A Process Flow for Classification and Clustering of Fruit Fly Gene Expression Patterns

Andreas Heffel, Peter F. Stadler, Sonja J. Prohaska,

Gerhard Kauer, Jens-Peer Kuska

Santa Fe Institute

Department of Computer Science,
University of Leipzig

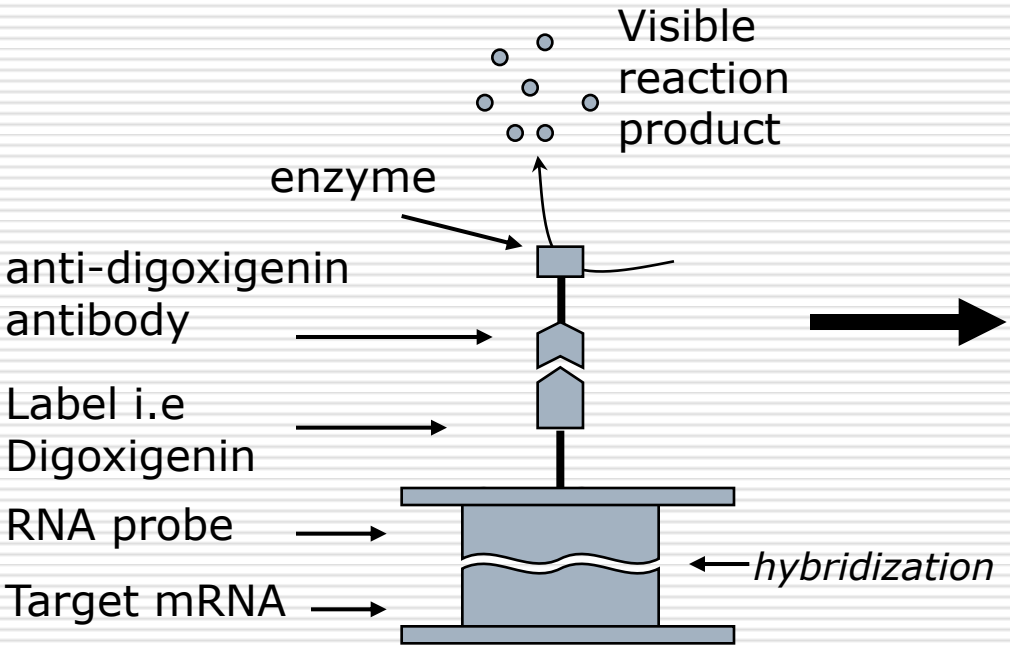
RNomics Group, Fraunhofer Institute
for Cell Therapy and Immunology,
Leipzig

Outline

- Gene expression images
  - Background and motivation
- The elaborated Processing Pipeline
  - Embryo Shape Segmentation
  - Allignement / Registration
  - GEP Extraction
  - GEP Classification / Representation
  - GEP Clustering
- Conclusion

Drosophila melanogaster Gene Expression Patterns (GEP)

- **whole-mount mRNA In situ hybridization**



Gene example: Slp1 (Stage 4)

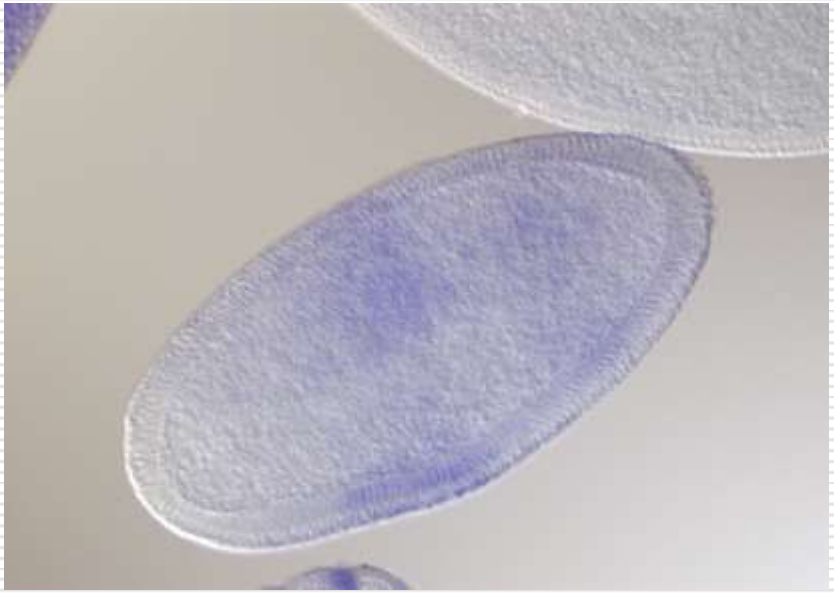
Drosophila GEP Projects

- Data management, storage, access, and integration:
  - Berkeley Drosophila Genome Project
  - FlyBase
  - …
- Data processing, analysis and observation:
  - FlyExpress
  - …
  - Our Appoach

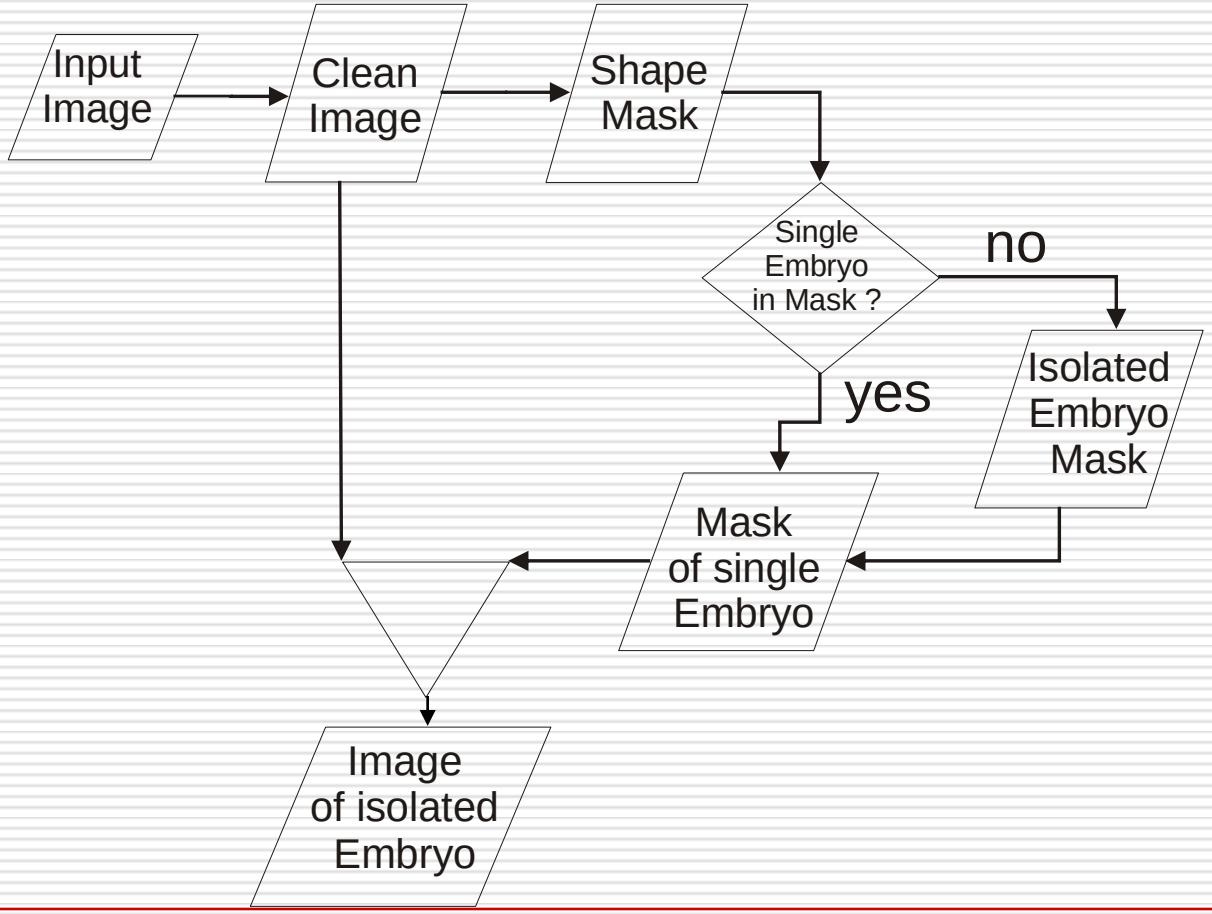
Imaging Complications

poor contrast &
background shading

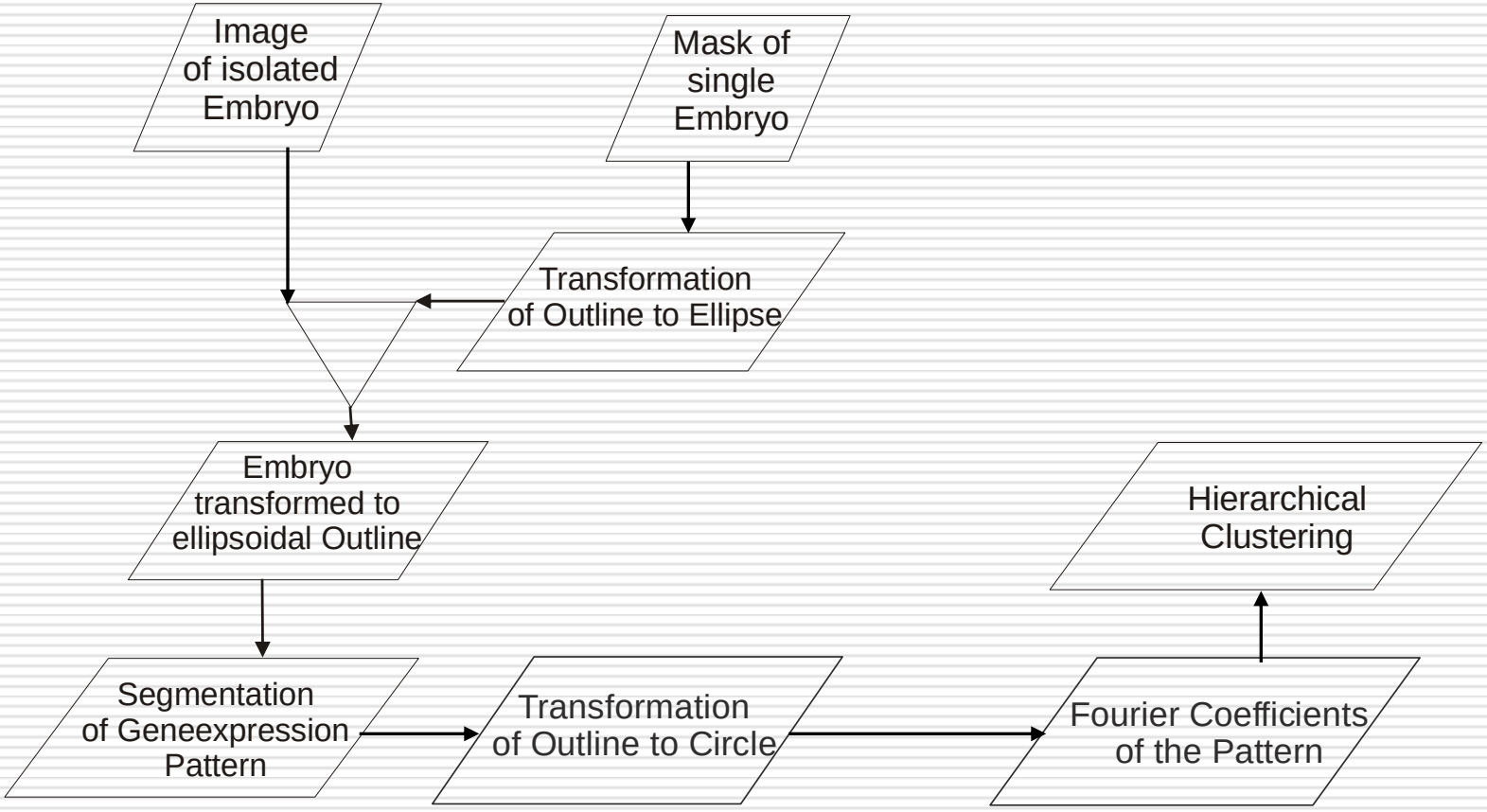
coherent partial embryos



Processing Pipeline Overview



Processing Pipeline Overview



Preprocessing

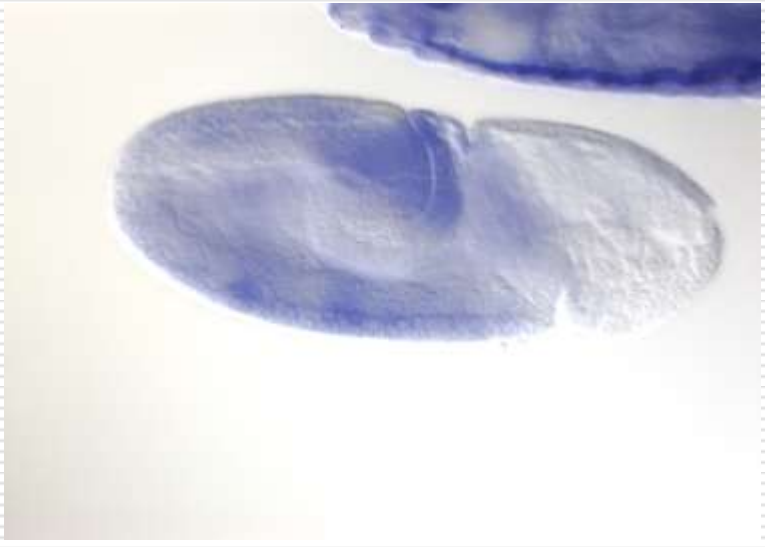
- Shading correction
- Contrast optimization

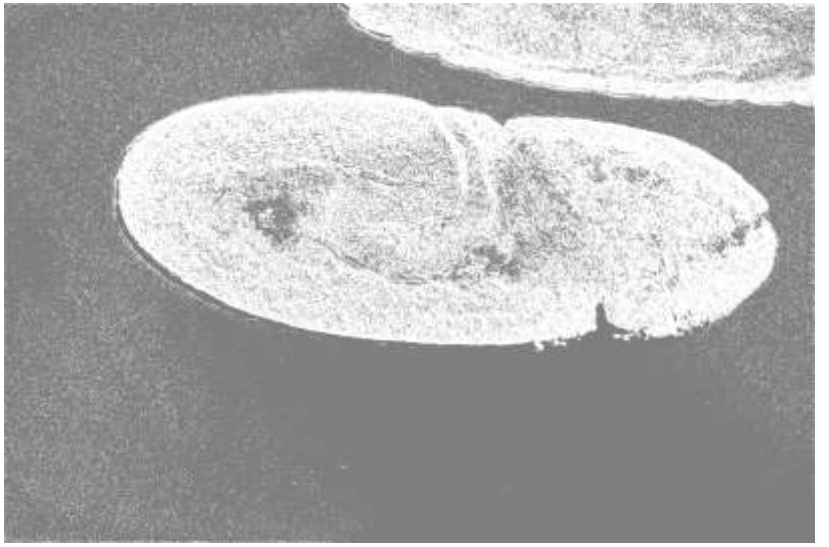
Shape Segmentation

- Feature space:

  gradient magnitude
- Method:

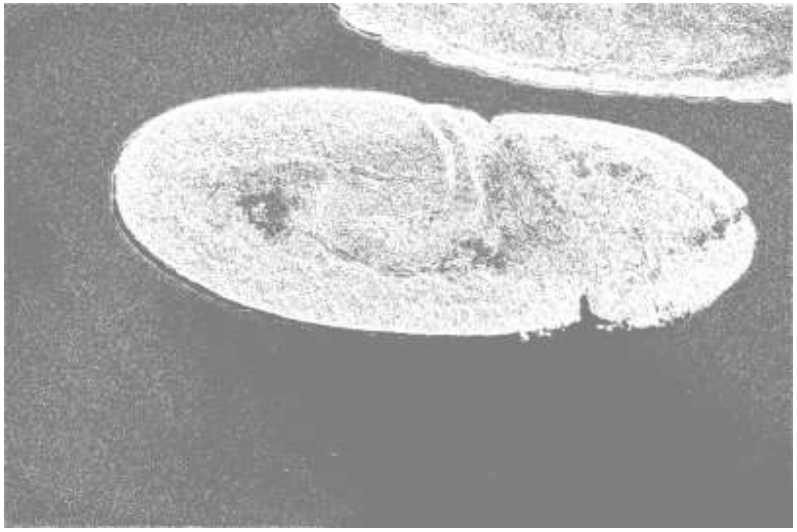
  Estimating Gaussian Mixture Densities with EM

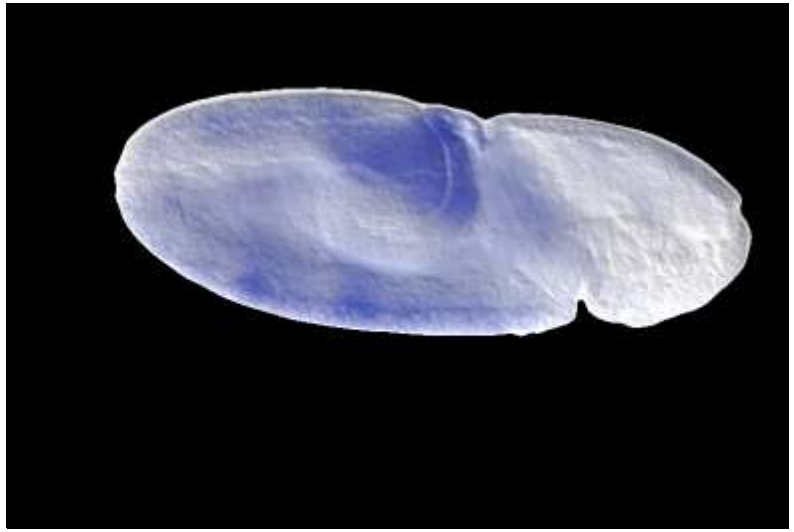




Shape Segmentation

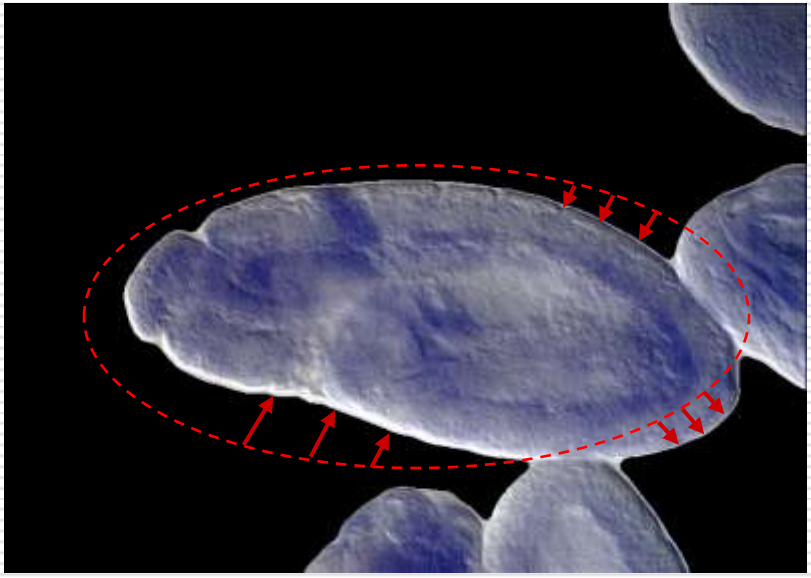
- Denoising: Total variation filter
- Close holes
- Remove other partial embryos

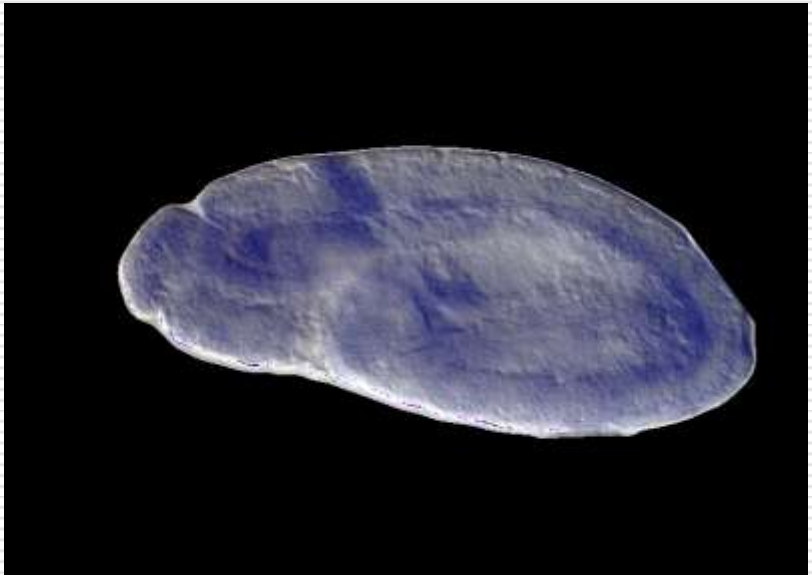




Shape Segmentation - Isolate Coherent Embryos

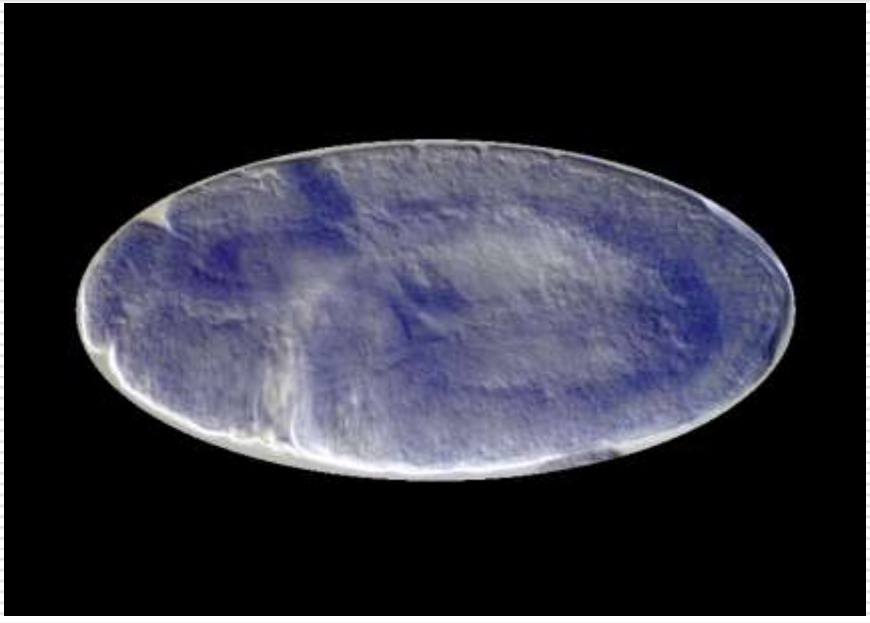
- Active Contour Approach - Snake Segmentation
- Marker particles are placed along an initial ellipsoidal contour.
  - -> Evolution toward maximum gradient regions





Transformation of Outline to Ellips

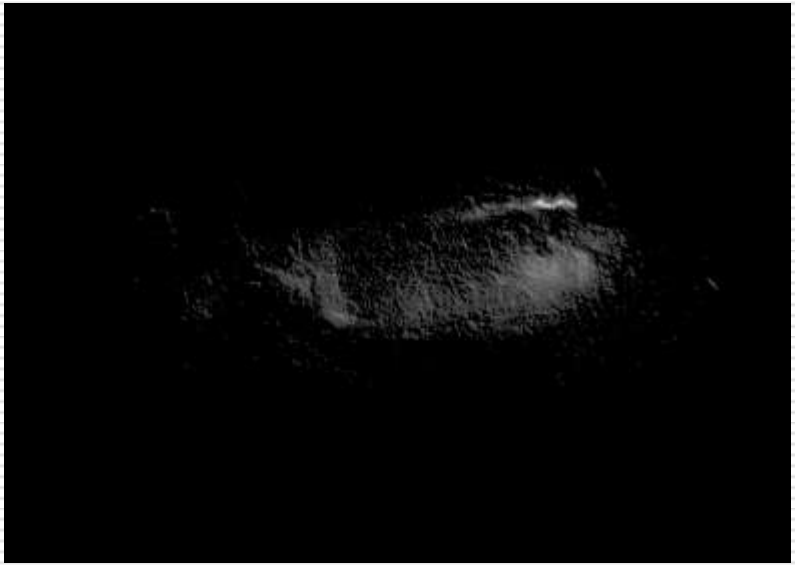
- Rigid Registration
- curvature based Nonlinear Registration



Segmentation of the GEP

- HSV Colorspace Transformation

V–channel, T=20%



GEP Classification
Fourier Coefficients

- The patterns are described by a set of Fourier coefficients.

$$\mathcal{P}(r,\phi) = \sum_{j=1}^{\infty}\sum_{k=0}^{\infty} a_{j,k}\,\psi_{j,k}(r,\phi)$$

- As basis, the eigenfunctions of the Laplace operator on a circle of radius l are used.

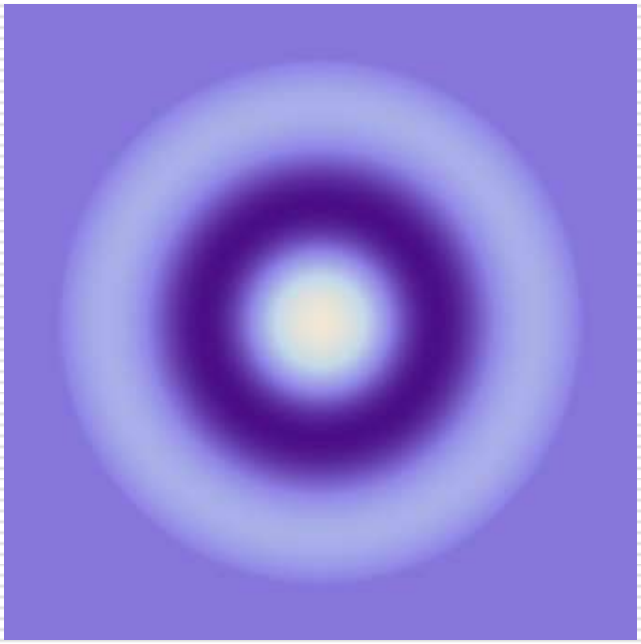
$$\psi_{j,k}(r,\phi) = N_{j,k}\, e^{ik\phi} J_k\left(\frac{r j_{k,j}}{\ell}\right)$$

Complete orthonormal system

$$\psi_{j,k}(r,\phi) = N_{j,k}\, e^{ik\phi} J_k\left(\frac{r j_{k,j}}{\ell}\right)$$

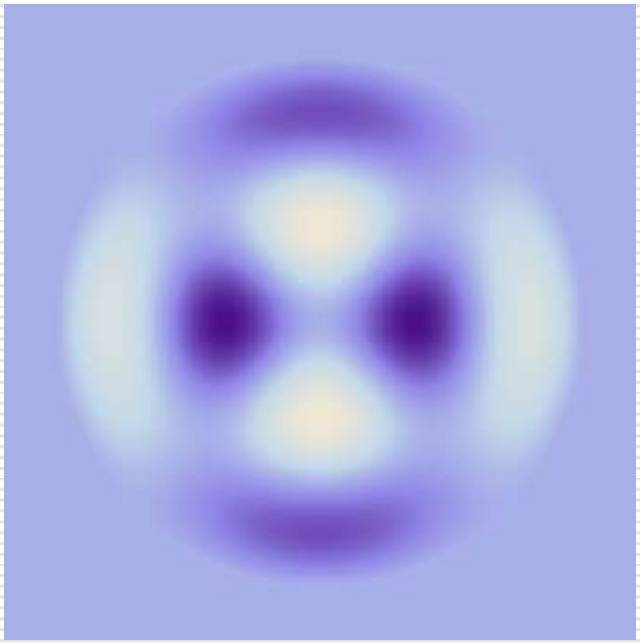
k=0

j=3



k=2

j=2



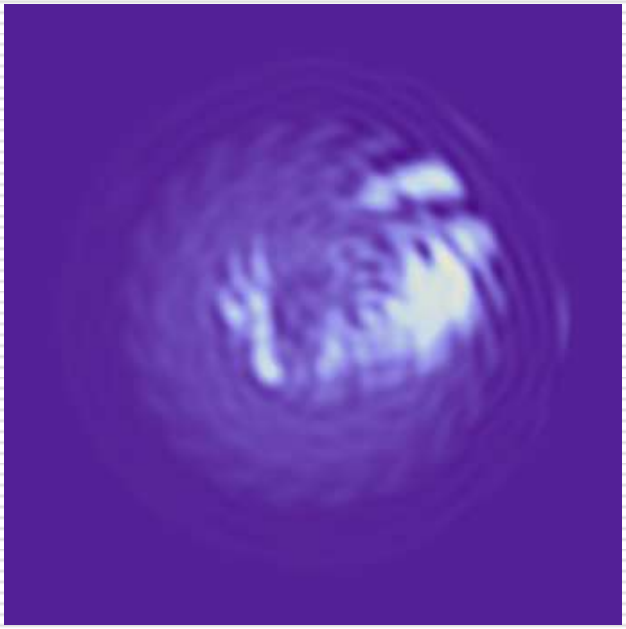
Representation with a set of 420 Fourier coefficients

$$a_{j,k} = \int_0^{\ell} \int_0^{2\pi} \psi^*_{j',k'}(r,\phi) g(r,\phi)\, r\, \mathrm{d}\phi \mathrm{d}r$$

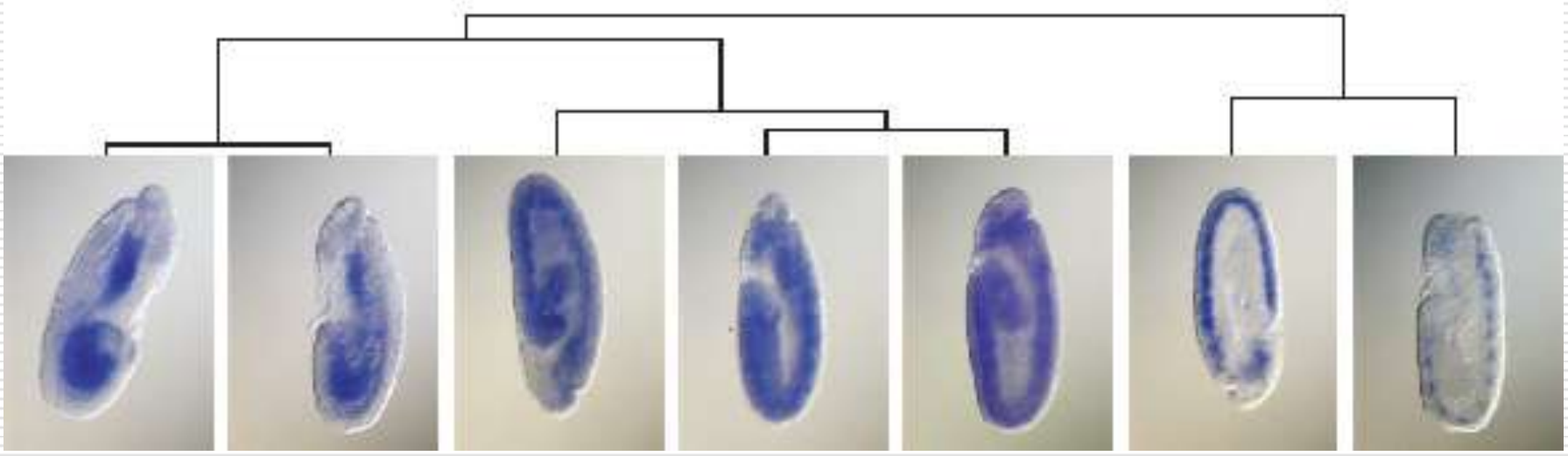
420 Fourier Coefficients

$k \in [0,\ldots,20]$

$j \in [1,\ldots,20]$



GEP Clustering



- Hierarchical clustering of the absolute values of the coefficient sets using Euclidean norm.

Conclusion

- Clustering results show agreement with the visual expectation
- The snake segmentation accuracy can be improved
- The “orientation problem“ should be solved
- Future Work:
  Investigate biological relevance of the results

Thank you for your attention!

[1] "Bdgp: Berkeley drosophila genome project," http://www.fruitfly.org/

[2] "Flybase: A database of drosophila genes and genomes," http://flybase.bio.indiana.edu/

[3] Sudhir Kumar, Karthik Jayaraman, Sethuraman Panchanathan, Rajalakshmi Gurunathan, Ana Marti-Subirana, and Stuart J. Newfeld, "Best: a novel computational approach for comparing gene expression patterns from early stages of drosophila melanogaster development," Genetics, vol. 162, no. 4, pp. 2037–2047, 2002.

[4] Chenyang Xu and Jerry L. Prince, "Gradient vector flow: A new external force for snakes," IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR), Los Alamitos: Comp. Soc. Press, vol. 1997, pp. 66–71, June.

[5] Franz Pernkopf and Djamel Bouchaffra, "Genetic-based EM algorithm for learning Gaussian mixture models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1344–1348, 2005.

[6] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," International Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, November 2004.

[7] Ulf-Dietrich Braumann and Jens-Peer Kuska, "A new equation for nonlinear image registration with control over the vortex structure in the displacement field," in Proceedings of the IEEE International Conference on Image Processing. Oct. 2006, pp. 329–332, IEEE Signal Processing Society.

...