of the nuclei ’s boundaries

Now, the position of the nucleus is known *(white cross)*

The goal is to find the boundary of the nucleus *(blue and green line)*



Watershed

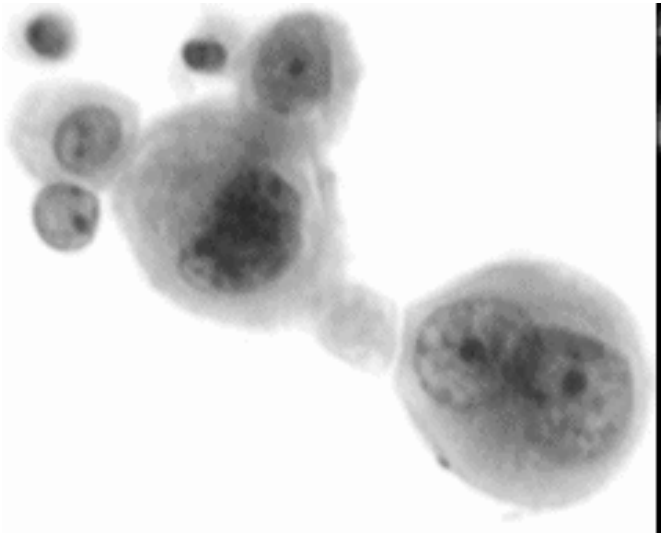


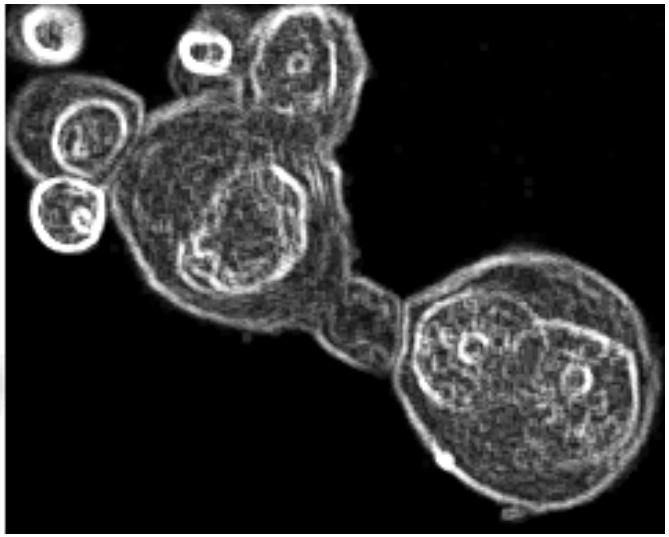
Radius

Using Watershed[Lezoray98]

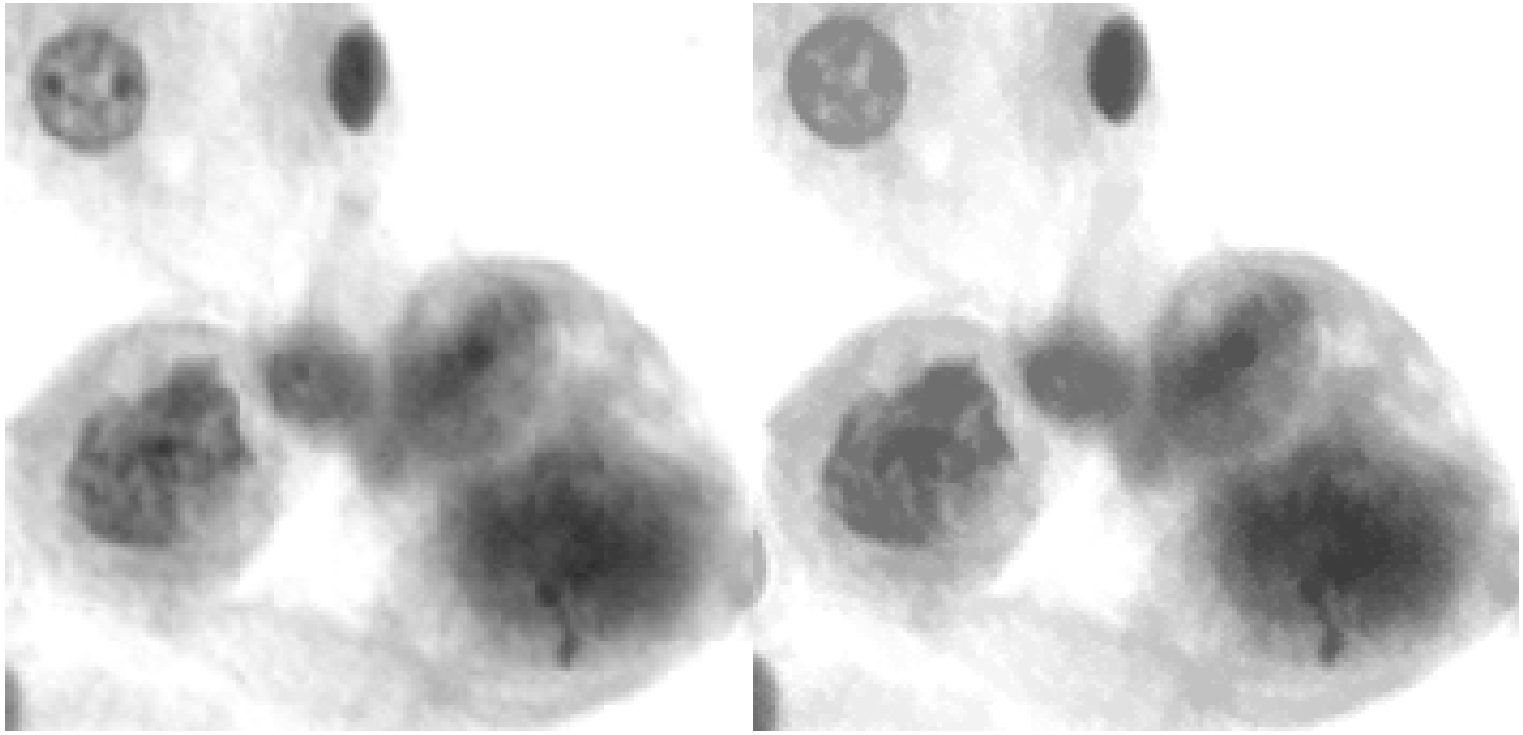
- The markers *(where the water comes from)* are the position of of the nuclei
- The **gradient** of the image is needed
- The 'water' should not be stopped by the **impurities** of the cell

Beucher ’s Gradient



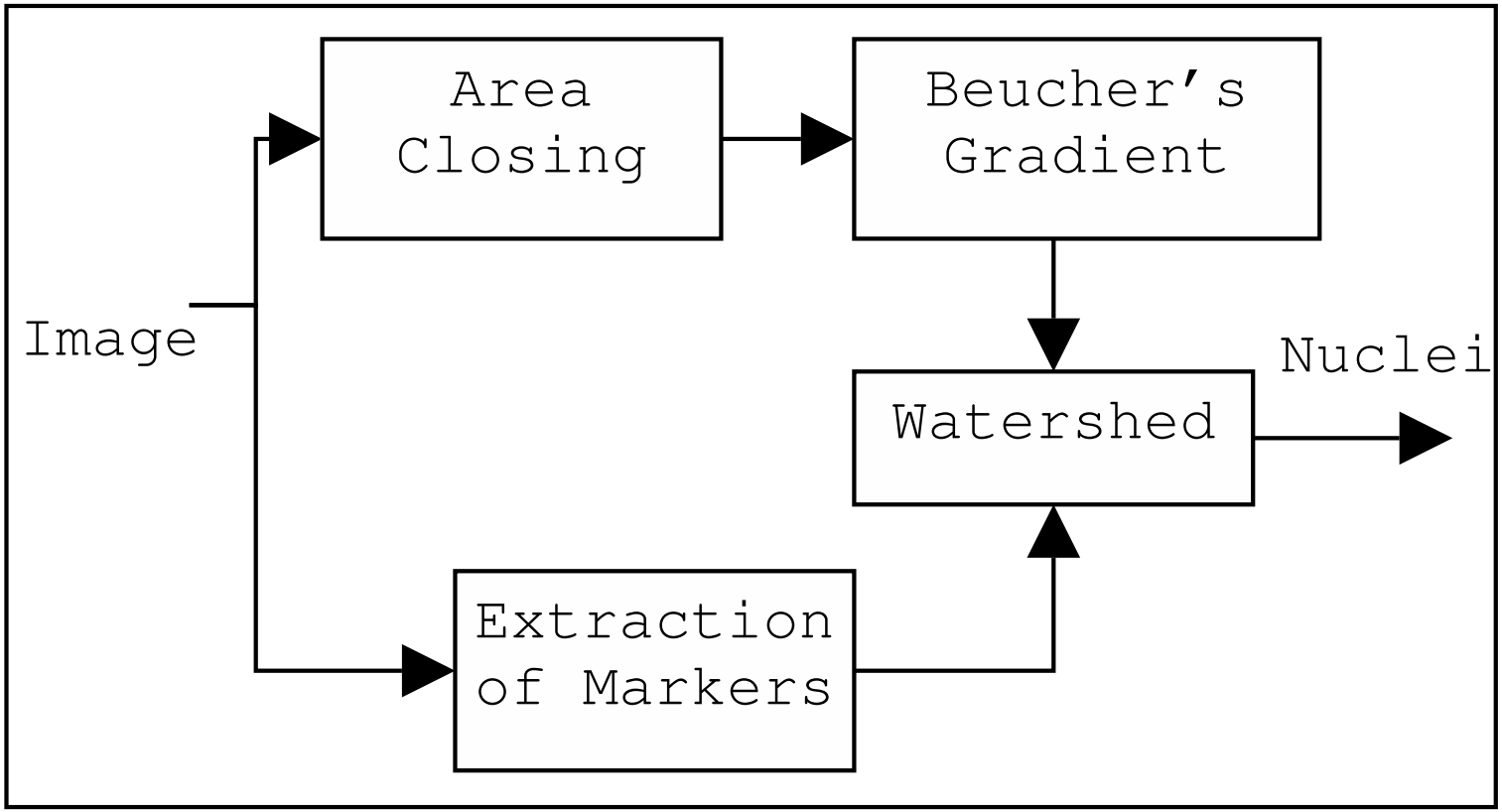


Area Closing

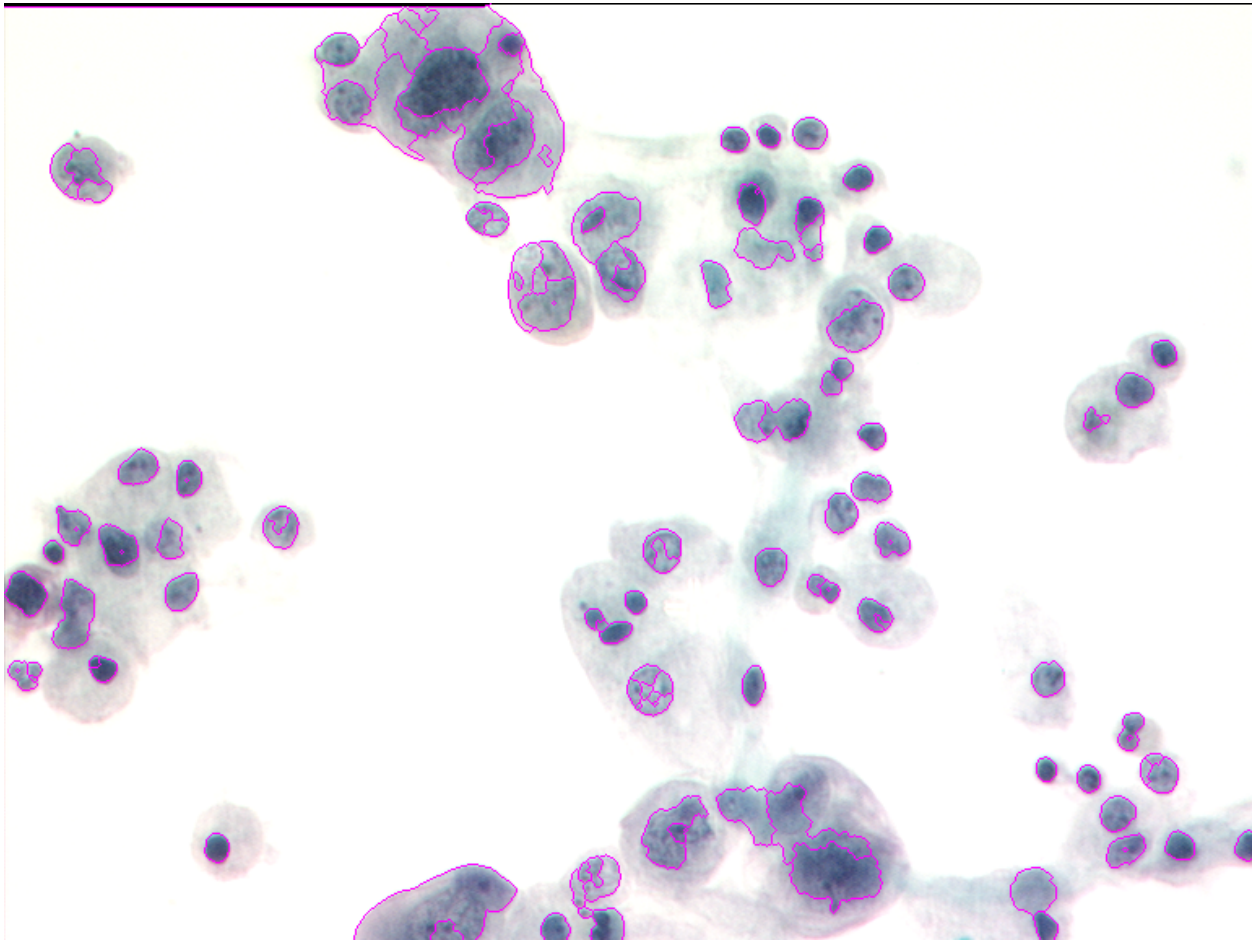


Clear small dark areas

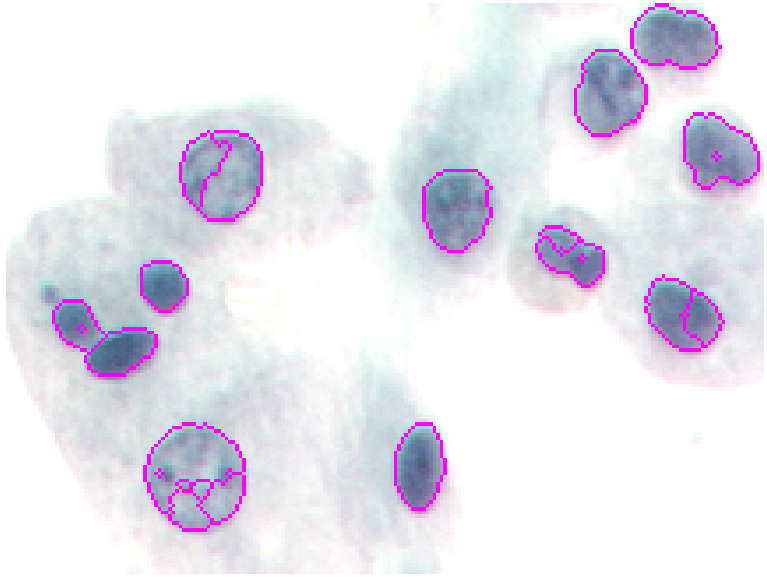
Extraction of nuclei ’s boundary using watershed



Result



Over-segmentation



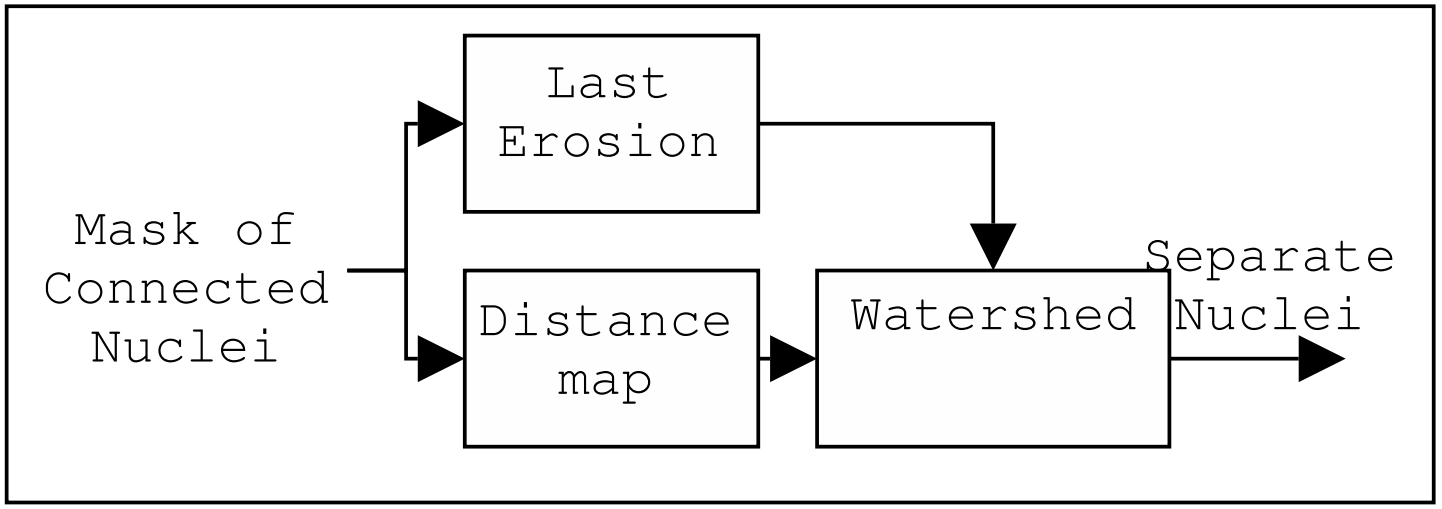
Over-segmentation within the nuclei!

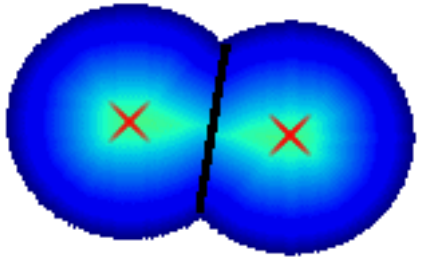
$\Rightarrow$ **Union of regions that share a boundary**

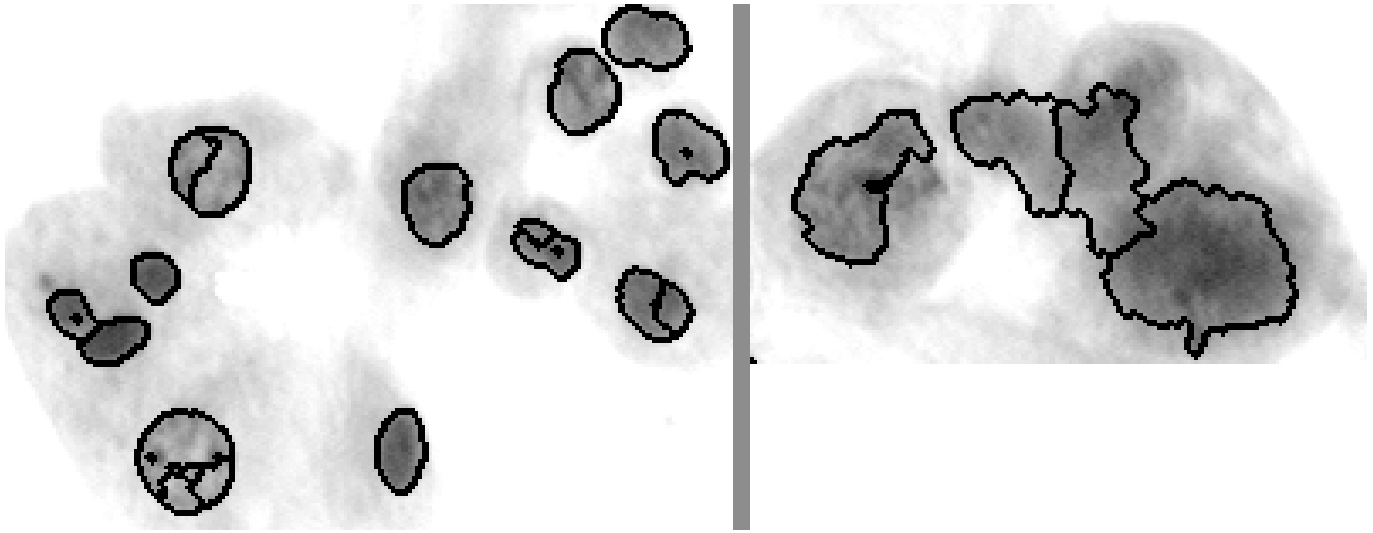
$\Rightarrow$ Many nuclei sticked together share the same region!

$\Rightarrow$ **Separation of these regions with distance map**

Separation of interconnected nuclei







Do not work well on heterogenous cells. Do not work well on fuzzy cells

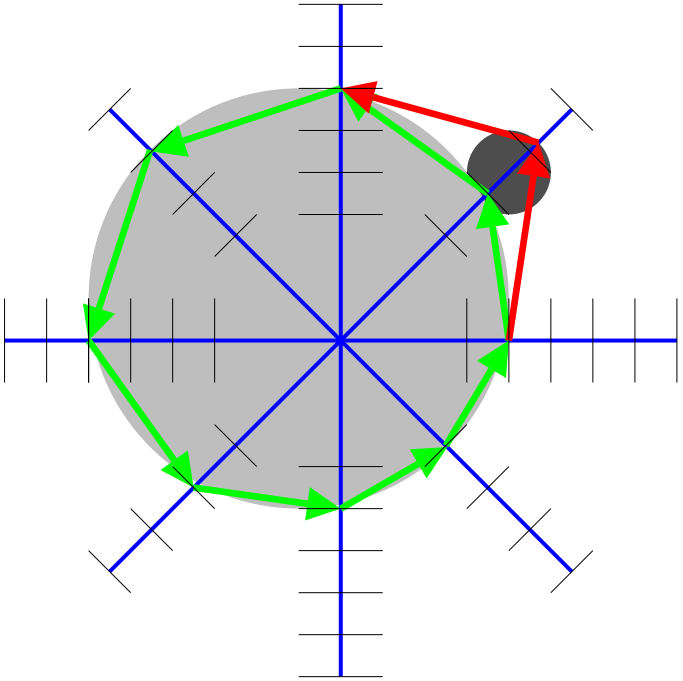
Shape ’s information

Research of elliptic objects[Wu, Barba, Gil 98]

Snakes [lee, Street 99]

Research of round objects[Bamford/Lovell]

Radius



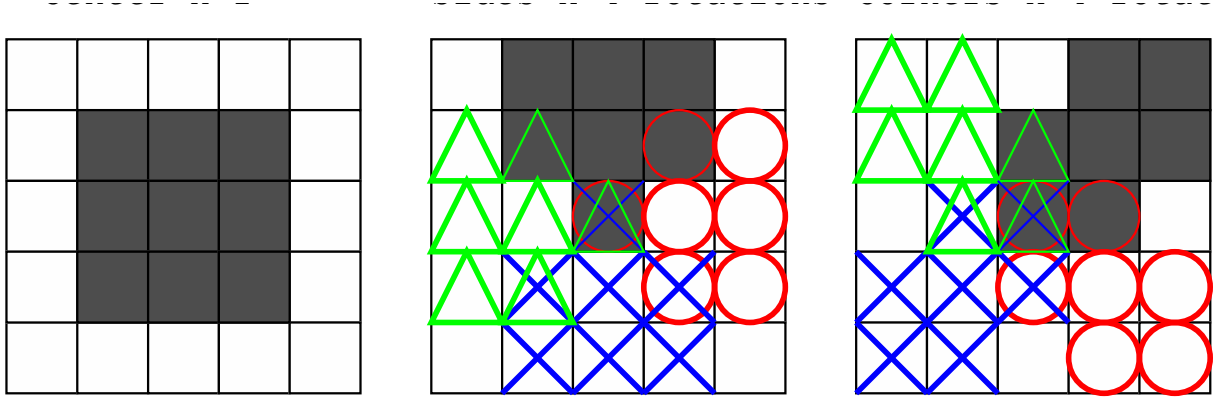
- α coefficient

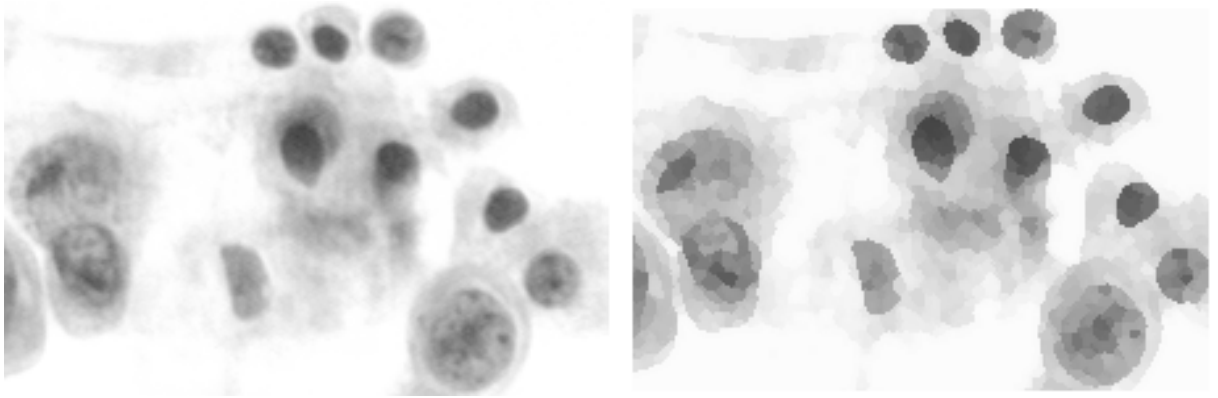
$$Cost = \alpha * circular\_shape + (1 - \alpha) * bondary\_matching$$

- $\Delta angle$ coefficient
- $range$

Problem: Our images are much more **heterogeneous**

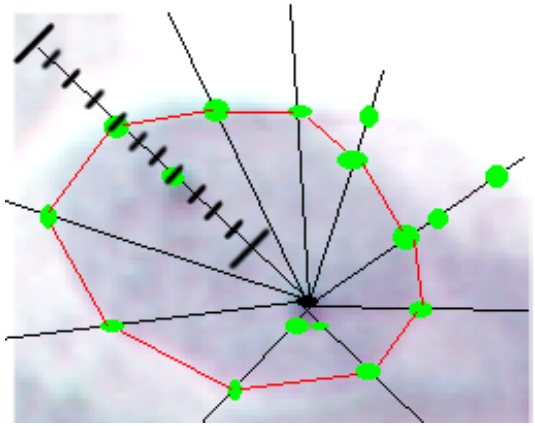
Nagao filter





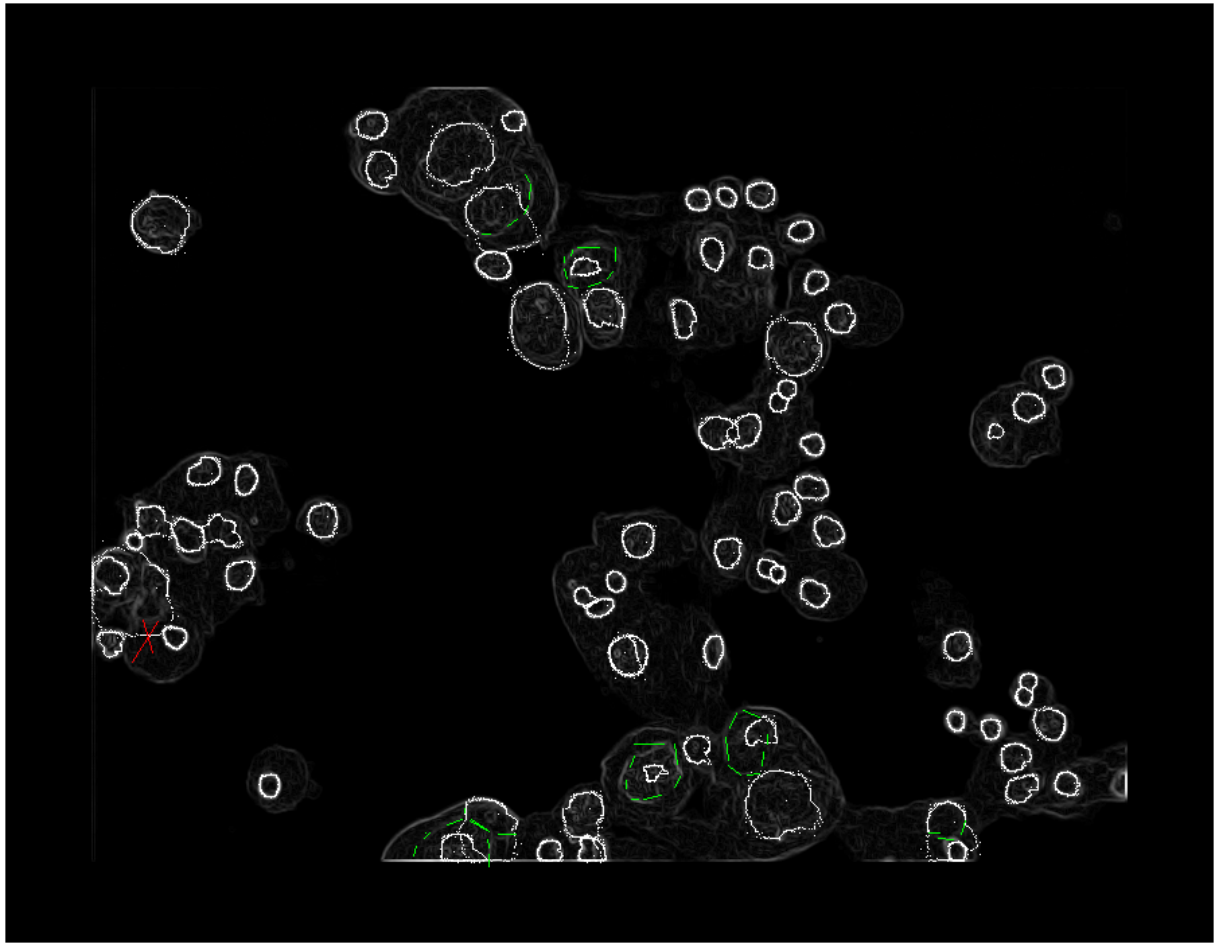
Without filtered After filtered

Problems of the radius technique



- The **centers** of the nuclei are not exact
- Cancerous cells are (sometime) **oval** and not circular
- Huge difference of **size**
  $\Rightarrow$ **More radius, longer radius** $\Rightarrow$ **Too slow to compute every path and harder to control**

Results



Improving the segmentation 1/3

- **Two** segmentations [Bamford/Lovell]
- Using a **trust** degree (cost of the path)
- Detecting errors after the segmentation

Improving the segmentation 2/3

Using **more information**; example with the radius technique:

- Wider circles: Using the inclusion information
- Local information: Same gray level for the same nucleus / cytoplasm

Improving the segmentation 3/3

- **General** information (if x cells share 1 heap, at least 1 cell is bigger than $size(heap)/x$)
- Focus on **interesting parts** of the image (dark)

Conclusion

- Segmentation of **abnormal** cells $\neq$ segmentation of normal cells
- Classification of the result can remove the wrong cells
- Using more informations
- **The goal is not to replace the pathologist**