M.Sc. Thesis Defense

Perceptual Watersheds for Cell Segmentation in Fluorescence Microscopy Images

Presenter: Salim Arslan

Supervisor: Çiğdem Gündüz Demir

Department of Computer Engineering
Bilkent University
August 6th, 2012

Outline

- Introduction
- Background
  - Problem Domain
  - Related Work
- Motivation
- Method
- Experiments
- Conclusion
- Future Work

Introduction

- High content screening helps scientists
  - Analyze complex biological systems
  - Collect quantitative data
- Advances in automated fluorescence microscopy imaging systems lead to
  - Rapid high-throughput screening
  - Better reproducibility

High Content Screening

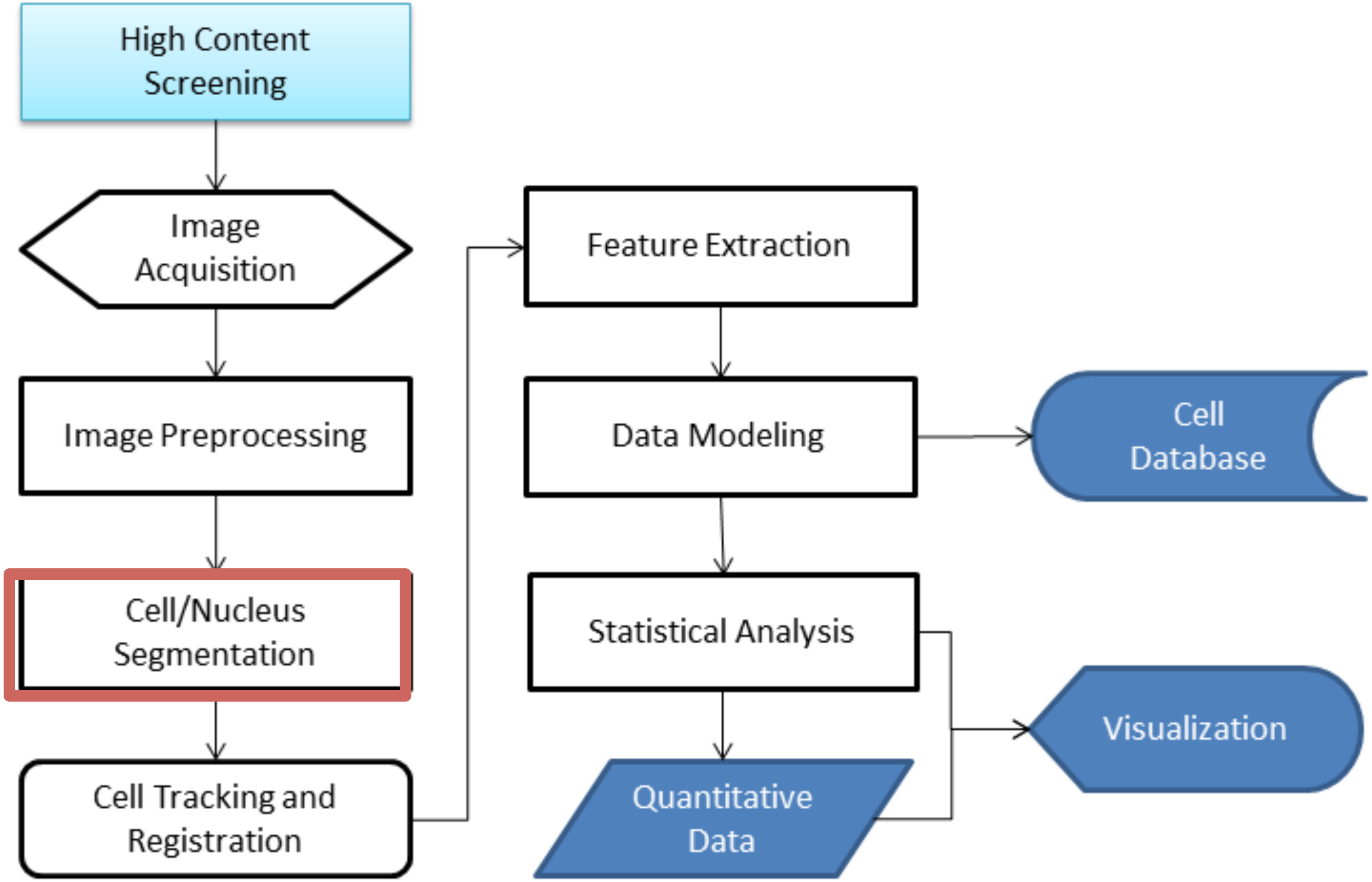

High Content Screening

Image Acquisition

Image Preprocessing

Cell/Nucleus Segmentation

Cell Tracking and Registration

Feature Extraction

Data Modeling

Statistical Analysis

Quantitative Data

Cell Database

Visualization

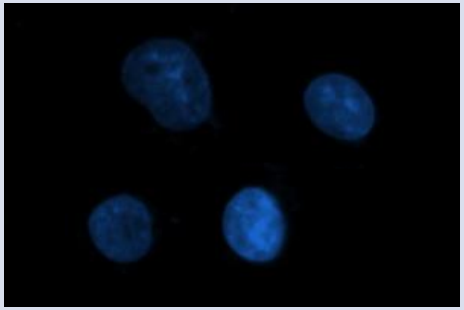
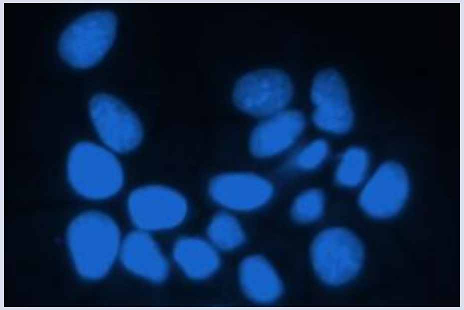
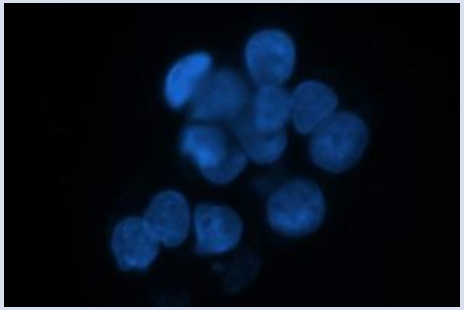
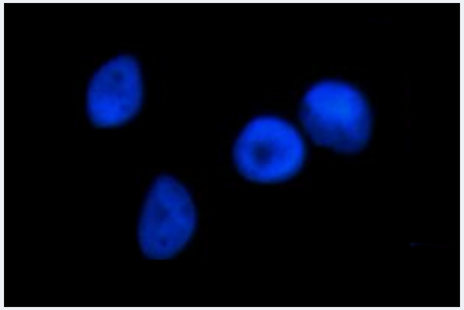
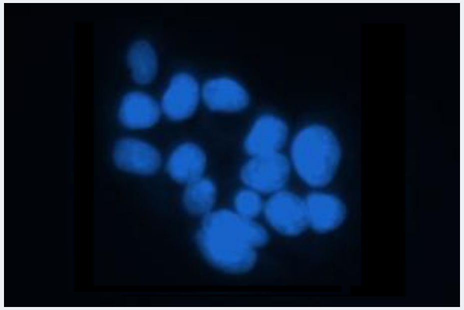
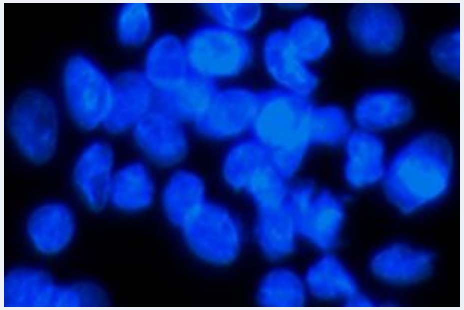
Cell Segmentation

- The core step is cell/nucleus segmentation
  - Results directly affect the performance of the system

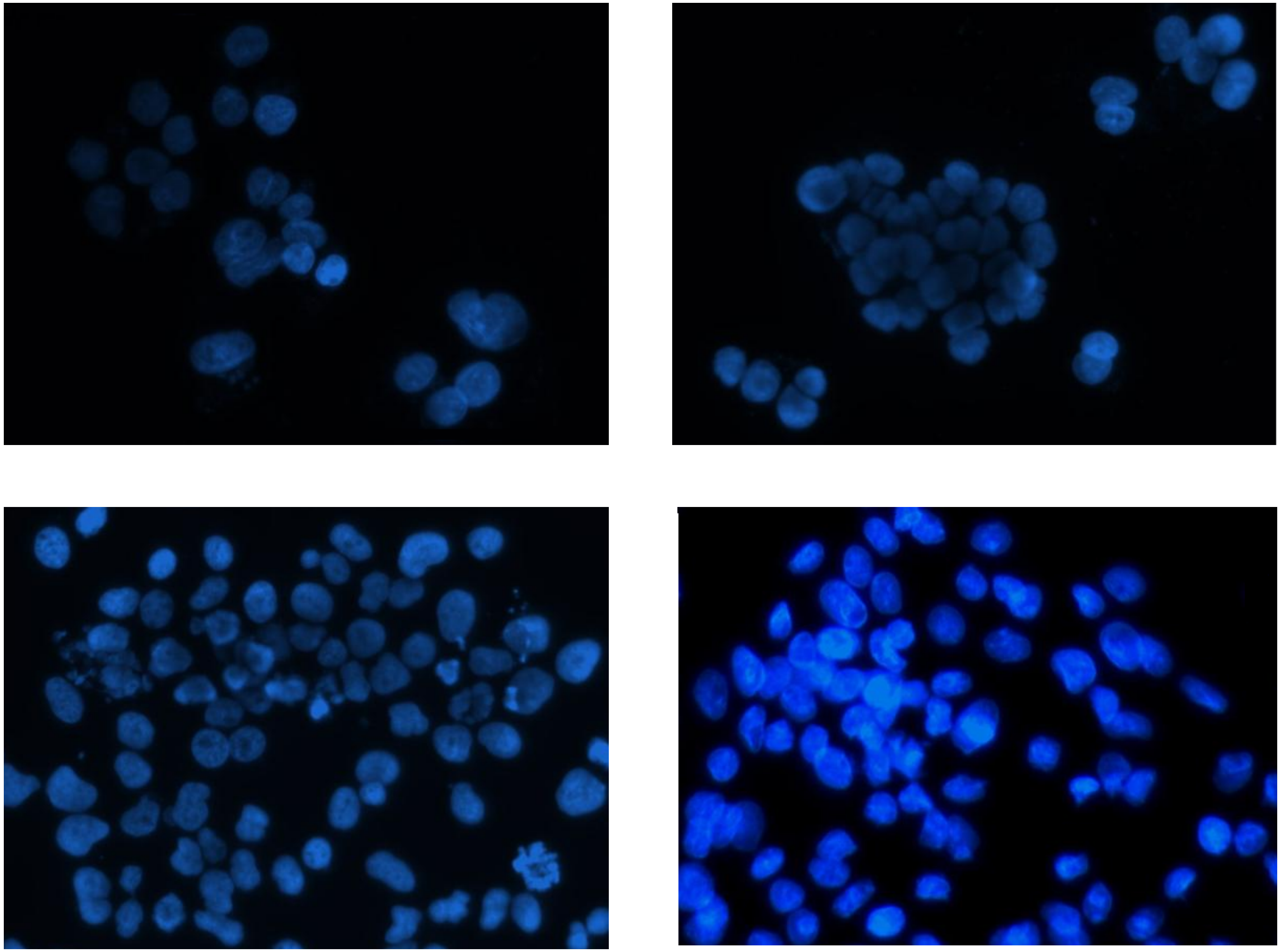
- It is of great importance to develop accurate segmentation algorithms both for
  - Monolayer isolated cells
  - Confluent cells growing in aggregates as over-layers

- These algorithms should be robust to
  - Uneven illumination
  - Noise

Isolated vs. Confluent Cells

Monolayer / Isolated	Touching / Adjacent	Confluent (Clustered)
		
		

Uneven Illumination & Noise



Related Work

- Thresholding
    - Global [1] and adaptive (local) [2]
- Active contours (deformable models)
    - Edge based active contours (snakes) [3]
    - Region based active contours [4]
- Watershed-based methods
    - Marker-controlled watersheds [5][6]

[1] **Chen et al.**, 2006, *IEEE T Biomed Eng*, [2] **Kass et al.**, 1988, *Int J Comput Vision*,
[3] **Kass et al.**, 1988, *Int J Comput Vision*, [4] **Zimmer et al.**, 2006, *IEEE T Med Imag*,
[5] **Cheng et al.**, 2009, *IEEE T Biomed Eng*, [6] **Yang et al.**, 2006, *IEEE T Circuits-I*

Marker–Controlled Watershed

- Water rises from only previously defined seed points, a.k.a. markers

- Marker detection and marking function
  - Define markers/marking function on the gradients/intensities or distance transform map calculated after binarization
    - H-minima transform to suppress undesired minima
  - Common to refine results afterwards
    - Validate, merge, split the segmented regions

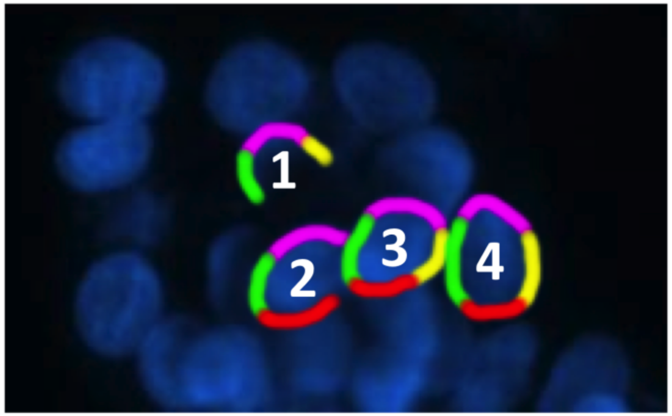
- Gives accurate results when
  - One-to-one match between cells and markers

Motivation

- Previous studies lead to promising results, but there still remain challenges to overcome
- Segmentation of both
  - Isolated cells
  - Confluent cells
- Noise and uneven lighting conditions in images

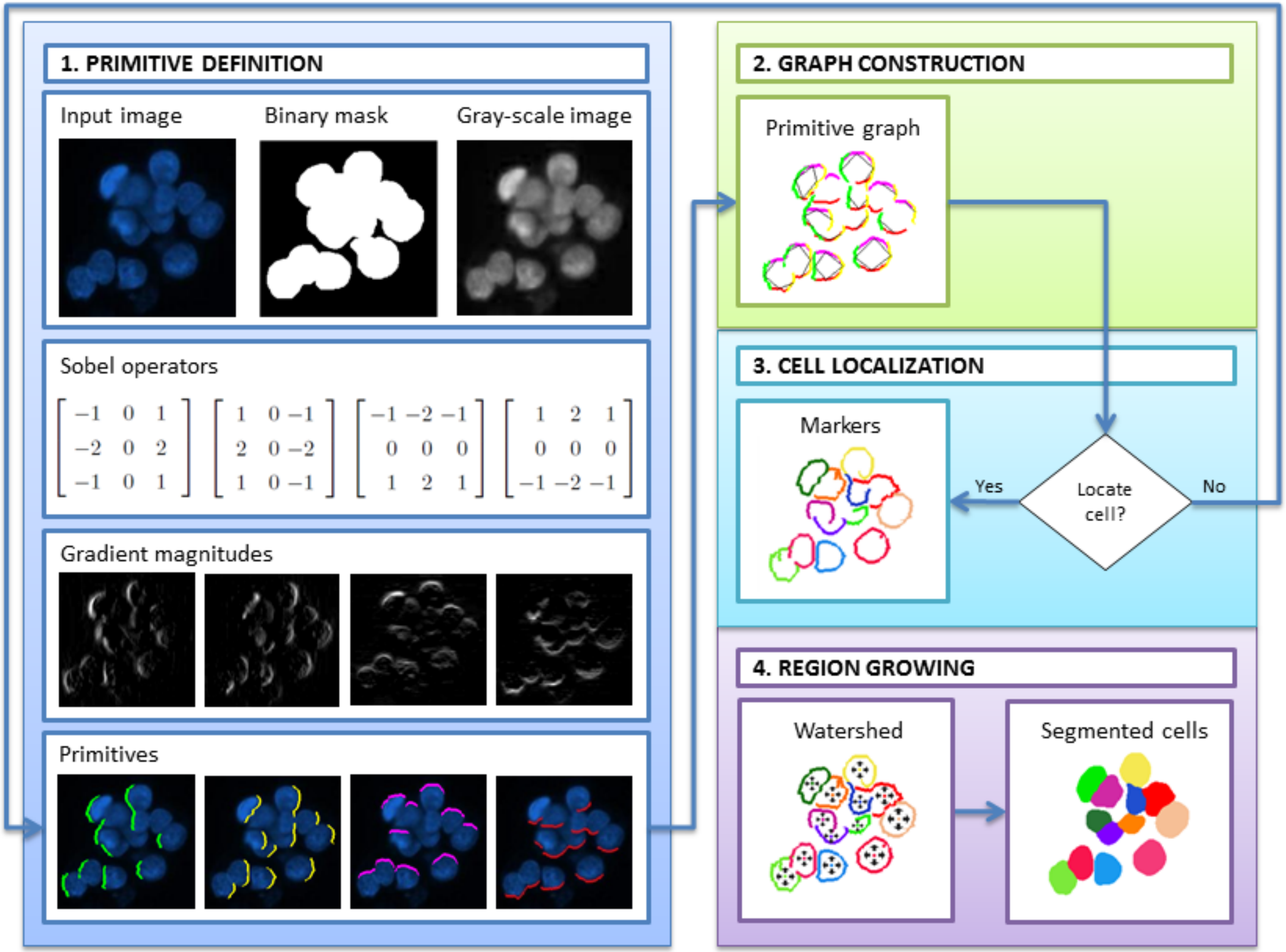
Motivation

- A new algorithm that incorporates human perception into cell segmentation
- The algorithm relies on modeling a very trivial fact a human uses
- Each cell should have a
  - Left boundary (yellow)
  - Right boundary (green)
  - Top boundary (magenta)
  - Bottom boundary (red)



- These boundaries should be in the correct position with respect to each other

Method Overview



1. PRIMITIVE DEFINITION

Input image

Binary mask

Gray-scale image

Sobel operators

$$\begin{bmatrix}-1&0&1\\-2&0&2\\-1&0&1\end{bmatrix}\quad\begin{bmatrix}1&0&-1\\2&0&-2\\1&0&-1\end{bmatrix}\quad\begin{bmatrix}-1&-2&-1\\0&0&0\\1&2&1\end{bmatrix}\quad\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}$$

Gradient magnitudes

Primitives

2. GRAPH CONSTRUCTION

Primitive graph

3. CELL LOCALIZATION

Markers

Locate cell?

Yes

No

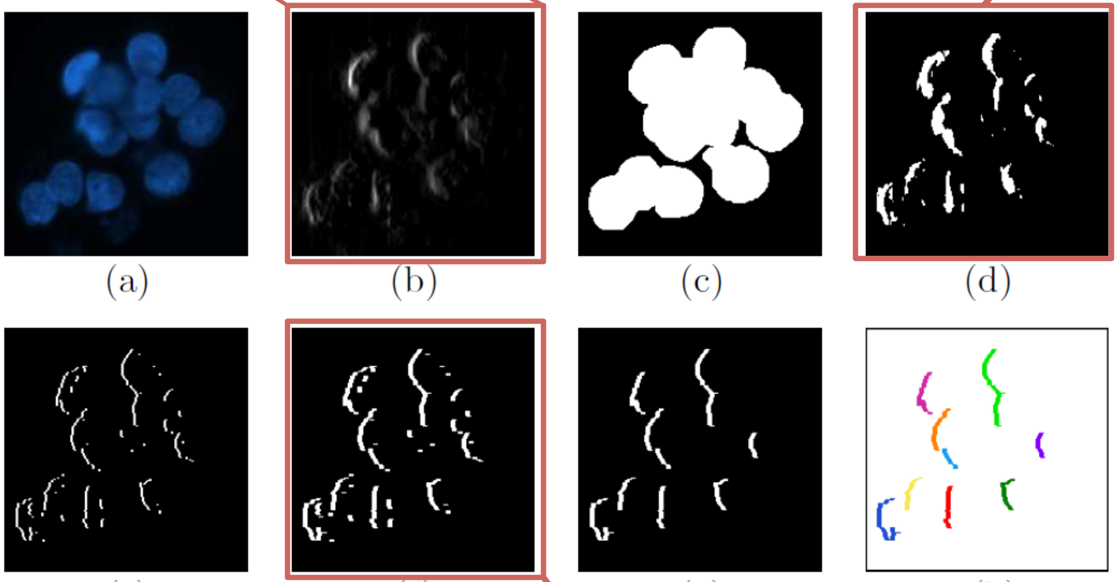
4. REGION GROWING

Watershed

Segmented cells

Primitive Definition

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$



(a) (b) (c) (d)

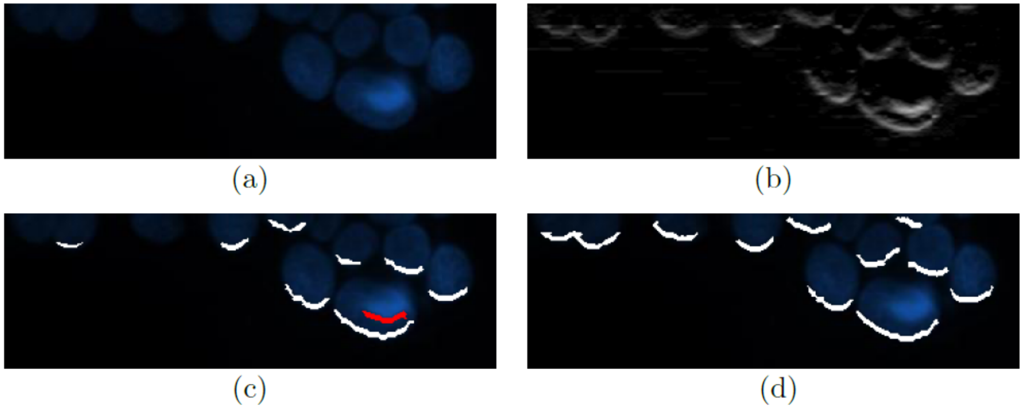
(e) (f) (g) (h)

Take *d*-leftmost pixels only

Eliminate components whose sizes are smaller than $t{\downarrow}size$

Benefit of the Mask

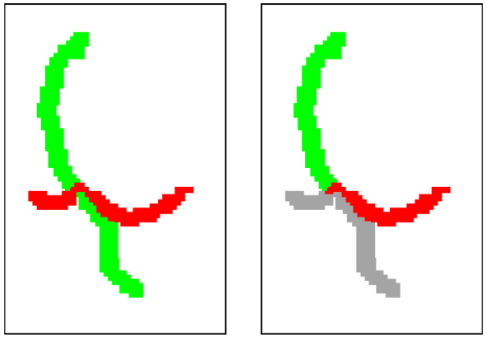
- Used to locally threshold gradients for primitive definition
  - Removes noise from the image
  - Preserves local gradient information
- Used nowhere else throughout the algorithm
- Any binarization method is possible as soon as it captures most of the cellular regions


(a) (b)

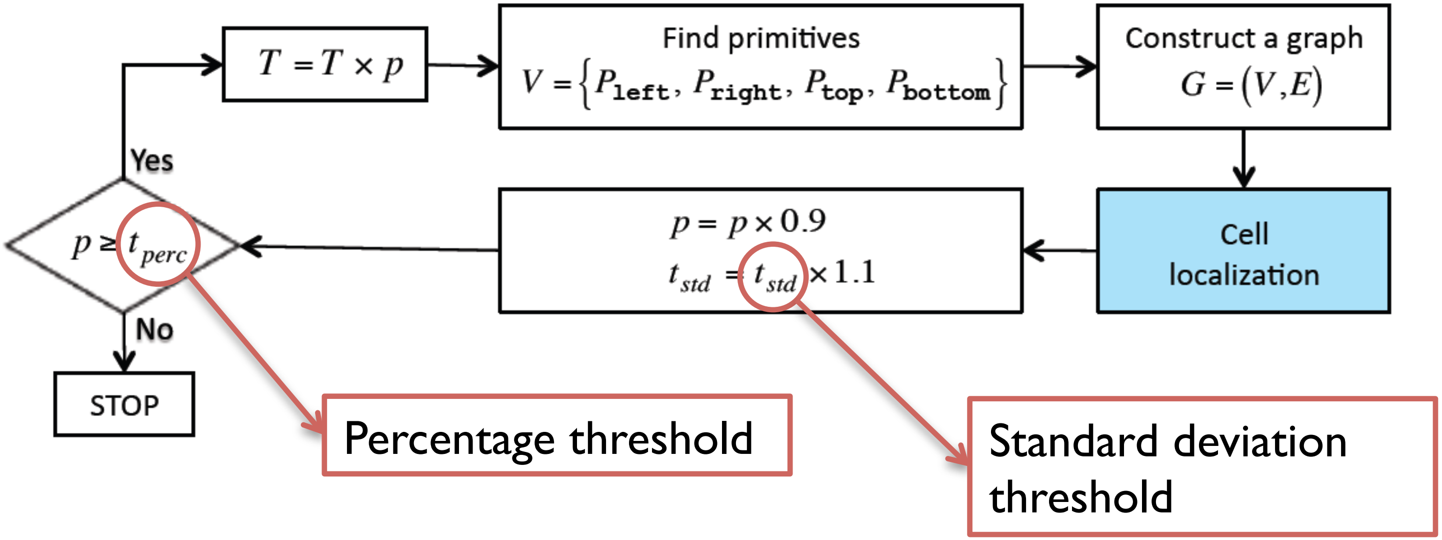
(c) (d)

Graph Construction

- Primitive graph $G=(V, E)$
- $V=\{P_{left}, P_{right}, P_{top}, P_{bottom}\}$
- An edge $e=(u, v)\in E$ is assigned if
    - Primitives have overlapping pixels
    - One primitive is vertical (left or right) and the other is horizontal (bottom or top)
    - Each primitive has a “large enough” segment in the correct side of the other



Iterative Search Algorithm



$T = T \times p$

Find primitives

$V = \{P_{\mathtt{left}}, P_{\mathtt{right}}, P_{\mathtt{top}}, P_{\mathtt{bottom}}\}$

Construct a graph

$G = (V, E)$

Cell localization

$p = p \times 0.9$

$t_{std} = t_{std} \times 1.1$

$p \geq t_{perc}$

Yes

No

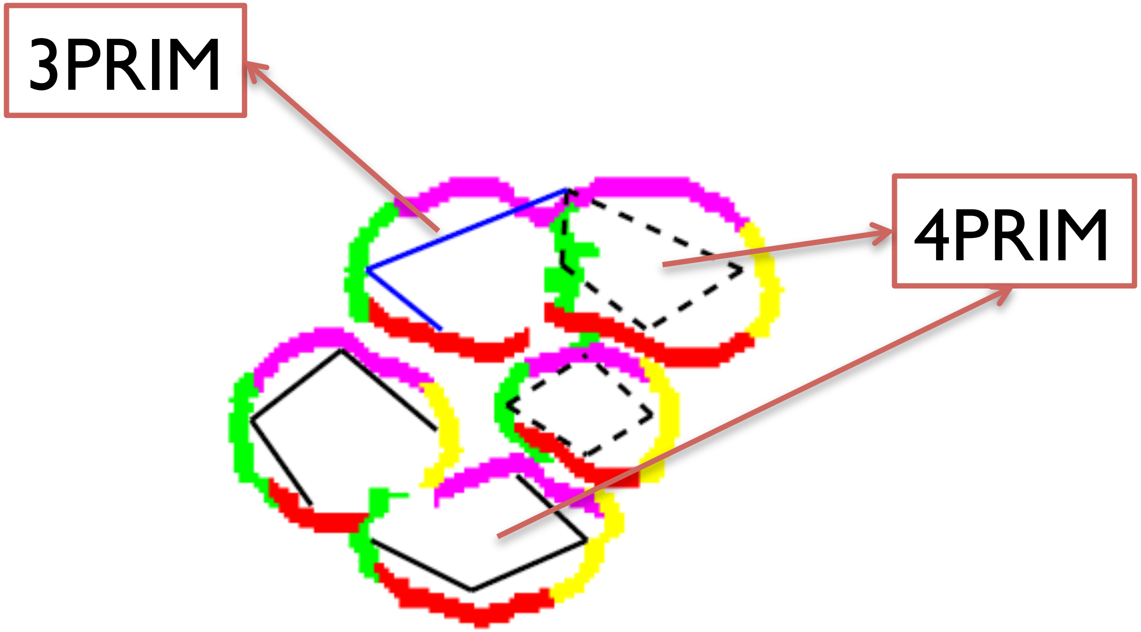
STOP

Percentage threshold

Standard deviation threshold

Cell Patterns

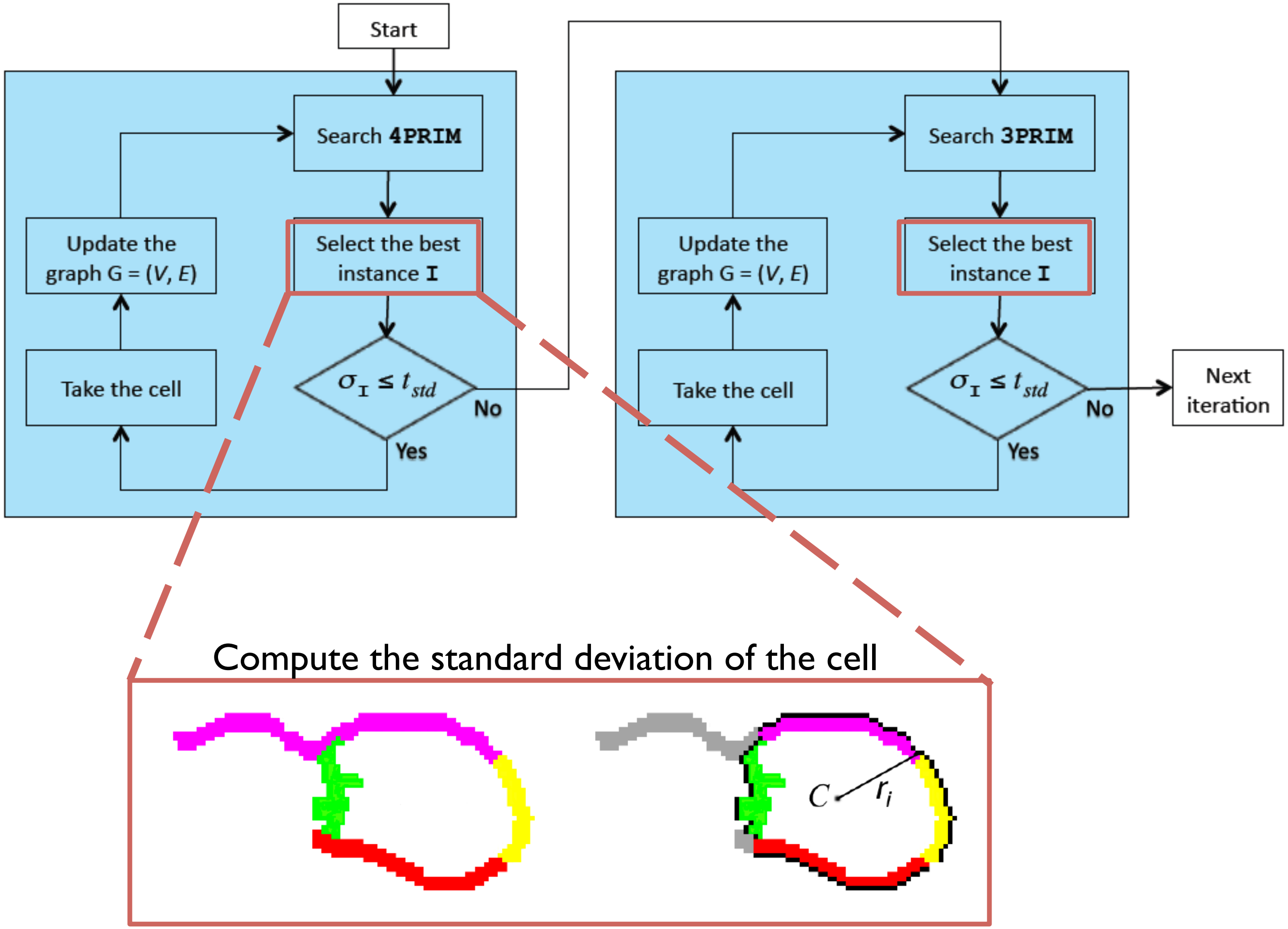
- Cell localization searches structural patterns in the constructed graph



3PRIM

4PRIM

Cell Localization



Start

Search **4PRIM**

Select the best instance **I**

$\sigma_{\mathbf{I}} \leq t_{std}$

No

Yes

Take the cell

Update the graph G = (*V*, *E*)

Search **3PRIM**

Select the best instance **I**

$\sigma_{\mathbf{I}} \leq t_{std}$

No

Yes

Take the cell

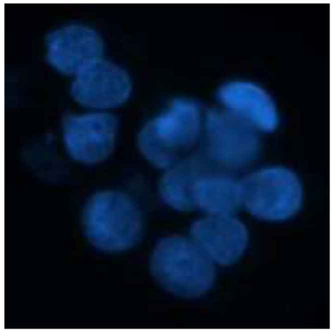
Update the graph G = (*V*, *E*)

Next iteration

Compute the standard deviation of the cell

Region Growing

- Starts flooding from the centroids of cell primitives in accordance to
  - A pixel <u>cannot</u> flood to outer boundaries of a primitive
  - For a cell, pixels reached after the last point that it floods to a primitive pixel are excluded



- Uses geodesic distance from a pixel to a marker as criterion

- Smoothes results by majority filtering

Experiments

- Images taken from Huh7 and HepG2 cell lines
  - A total of 2661 cells in 37 images
    - 16 Huh7 images with 1378 cells
    - 21 HepG2 images with 1283 cells

- HepG2 images are more confluent than the Huh7 images

- Gold standard images were annotated by experts

Comparisons

- Adaptive h-minima [1]
  - Binarization via active contours without edges
  - Identify markers with applying h-minima on inner distance map
  - Marking function based on outer distance transform and intensities

- Conditional erosion [2]
  - Binarization via global thresholding
  - Markers obtained by iterative erosions on the binary map

[1] **Cheng et al.**, 2009, *IEEE T Biomed Eng*, [2] **Yang et al.**, 2006, *IEEE T Circuits-I*

Evaluation

- Cell-based evaluation
  - How good an algorithm is in segmenting cells
  - Match computed and annotated cells
  - Compute precision, recall, and f-score
- Pixel-based
  - How good an algorithm is in delineating cells
  - Find true positive pixels of one-to-one matches
  - Compute precision, recall, and f-score

Cell-based evaluation

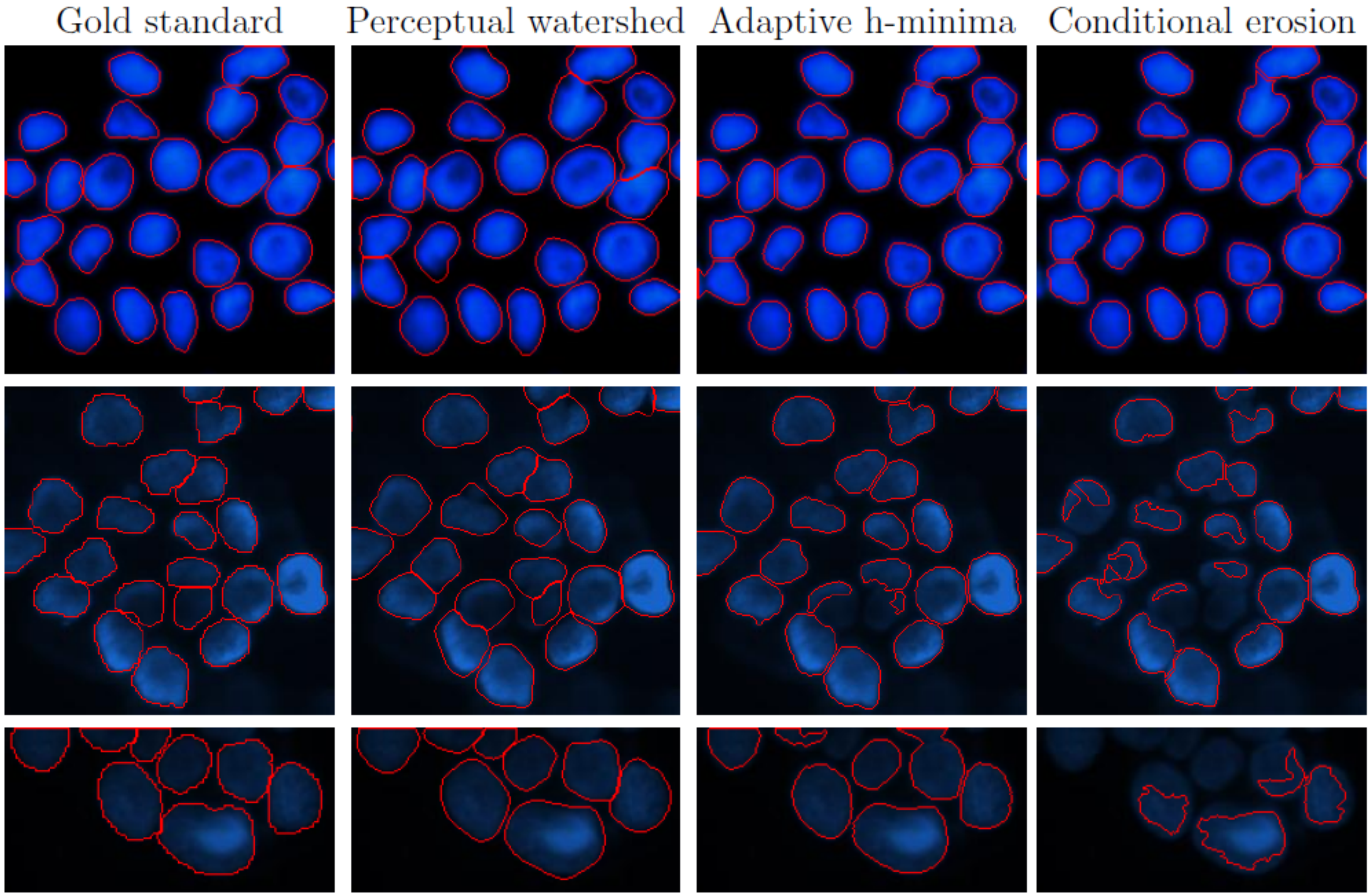
- One-to-one match: a computed cell and an annotated cell make a unique pair
  - Consider the overlapping pixels as true positive in *pixel-based evaluation*
- Oversegmention: an annotated cell matches with more than one computed cell
- Undersegmentation: a computed cell matches with more than one annotated cell
- Miss: an annotated cell matches with no computed cell
- False: a computed cell matches with no annotated cell

Parameter Selection

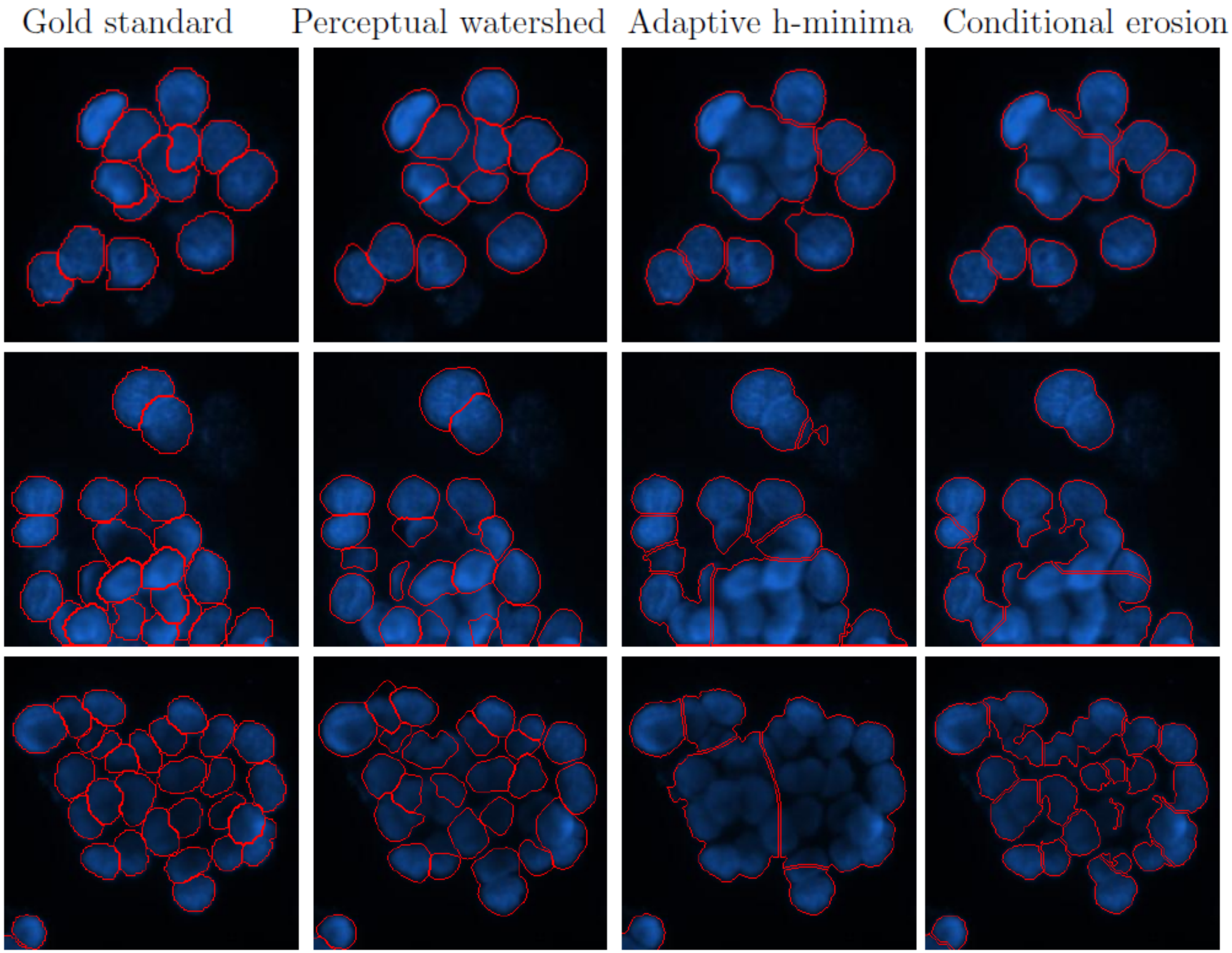
- Proposed algorithm has four external model parameters
  - The primitive length threshold $t_{size}=15$
  - The percentage threshold $t_{perc}=0.3$
  - The standard deviation threshold $t_{std}=4.0$
  - The radius of the structuring element $W=5$

- Internal parameter $d=3$
  - For selecting only *d*-outermost pixels in primitive definition

Visual Results - 1



Visual Results - 2



Quantitative Results

Cell based evaluation results for the Huh7 dataset

	One-to-one	Overseg.	Underseg.	Misses	Falses
Perceptual watershed	**1232**	**44**	**72**	**28**	**30**
Adaptive h-minima	1140	14	205	45	19
Conditional erosion	1131	35	138	61	74

Cell based evaluation results for the HepG2 dataset

	One-to-one	Overseg.	Underseg.	Misses	Falses
Perceptual watershed	**1002**	**53**	**153**	**62**	**75**
Adaptive h-minima	867	4	390	51	22
Conditional erosion	820	19	364	63	80

Quantitative Results

Precision, recall, f-score measures for the Huh7 dataset

	Cell-based			Pixel-based		
	Prec.	Recall	F-score	Prec.	Recall	F-score
Perceptual watershed	**89.15**	**89.40**	**89.28**	**77.87**	**86.11**	**81.78**
Adaptive h-minima	87.90	82.73	85.23	79.97	77.35	78.64
Conditional erosion	85.17	82.08	83.59	84.06	71.20	77.10

Precision, recall, f-score measures for the HepG2 dataset

	Cell-based			Pixel-based		
	Prec.	Recall	F-score	Prec.	Recall	F-score
Perceptual watershed	**80.22**	**78.10**	**79.15**	**65.22**	**74.54**	**69.57**
Adaptive h-minima	80.20	67.58	73.35	65.45	65.10	65.27
Conditional erosion	76.35	63.91	69.58	66.15	57.42	61.48

Discussion

- Our proposed method relies on boundary primitives
  - Adequate to locate cells even if their boundaries are partially present
  - Segmentation is not based on a binary map
    - Both conditional erosion and adaptive h-minima rely on their binary maps
    - If they cannot capture a foreground pixel it will be never segmented

Conclusions

- In this thesis, we proposed a new marker-controlled watershed algorithm inspired by human perception, which
  - Introduces high-level primitives for boundary representation
  - Constructs an attributed relational graph on these primitives to model their spatial relations
  - Searches for pre-defined patterns on the graph to locate markers
  - Guides flooding also using the primitives

Conclusions

- Both visual and quantitative results demonstrated that our proposed method
    - Can better handle segmentation problems in both isolated and confluent cells

- Its power comes from
    - Using the high-level boundary primitives
        - Instead of shape information via binarization

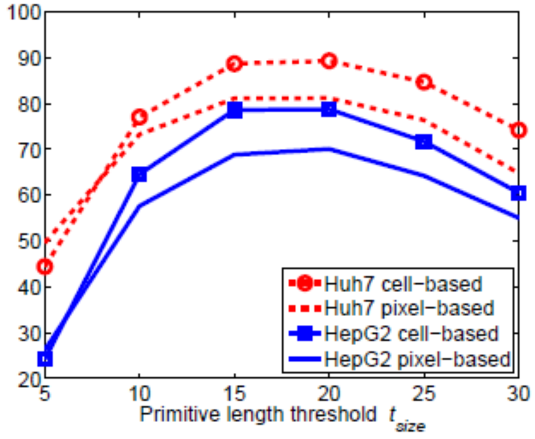
Future Work

- One future research direction is to implement a primitive-distance based marking function
    - This is expected to yield better segmentation boundaries
- Segmented cells may be further used for quantitative analysis in different applications of molecular biology research

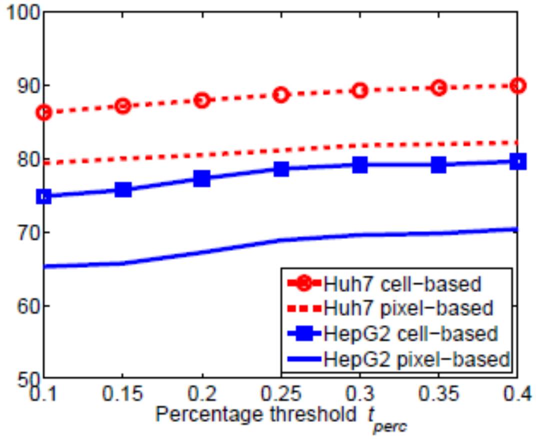
Q&A

- Thanks for listening
- Any questions are welcomed

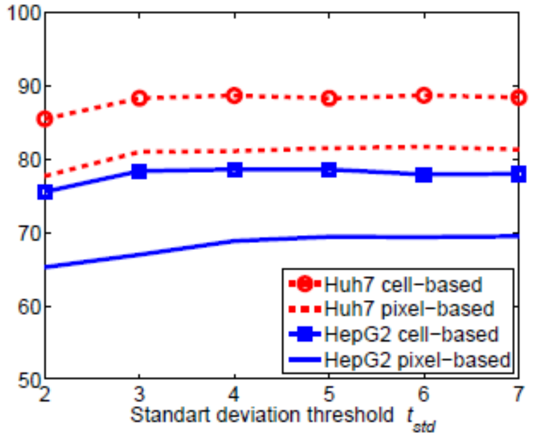
Parameter Analysis



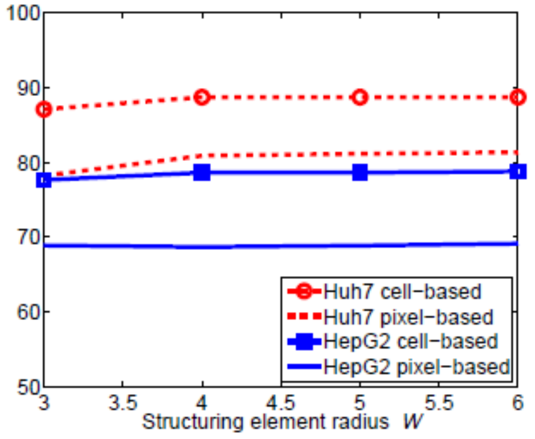
(a)



(b)



(c)



(d)