Unsupervised Unstained Cell Detection using SIFT Keypoint Clustering and Laplacian Boundary Potential (2)

Firas Mualla
27.01.2014
Pattern Recognition Lab (CS 5)

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

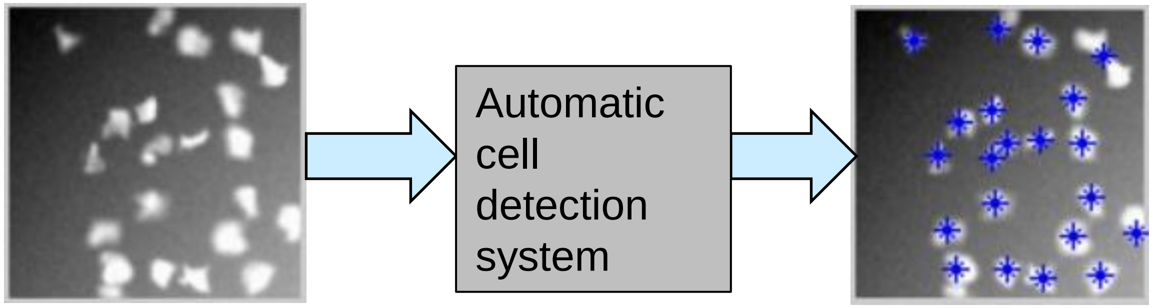
Outlines

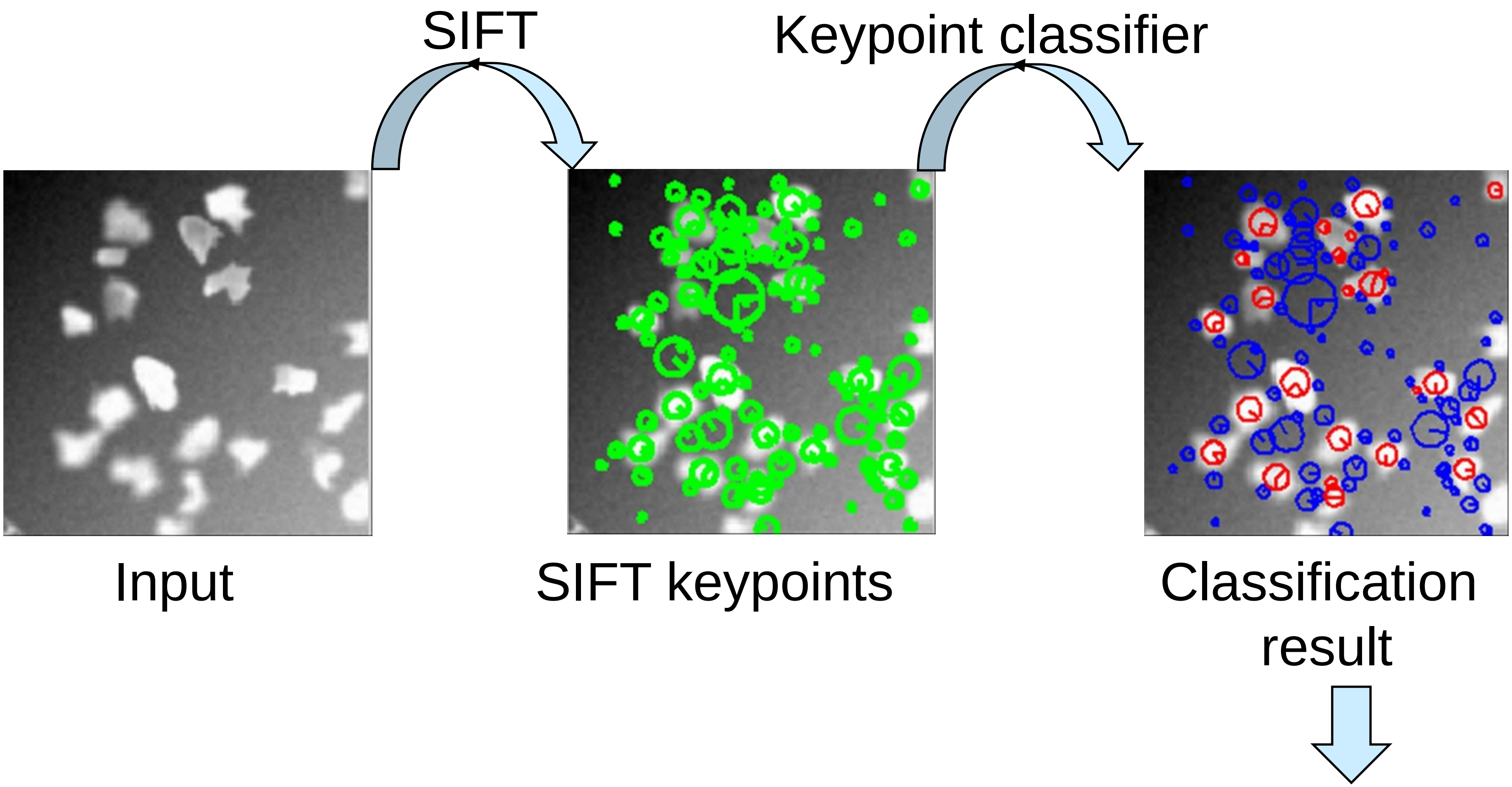
- Reminder of our supervised cell detection system
- Unsupervised approach
- Evaluation
- Conclusion & outlook

Reminder

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

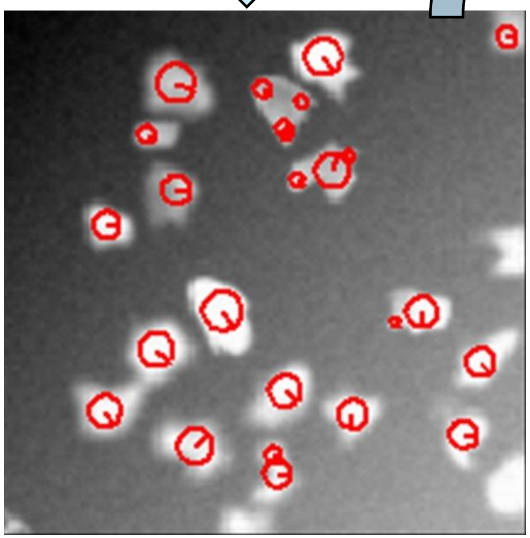
The system



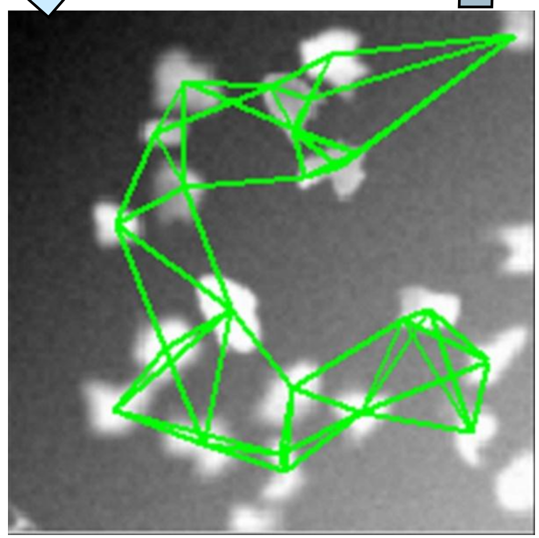

SIFT
Keypoint classifier
Input
SIFT keypoints
Classification result

Cut profiles

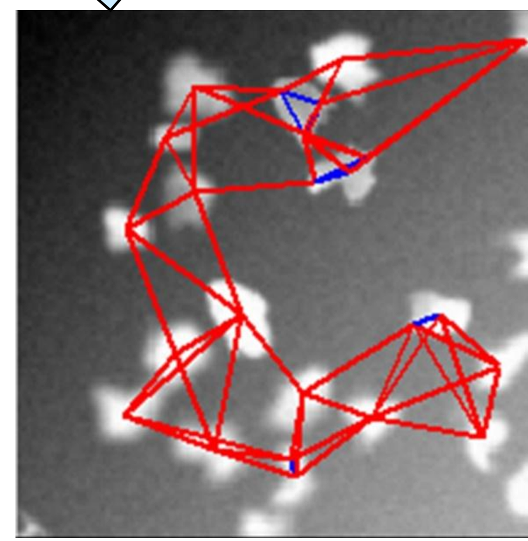
Profile classifier



Cell keypoints



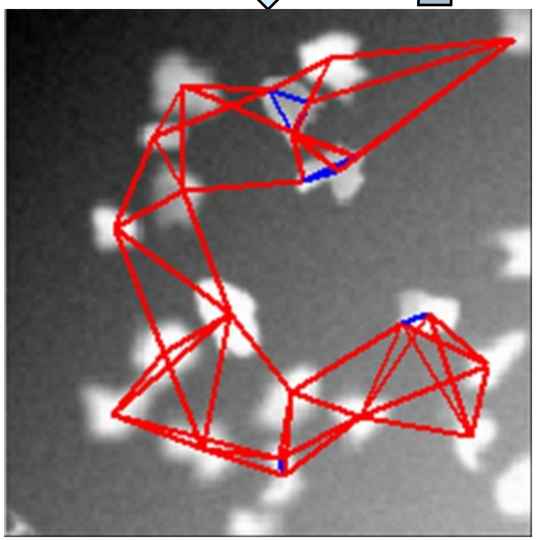
Profiles

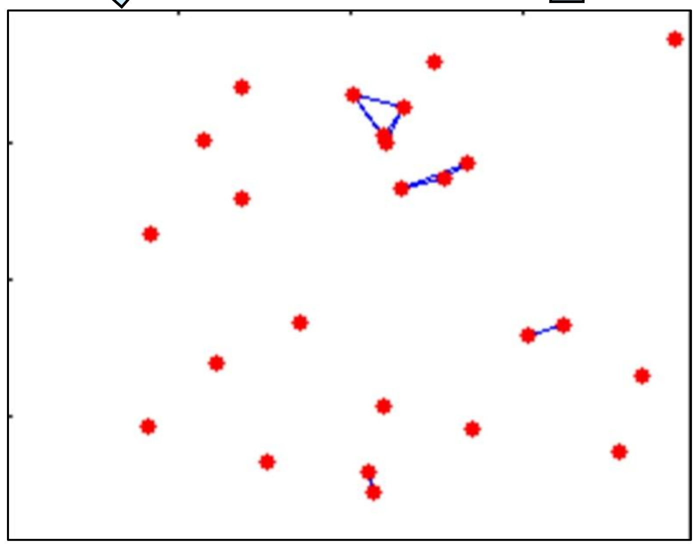


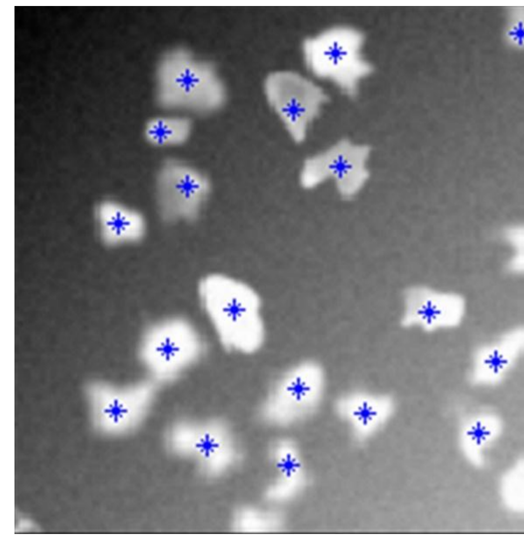
Classification result

Hierarchical clustering

Average







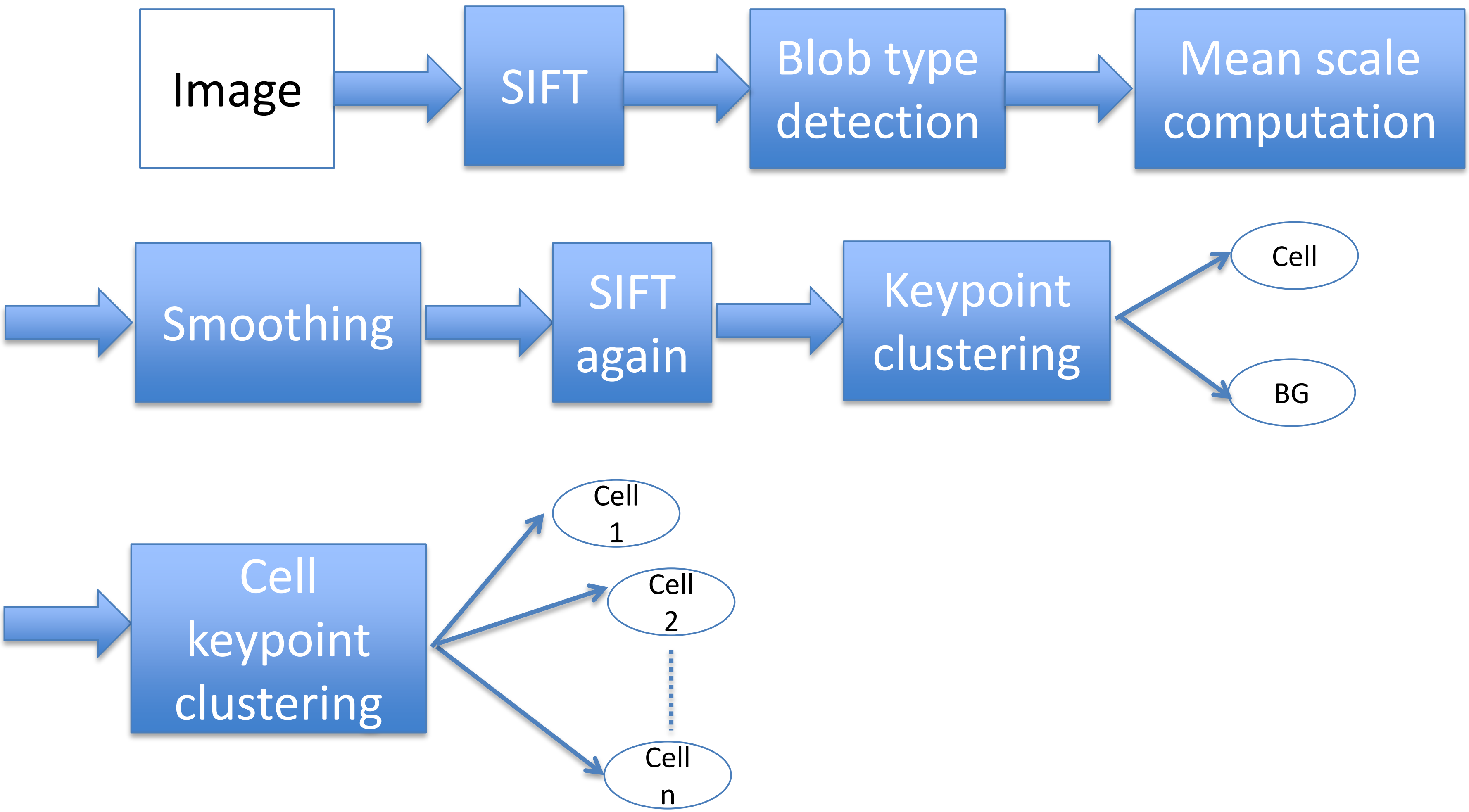
End

So far!

+ Fully automatic cell detection
+ High detection rate
+ High scores in scale-, orientation-, and illumination-invariance
+ General design
- Needs training

Unsupervised approach

System overview



Blob type detection

or

?

Positive

Negative

> 0.5

< 0.5

$$\omega_i = \frac{s(\mathbf{p}_i)}{\mathrm{score}(\mathbf{p}_i)}$$

- $s(\mathbf{p}_i)$: scale of the keypoint $\mathbf{p}_i$.
- $\mathrm{score}(\mathbf{p}_i)$: non-circularity measure.

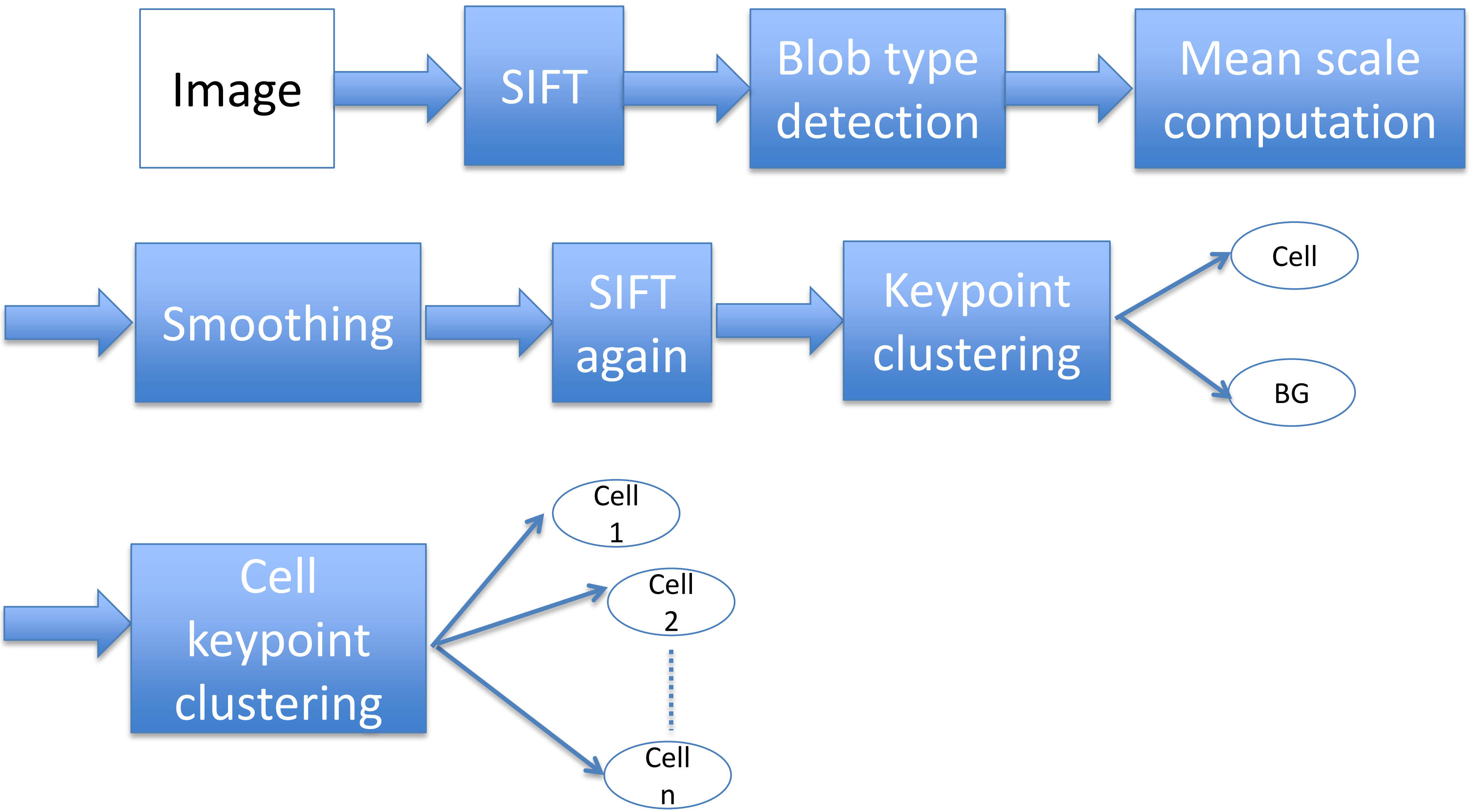
$$\mathrm{PositiveFit} = \frac{\sum_{i=1}^{N} \omega_i |DOG(\mathbf{p}_i)| H(DOG(\mathbf{p}_i))}{\sum_{i=1}^{N} \omega_i |DOG(\mathbf{p}_i)|}$$

- $\mathbf{p}_i$, $i = 1..N$ are the keypoints in the image.
- $DOG(\mathbf{p}_i)$: difference of Gaussians value.
- H is the Heaviside step function.

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

System overview

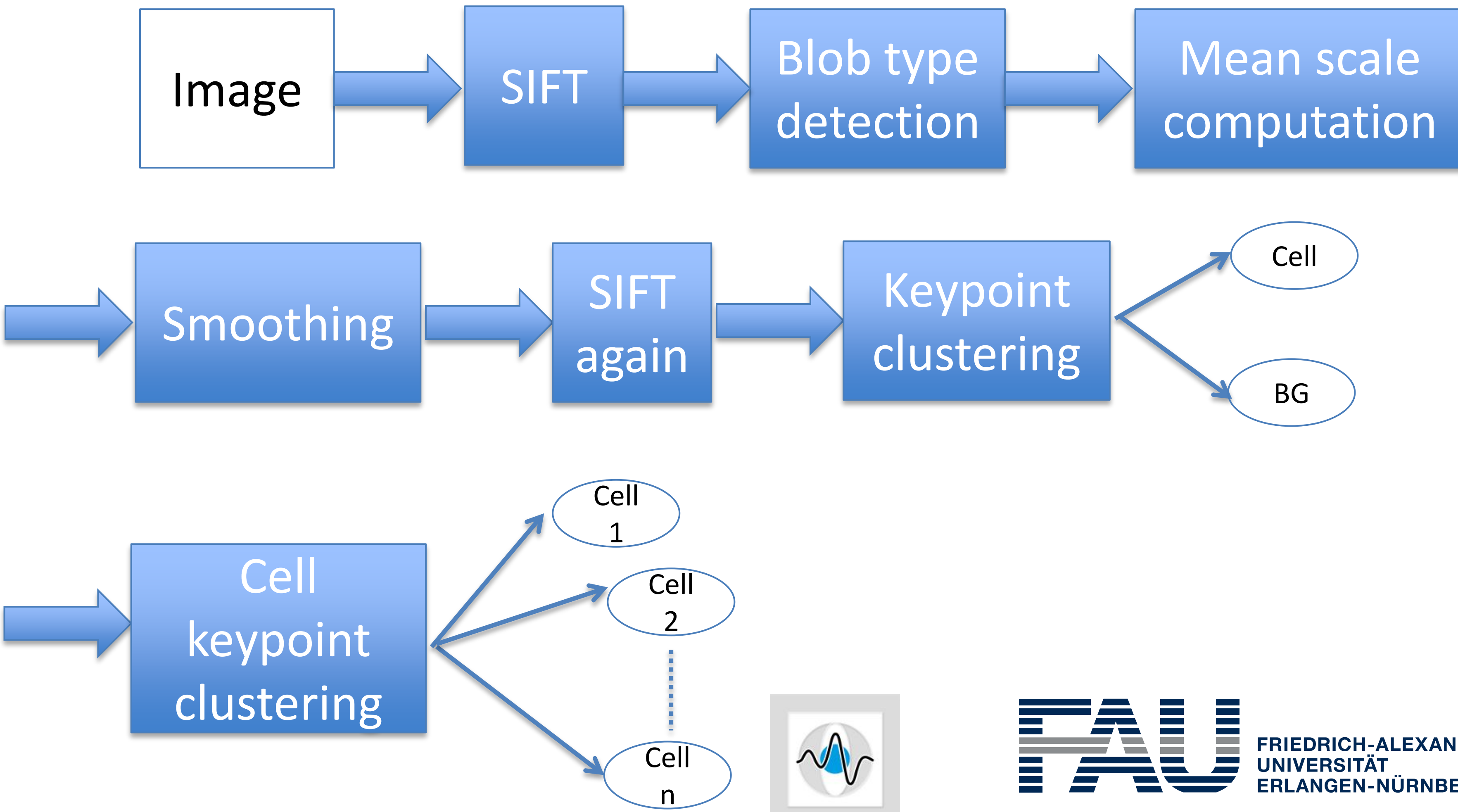


Mean scale computation

$$SE = \frac{\sum_{i=1}^{M} |DOG(\mathbf{p}_i)| s(\mathbf{p}_i)}{\sum_{i=1}^{M} |DOG(\mathbf{p}_i)|}$$

- $s(\mathbf{p}_i)$: scale of the keypoint $\mathbf{p}_i$.
- $DOG(\mathbf{p}_i)$: difference of Gaussians value.
- $\mathbf{p}_i$, $i = 1..M$ are the one-sided keypoints (after blob type detection).

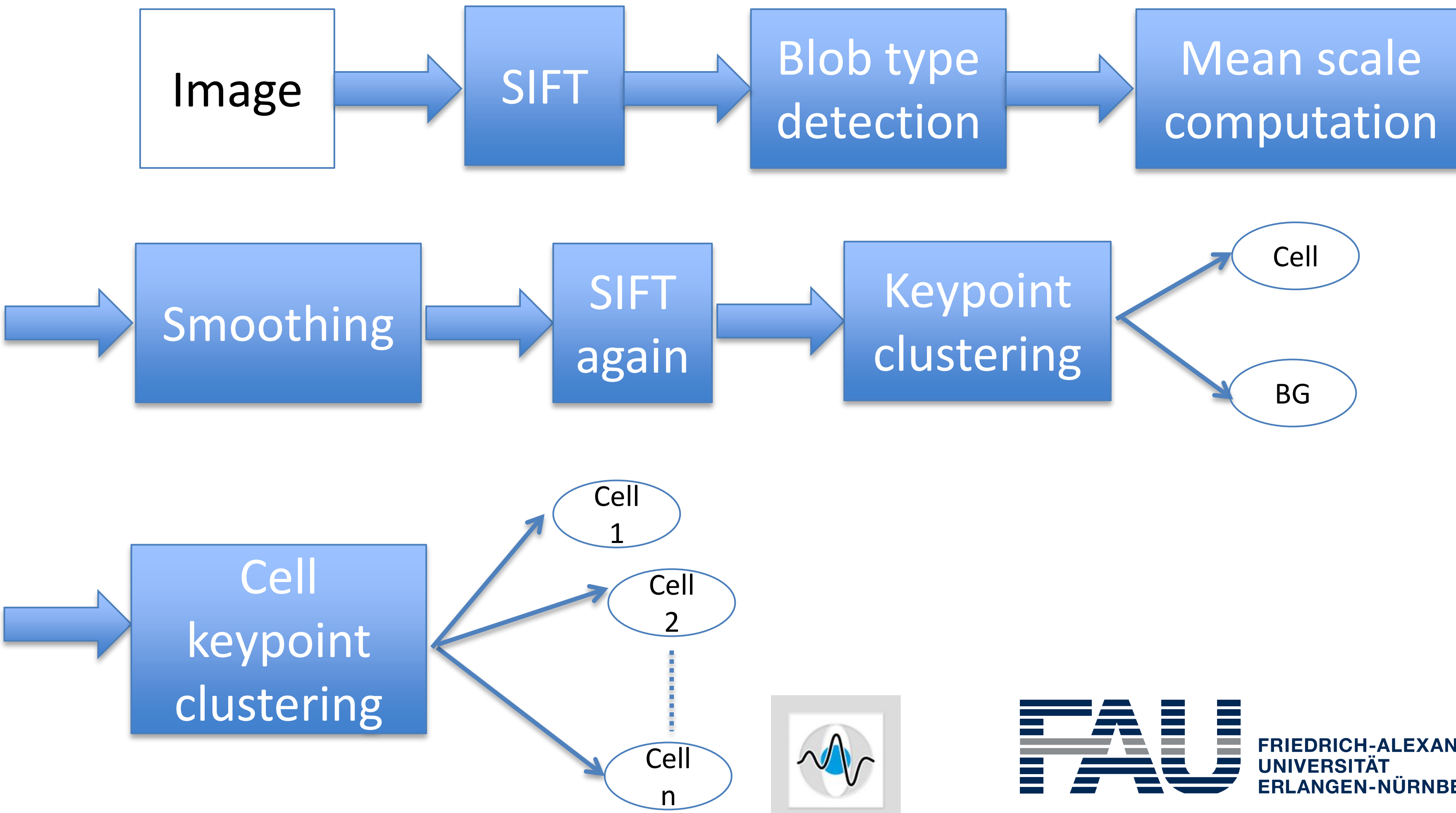
System overview



Smoothing

- Image is smoothed with a Gaussian kernel.
- Its standard deviation is the previous mean scale.
- Applying SIFT again:
  - The new keypoints are more stable.
  - Their number is considerably less.

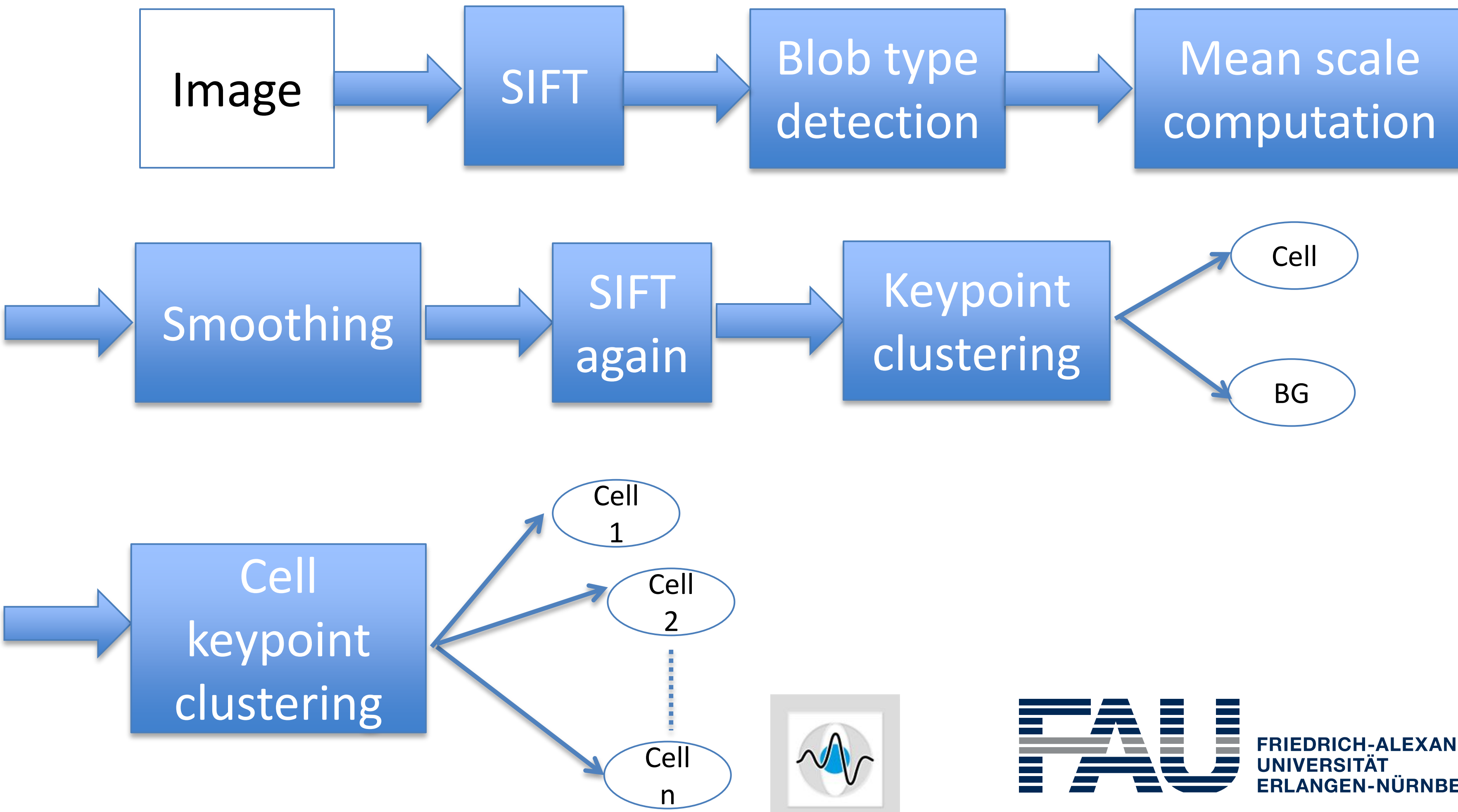
System overview



Keypoint clustering

- K-means clustering
- Clusters which correspond to 1D Otsu thresholding of the DOG values are used as initialization.
- City-block distance measure
- Modality-specific features:
  - DOG and score were used in phase contrast.
  - DOG and intensity were used in bright-field.
- Applying SIFT again seems to have considerable effect on the result!

System overview

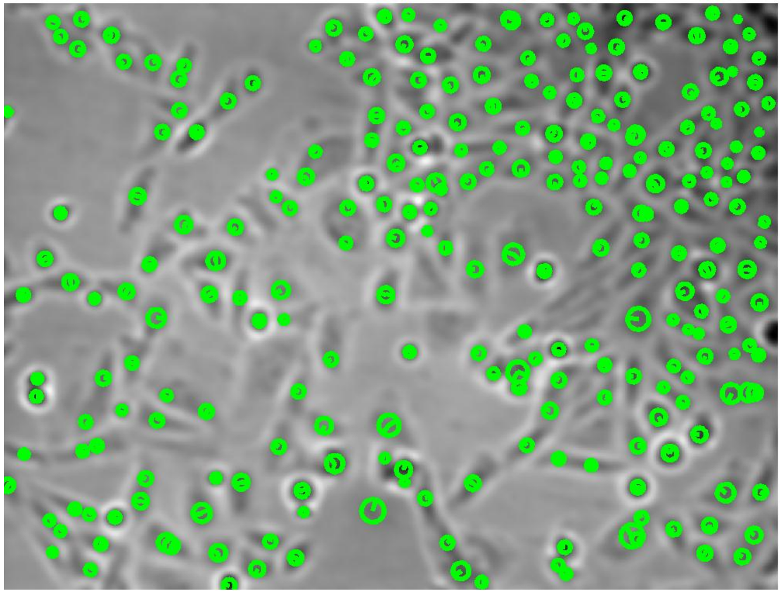


FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

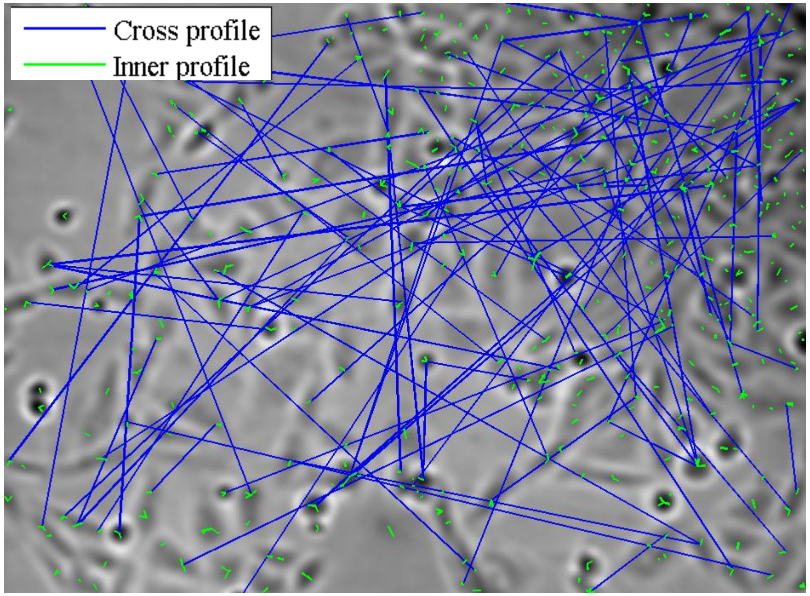
TECHNISCHE FAKULTÄT

Cell keypoint clustering

- Back to **intuition**:



Cell SIFT keypoints



Artificial inner and cross profiles

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Sample inner profiles

- Randomly sample M_1 keypoints from the resulting cell keypoints.
  - Use a triangular distribution in scale.
  - M_1 was set to 100 in our experiments.
- For each keypoint $\mathbf{p}_i$, pick a random angle θ_i and extract a profile between the following two points:
  - $\mathbf{p}_i$
  - $\mathbf{p}_i + (\mathrm{s}(\mathbf{p}_i)\cos(\theta_i), \mathrm{s}(\mathbf{p}_i)\sin(\theta_i))$
- Assign label "inner" to the extracted profiles.

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Sample cross profiles

- Compute the pairwise Euclidean distance matrix of the cell keypoints.
- Randomly pick M_2 distances more than $C \cdot SE$:
  - M_2 was set to 100.
  - C was set to 10.
- Assign label "cross" to the corresponding profiles.

Boundary potential

- Laplacian of Gaussian L of the **smoothed** image is used as boundary potential.
  - L(blob-type * L > 0) = 0, (blob-type = +1 or -1)
  - L = |L|

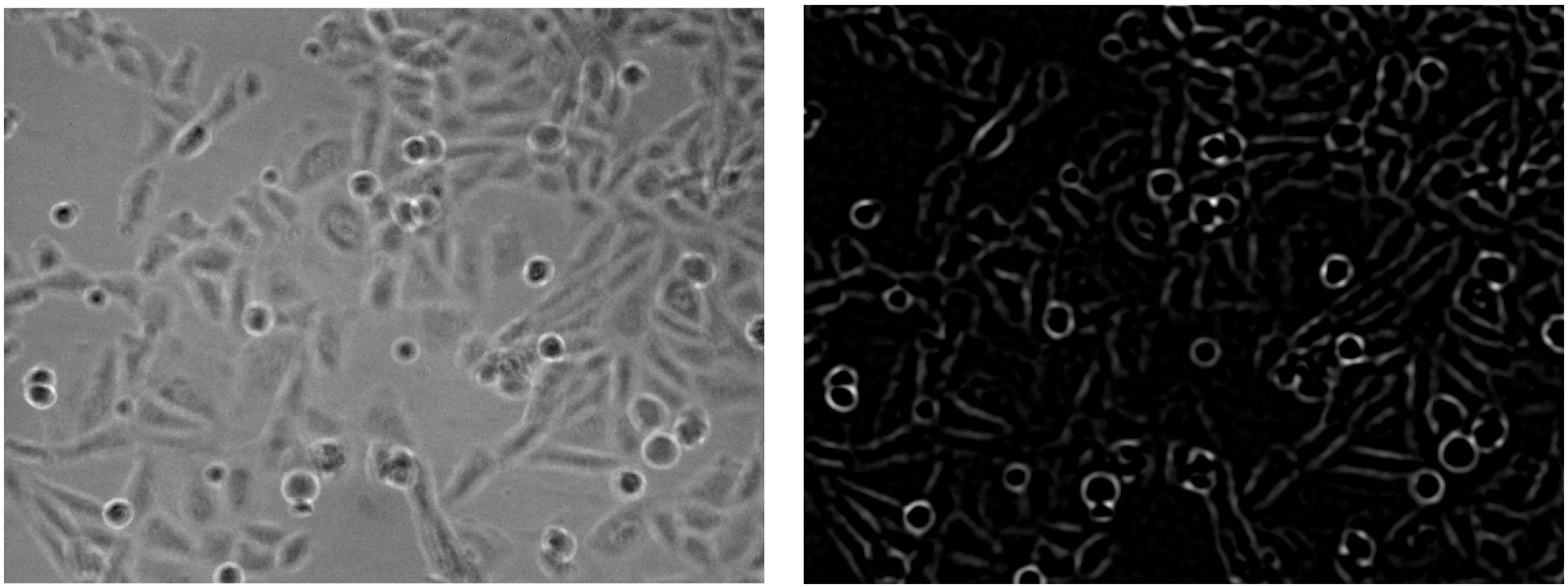


Image of CHO cells and its boundary potential

Training

- Extract features which are invariant to profile length.
  - Currently we use max(L) along the extracted profile.
- Use the extracted features to train a Bayesian classifier with Gaussian class conditional densities.

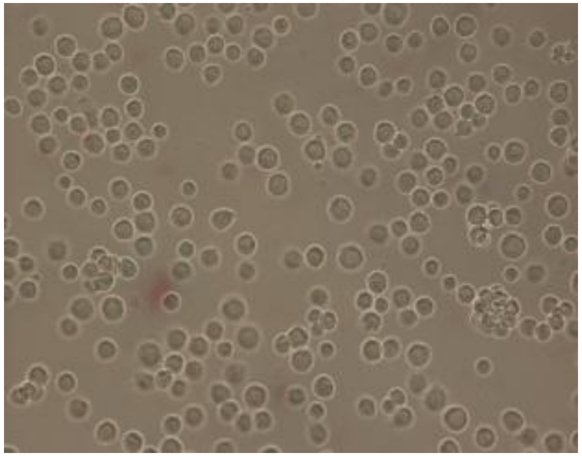
Cell keypoint clustering

- For each cell keypoint:
  - Find its K-nearest neighbors (K=3).
  - Classify the corresponding profiles using the Bayesian classifier.
- Similar to the supervised approach:
  - Agglomerative hierarchical clustering
  - Linkage method: average

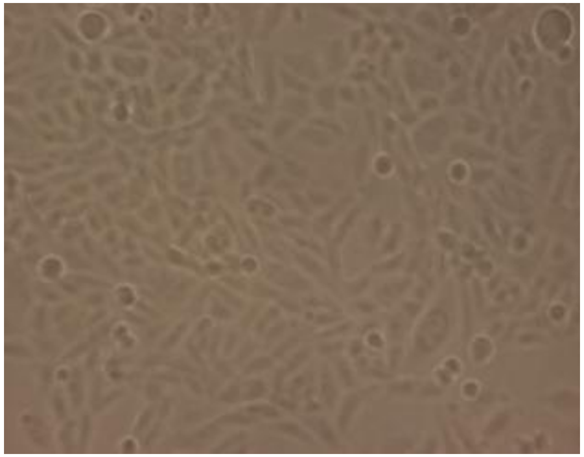
Evaluation

Bright-field

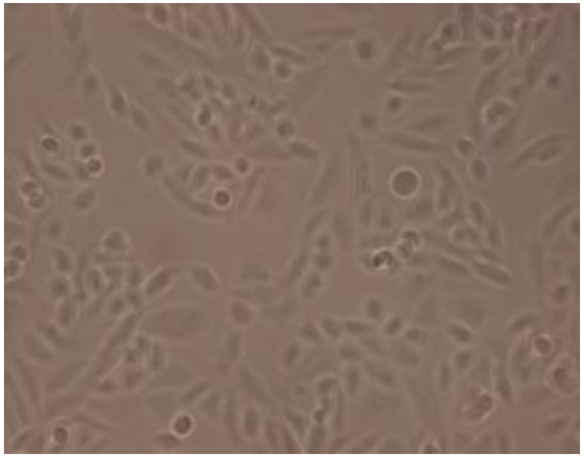
- 16 images
- Three cell lines
- More than 3500 manually labeled cells
- Ground truth type: border delineation



Sf21 cells



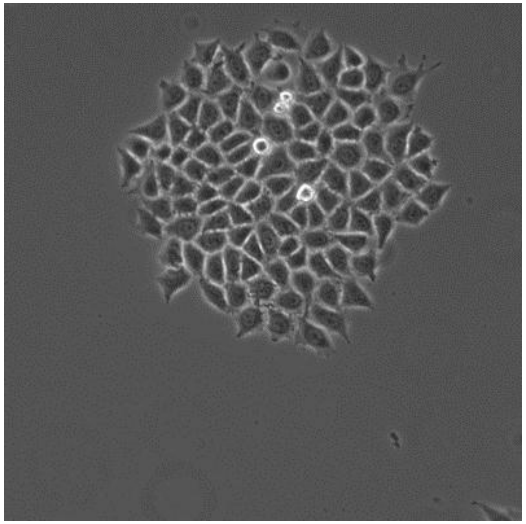
CHO cells



L929 cells

Phase contrast

- 11 images
- One cell line
- More than 1100 cells
- Ground truth type: dot at each cell center



HeLa cells

Results

	CHO	L929	Sf21	HeLa
F-Measure	80.58 %	84.31 %	88.26 %	85.72 %

Comparison with state-of-the-art: bright-field

	Mualla et al.[1]	Our approach	Becattini et al.[2]
F-measure	87.85 %	81.24 %	69.64 %

One CHO image was used to train [1] and the rest of the CHO images were used for evaluating the three approaches

[1] Mualla F, Schöll S, Sommerfeldt B, Maier A, Hornegger J (2013) Automatic cell detection in bright-field microscope images using SIFT, random forests, and hierarchical clustering. IEEE Trans Med Image 32(12):2274–2286.

[2] Becattini et al "A novel framework for automated targeting of unstained living cells in bright field microscopy," in Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, April 2011, pp. 195–198.

Comparison with state-of-the-art: phase contrast

	Arteta et al.[3]	Our approach
F-Measure	87.95 %	85.74 %

[3] C. Arteta, V.S. Lempitsky, J.A. Noble, and A. Zisserman, "Learning to detect cells using non-overlapping extremal regions", ;in Proc. MICCAI (1), 2012, pp.348-356.

Conclusions & outlook

Conclusions

- We have an unsupervised approach which has competitive performance compared to supervised state-of-the-art approaches on bright-field and phase contrast microscopy.
- Smoothing with mean image scale seems to provide:
  - Stable keypoints
  - Very good boundary potential between cells
- Use extreme cases when you do not have ground truth.

Outlook

- More features?
- More clever inner/cross profile selection?
- Bottleneck: Background / cell clustering

Thank you very much!