UNIVERSITÄT LEIPZIG

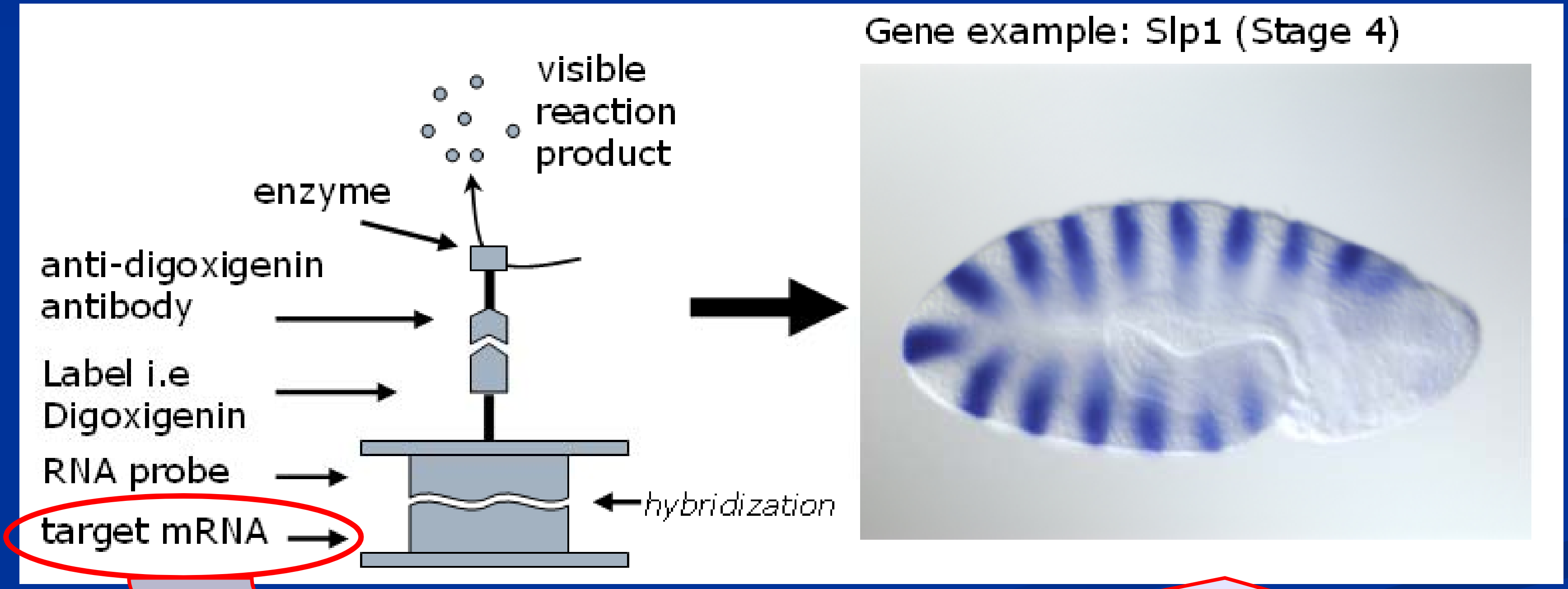
Fortgeschrittene Methoden in der Bioinformatik

IMAGE PROCESSING - FRUIT FLY GENE EXPRESSION PATTERN ANALYSIS

Andreas Heffel (Dipl. Bioinf.)
Raum 440.1
Tel: 0341 97 - 16665
E-Mail: andreas@bioinf.uni-leipzig.de

Drosophila melanogaster
Gene Expression Patterns (GEP)

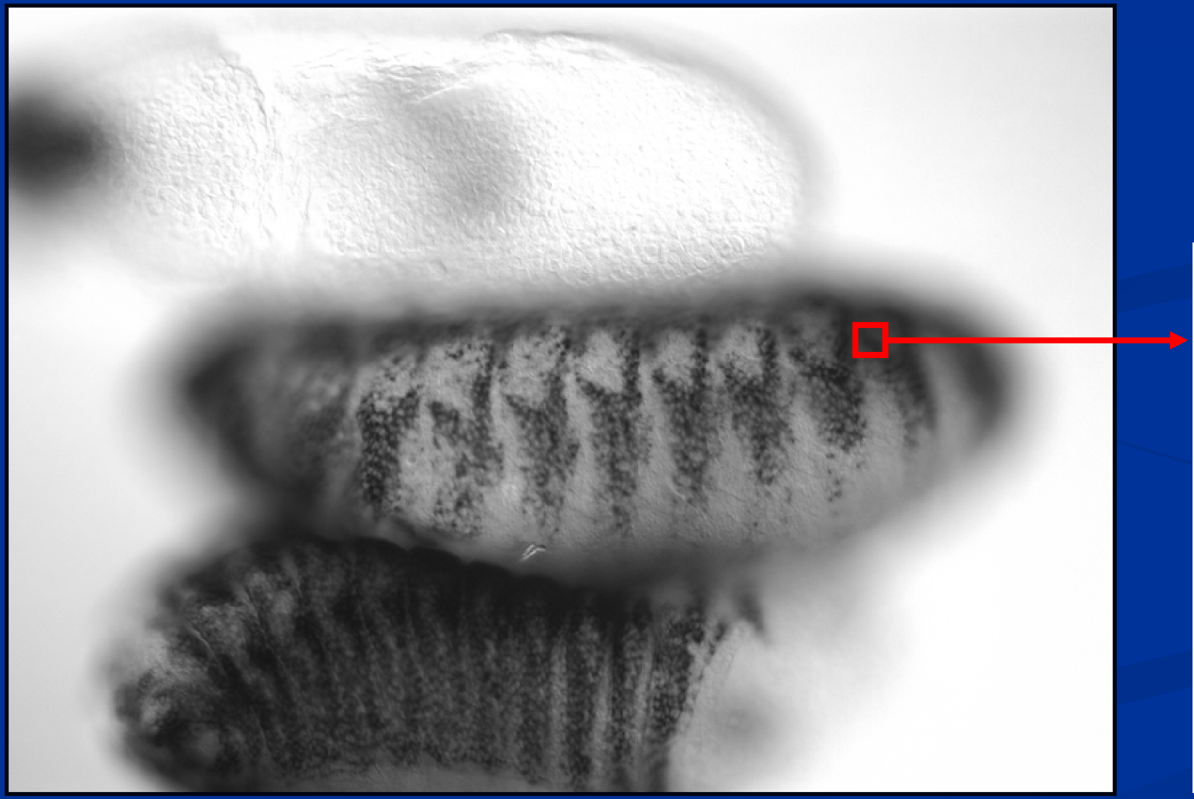
- whole-mount mRNA *in situ* hybridization



blue intensity =
spacial gene transcription level at one particular developmental stage

Image Processing Basics

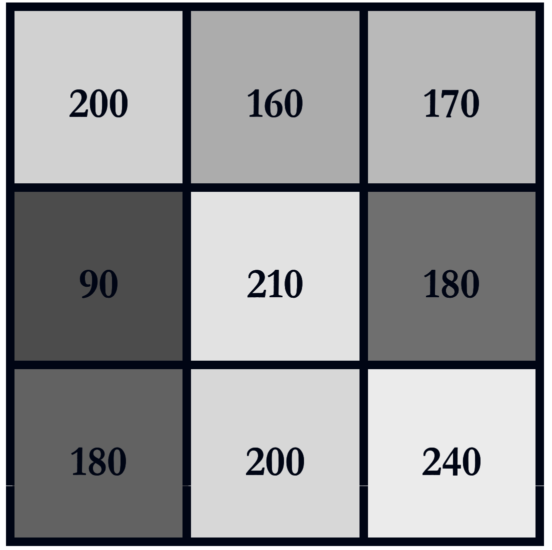
- Digital images are represented on regular grids



Picture x Element

Pixel values: {x}

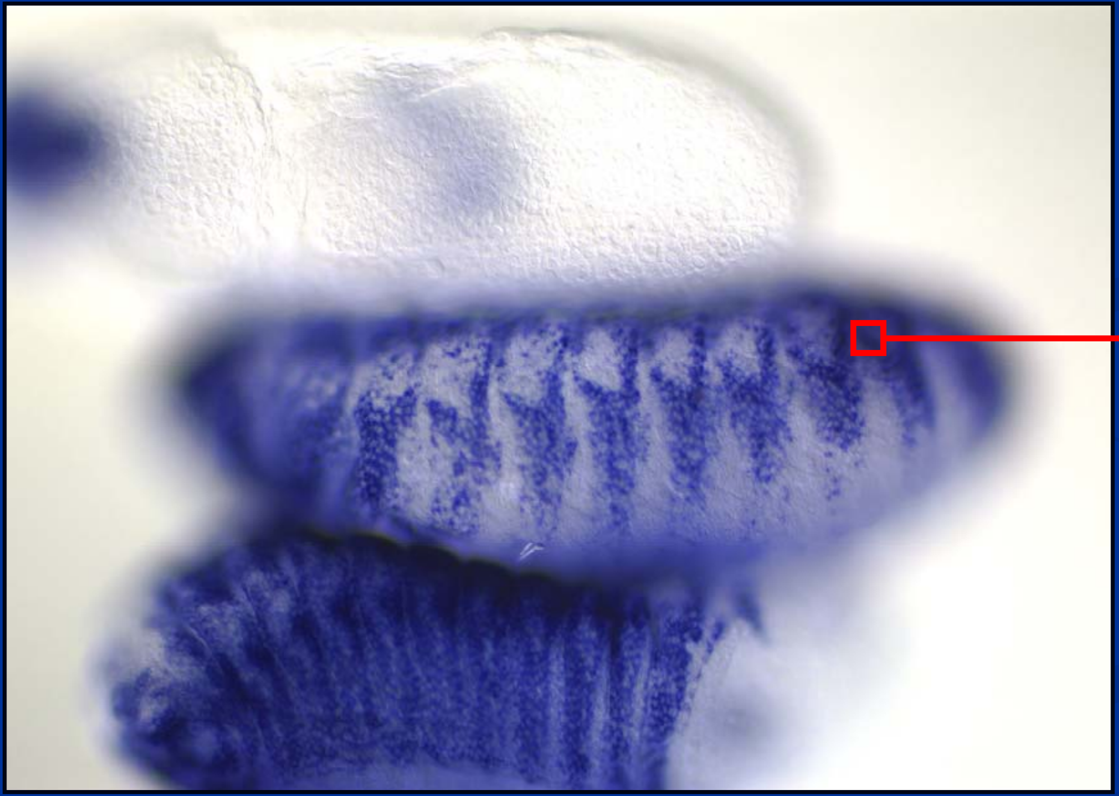
8-bit Range: [0,...,255]



200	160	170
90	210	180
180	200	240

Image Processing Basics

- Digital images are represented on regular grids



Picture x Element

Pixel values: {x,y,z}

8-bit Range: [0,...,255]

200 140 130	160 130 225	160 130 198
145 131 225	130 148 240	160 130 138
227 130 225	214 215 156	160 130 179

Image Processing Basics

- Digital images can be represented with discrete functions

$$f(x, y) = \begin{pmatrix} f(0,0) & f(0,1) & . & f(0,N-1) \\ f(1,0) & f(1,1) & . & f(1,N-1) \\ . & . & . & . \\ f(M-1,0) & f(M-1,1) & . & f(M-1,N-1) \end{pmatrix}$$

- Or with continues functions, after interpolation

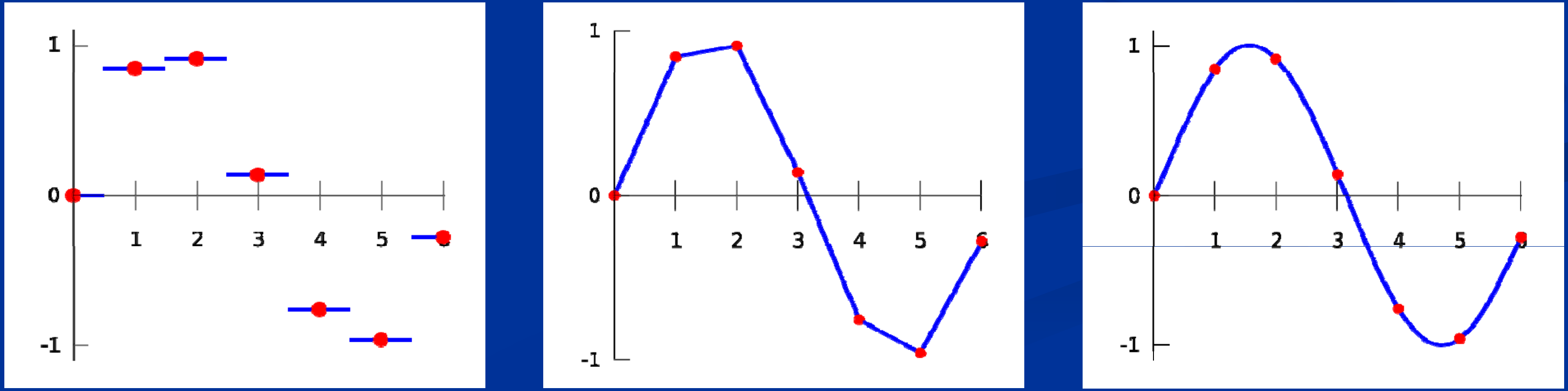


Image Processing Basics

- Various Color Models are available to code a pixel value
- The Gray Scale Model is a 1D List of GS Values

- The RGB Color Space is a 3 dimensional Cube

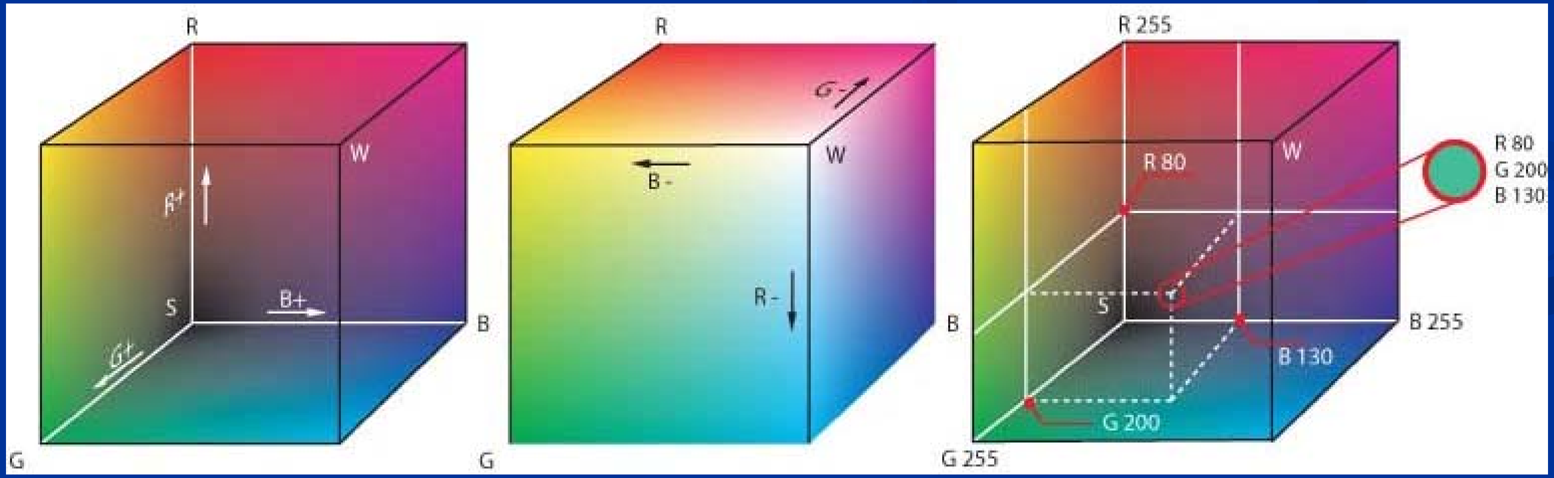
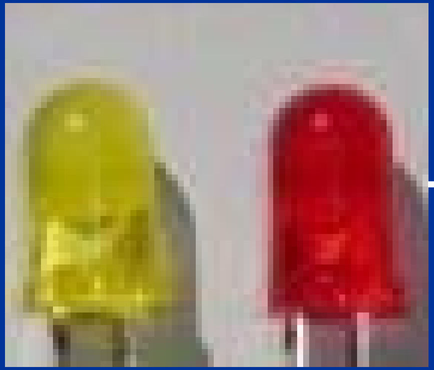


Image Processing Basics

- RGB image mapped to the RGB Color Space



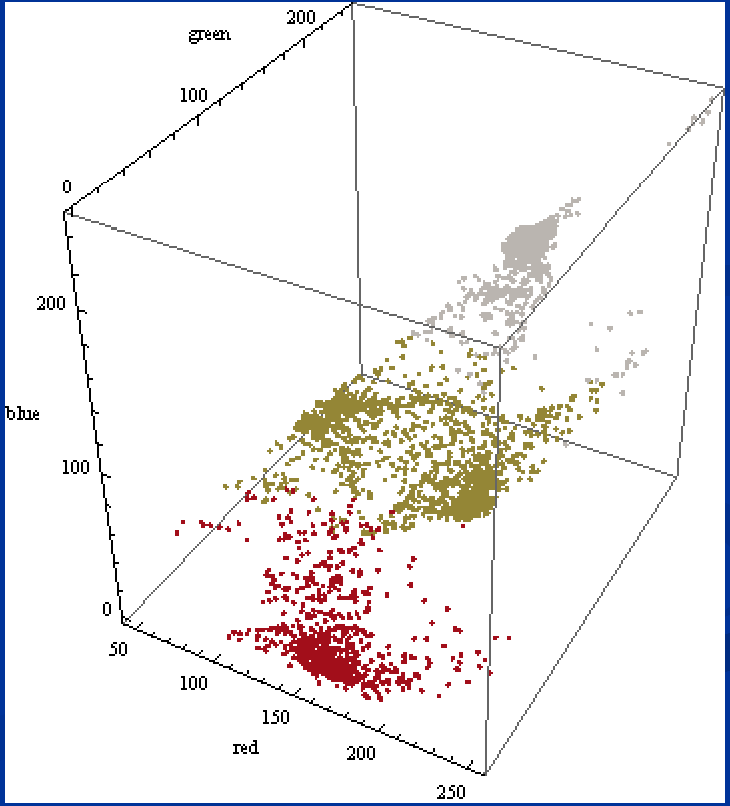


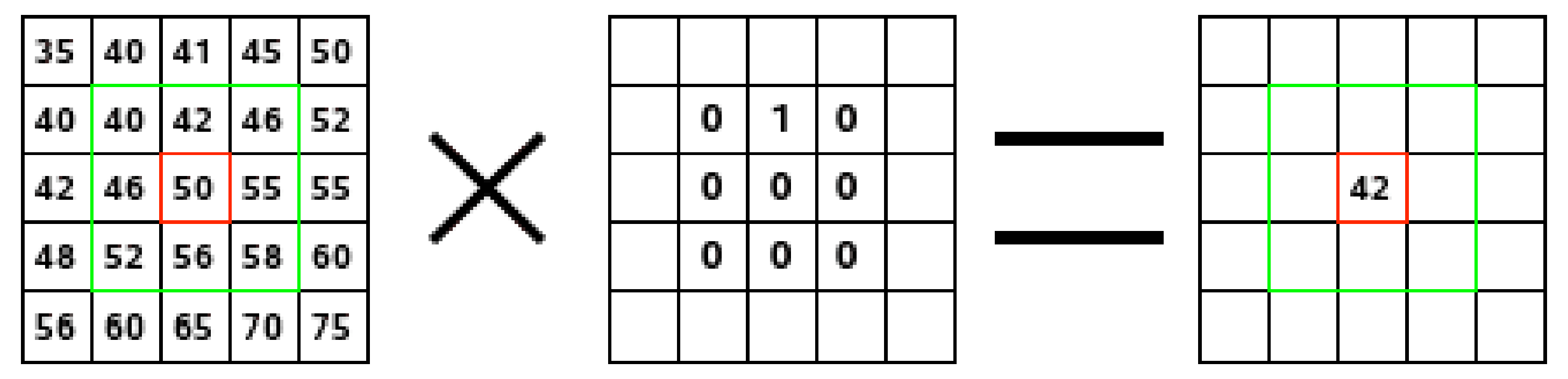
Image Processing Basics

- Image Convolution, Filtering

image matrix

convolution matrix
kernel

convolution result
for one pixel



35	40	41	45	50
40	40	42	46	52
42	46	50	55	55
48	52	56	58	60
56	60	65	70	75

×

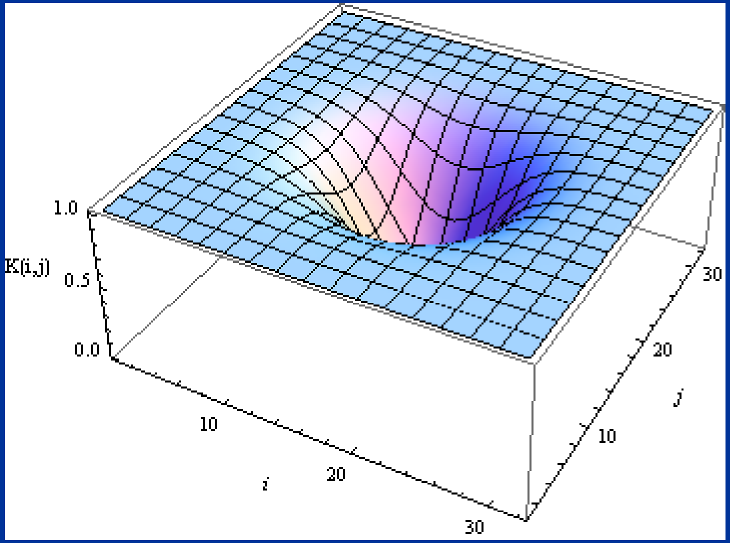
	0	1	0	
	0	0	0	
	0	0	0	

=

		42		

Image Processing Basics

- High-pass and low-pass filter *kernel*



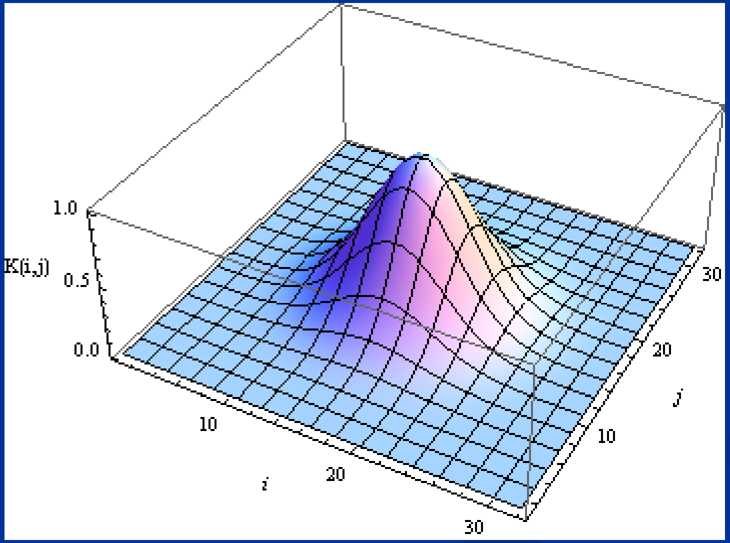


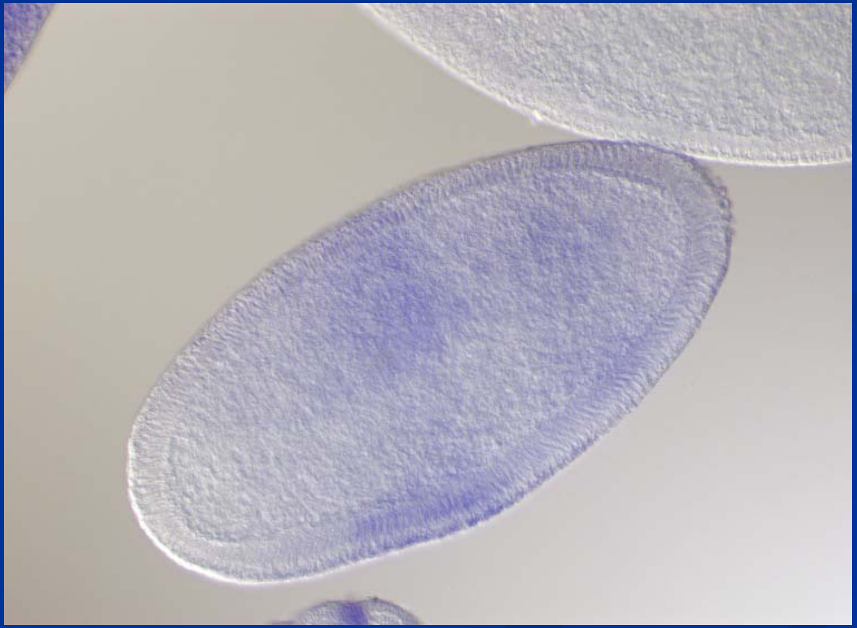
Image Processing Goal

- The creation of a process flow that allows an automatic image classification and analysis of embryonic gene expression patterns from fruit fly ISH experiments

GEP Imaging Complications

background shading

coherent partial embryos



GEP Imaging Complications

poor contrast

blurred contours

Requierements on the Processing pipeline

- Embryo Shape Segmentation
- Allignement of Shapes
- GEP Extraction
- GEP Representation
- Metadata generation
  - -Developmental stage
  - -Orientation
- GEP Clustering

Preprocessing

- Shading correction

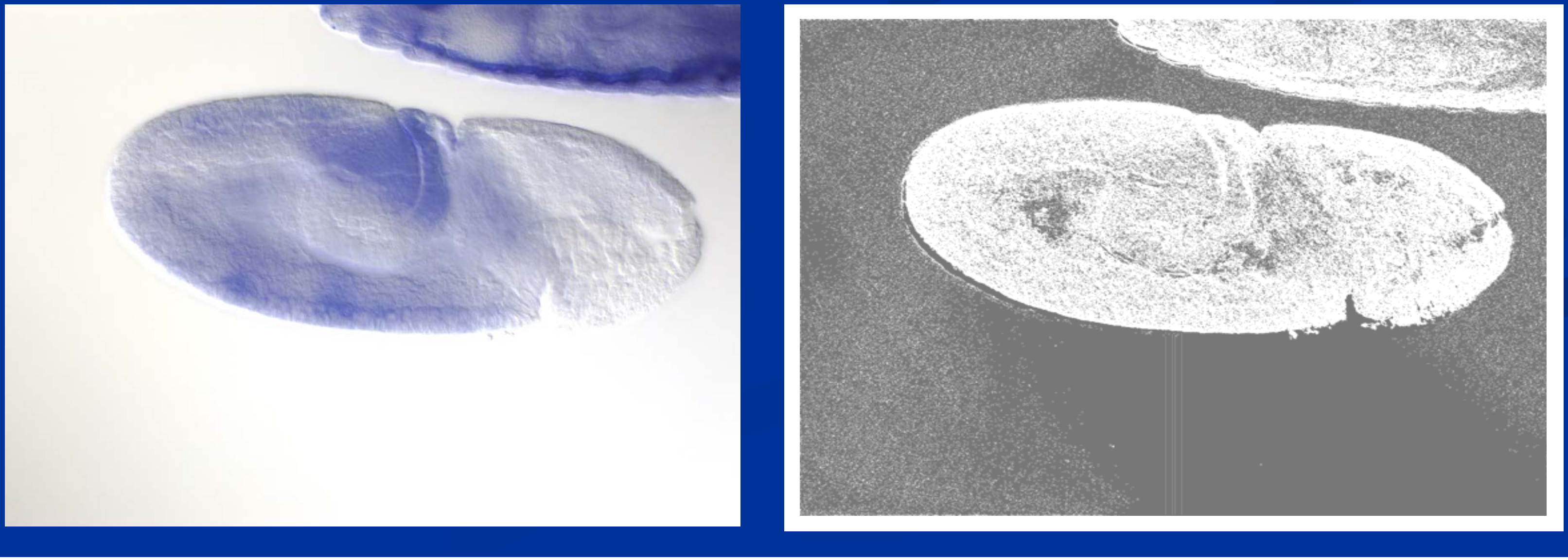
$$S_{cor}(x, y) = S(x,y) - S_{smoth}(x,y)$$

- Contrast optimization

$$P_{cont}(x, y) = (S(x,y) - Min(S) \times \frac{Max(S) - Min(S)}{255}$$

Shape Segmentation

- Feature space:
  gradient magnitude
- Method: Estimating Gaussian Mixture Densities with the Expectation-maximization algorithm



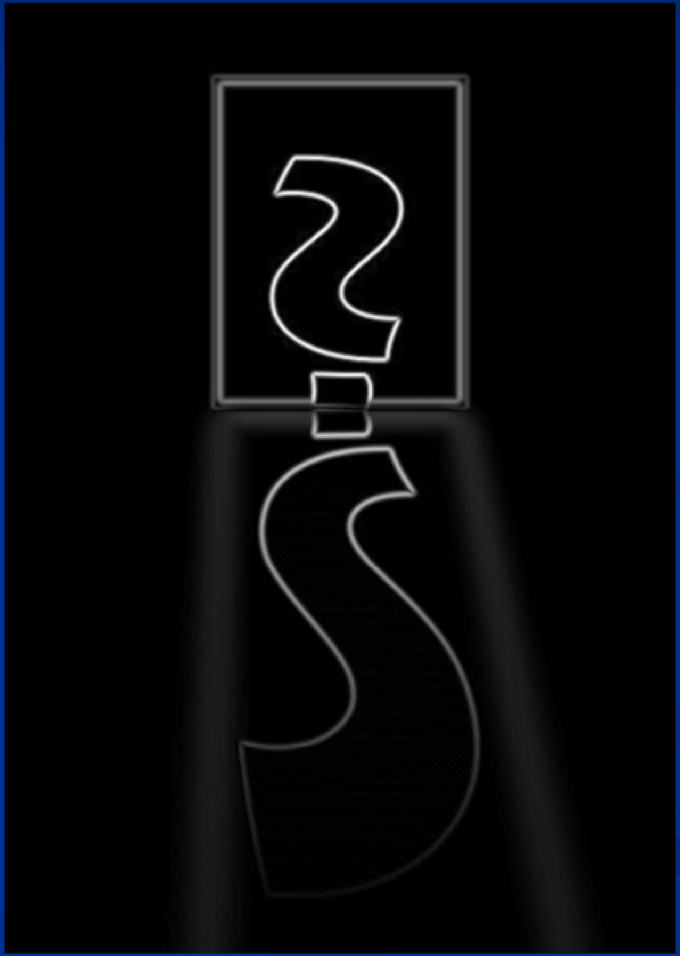
Feature: Mean

Mean

Feature: Standard Deviation

Standard Deviation

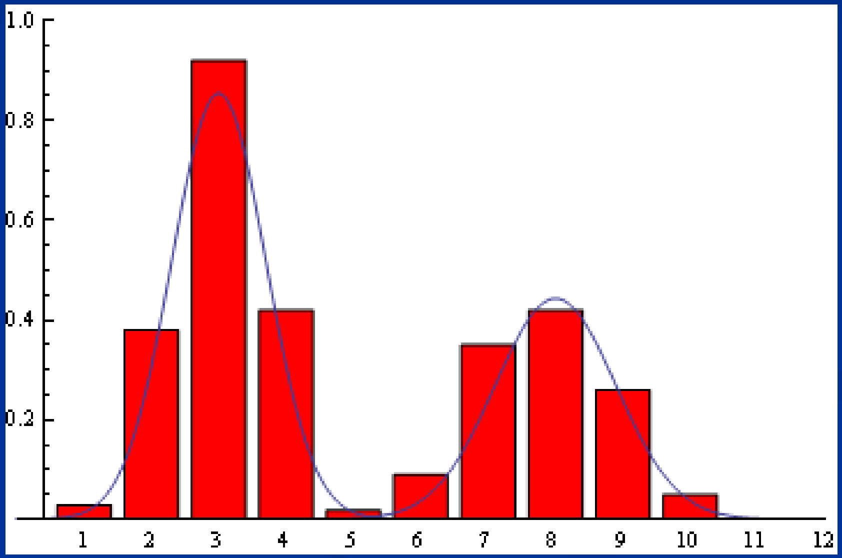
Feature: Gradient Magnitude



Gradient
Magnitude

Gaussian Mixture Model

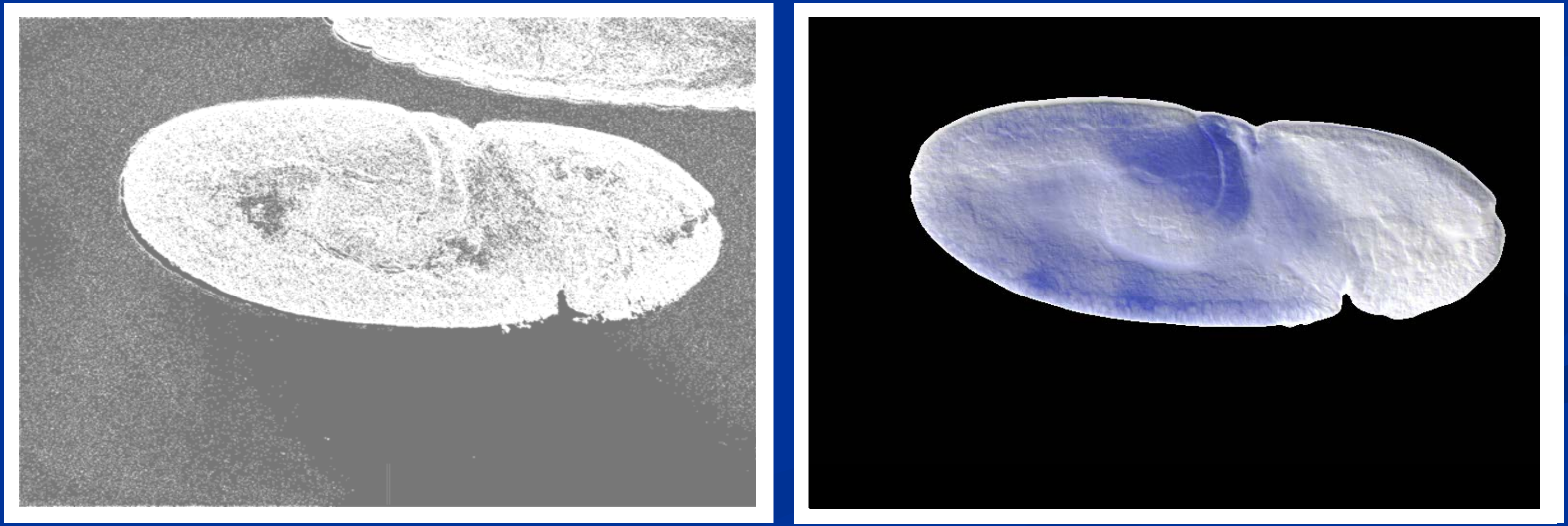
- Gaussian parameters are estimated with the *Expectation Maximasation* Algorithm via *maximum likelihood estimation*



$$P(x) = \sum_{i=1}^{n} a_i G(x;\theta)$$

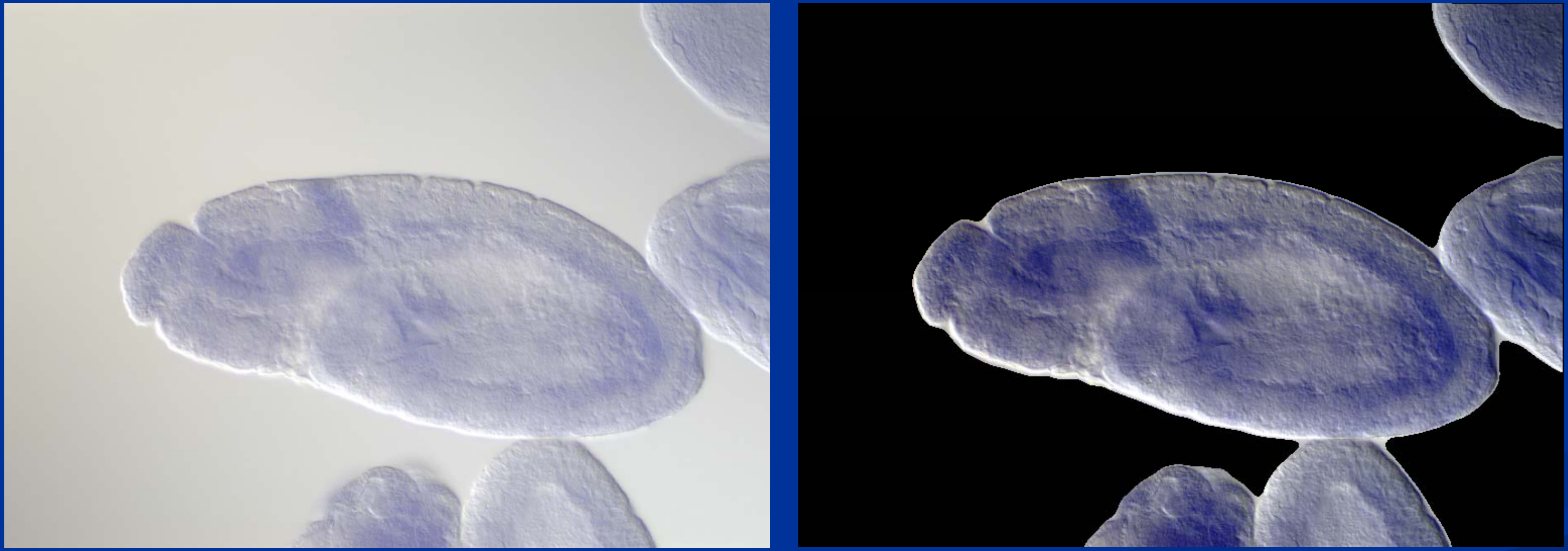
Shape Segmentation

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



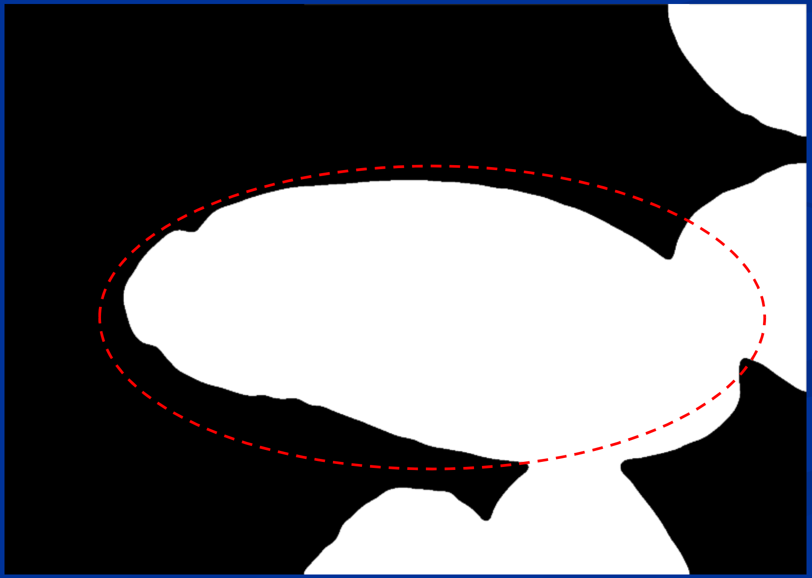
Shape Segmentation

- How to isolate coherent embryos?



Active Contour Approach
GVF Snakes

- Compute Gradient Vector Field
- Define parameterized curve along initial shape
- Define energy cost function along the curve



Shape Segmentation -
Active Contour Approach - Snakes

- Minimizing the energy cost function of the snake:

$$E_{snake} = \alpha \int_a^b E_{int}(v(s))ds + \beta \int_a^b E_{image}(v(s))ds + \gamma \int_a^b E_{con}(v(s))ds$$

with

$$E_{int} = (\alpha_1 |x_s(s)|^2 + \alpha_2 |x_{ss}(s)|^2)/2$$

$$E_{image} = (g(|\nabla I((x(s))|))^2$$

Shape Segmentation -
Active Contour Approach - Snakes

- A snake that minimizes E must satisfy the Euler equation

$$E_{snake} = \alpha_1 x_{ss}(s) - \alpha_2 x^{(iv)}(s) - \nabla E_{image} = 0$$

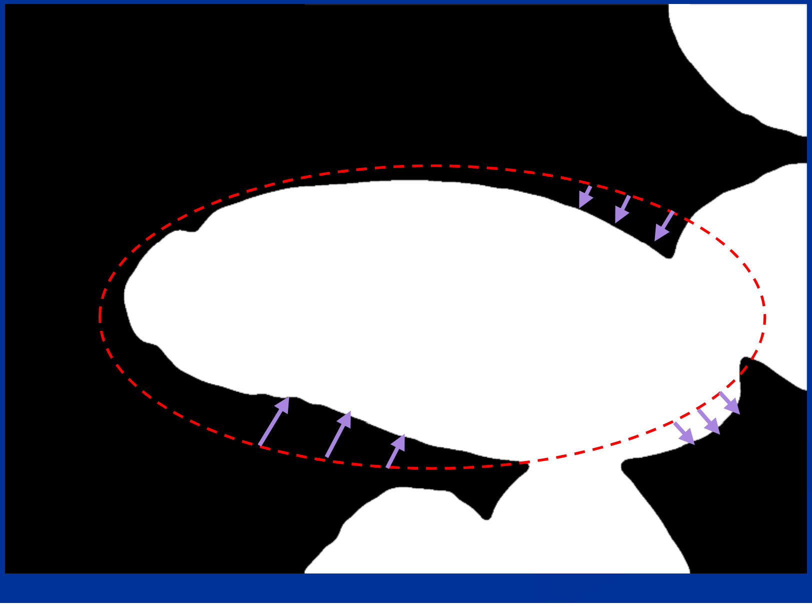
- Interpreted as a force balance problem this would mean

$$F_{int} + F_{image} = 0$$

- The numerical solution procedure is obtained by using a dynamic scheme. For this purpose an artificial time parameter is introduced.

Shape Segmentation - Isolate Coherent Embryos

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

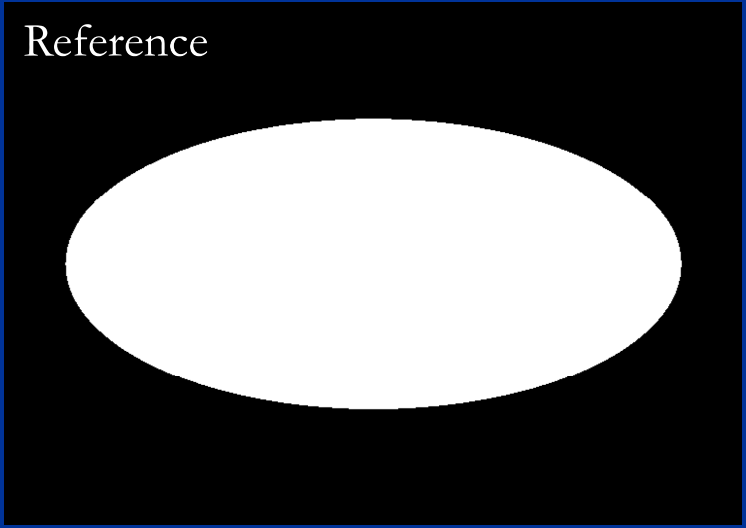


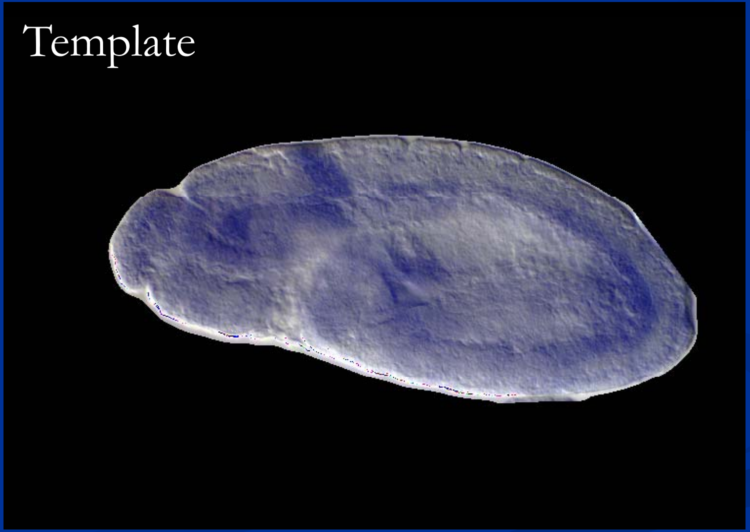
Alligning the Shapes

- Rigid Registration

Distance measure:

$$D(u) = \frac{1}{2}\int_{\Omega}(R(x) - T(x - u(x)))^2 dx$$


Reference


Template

Registration:

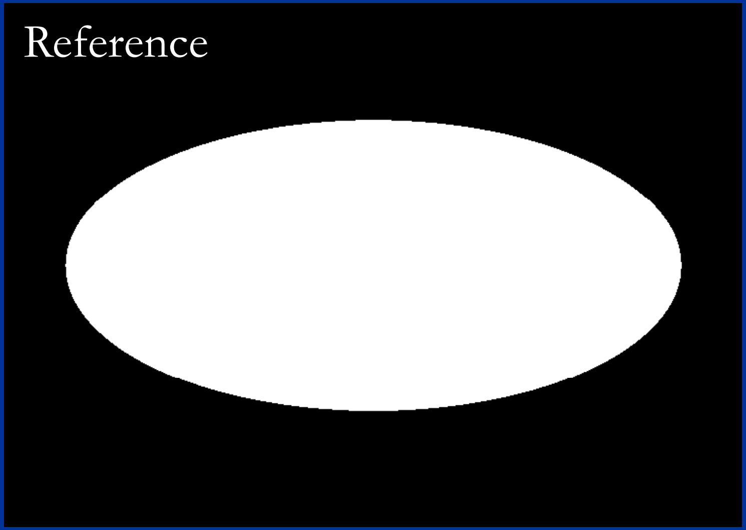
$$min(T(u)) = min(D(u) + \alpha S(u))$$

Alligning the Shapes

- Rigid Registration

Distance measure:

$$D(u) = \frac{1}{2}\int_{\Omega}(R(x) - T(x - u(x)))^2 dx$$



Reference

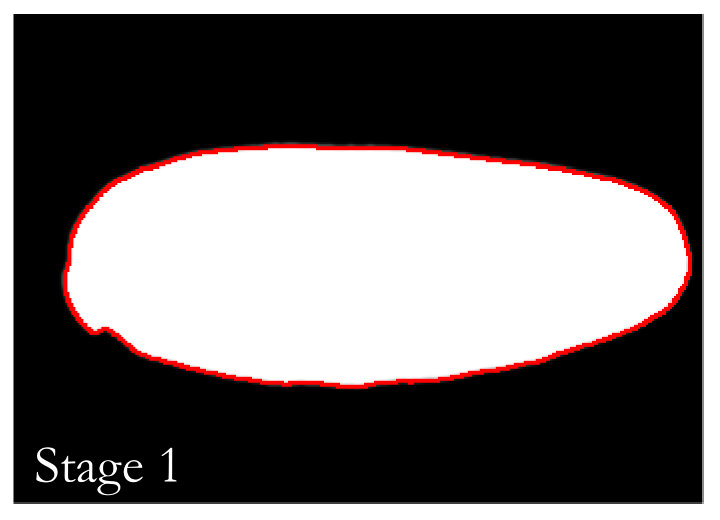
Template

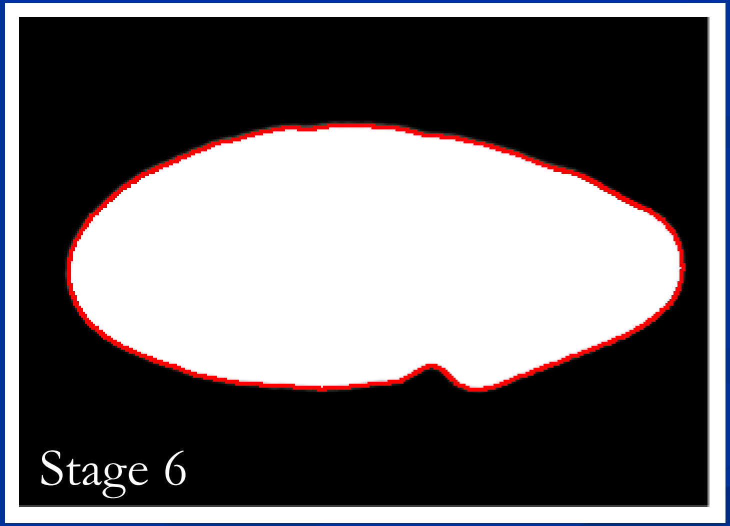
Registration:

$$min(T(u)) = min(D(u) + \alpha S(u))$$

Metadata extraction

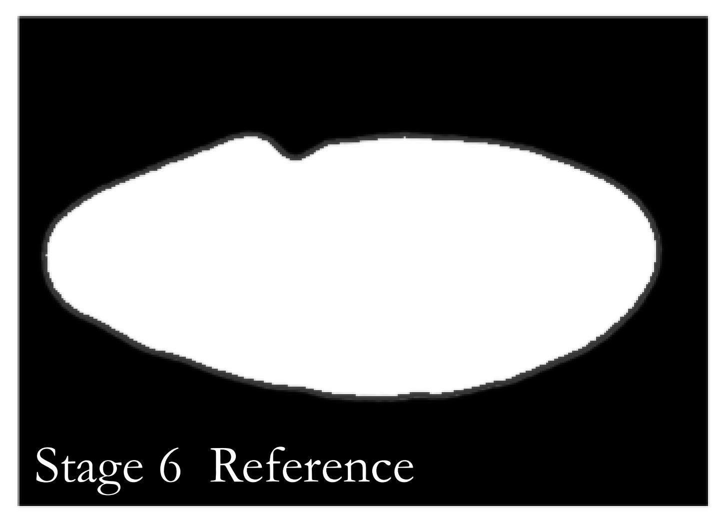
- The outline is used to anotate the stage of the embryo



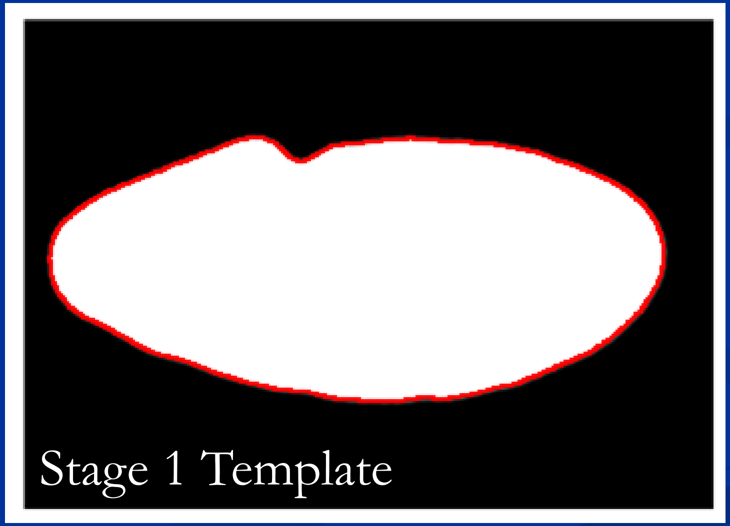


Metadata extraction

- The orientation is corrected by rigid registration to a stage specific standard shape



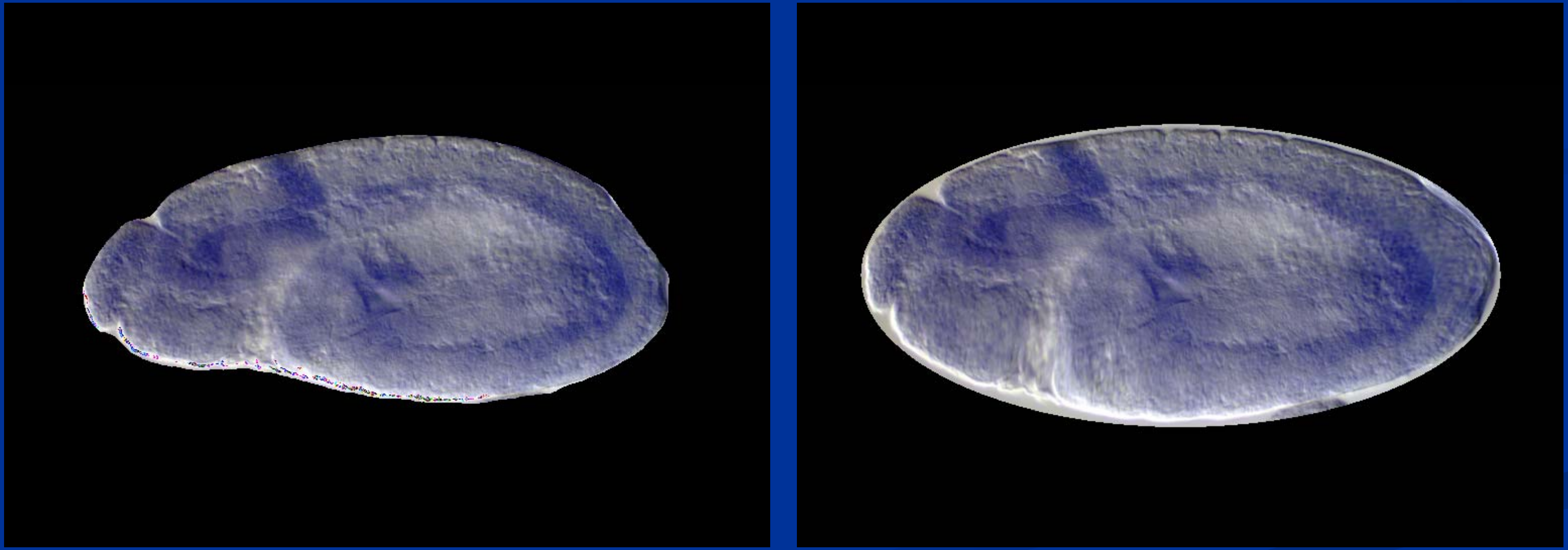
Stage 6 Reference



Stage 1 Template

Transformation of Outline to Ellips

- curvature based Nonlinear Registration



Curvature Based Nonlinear Registration

- Minimal difference between the template and reference shape

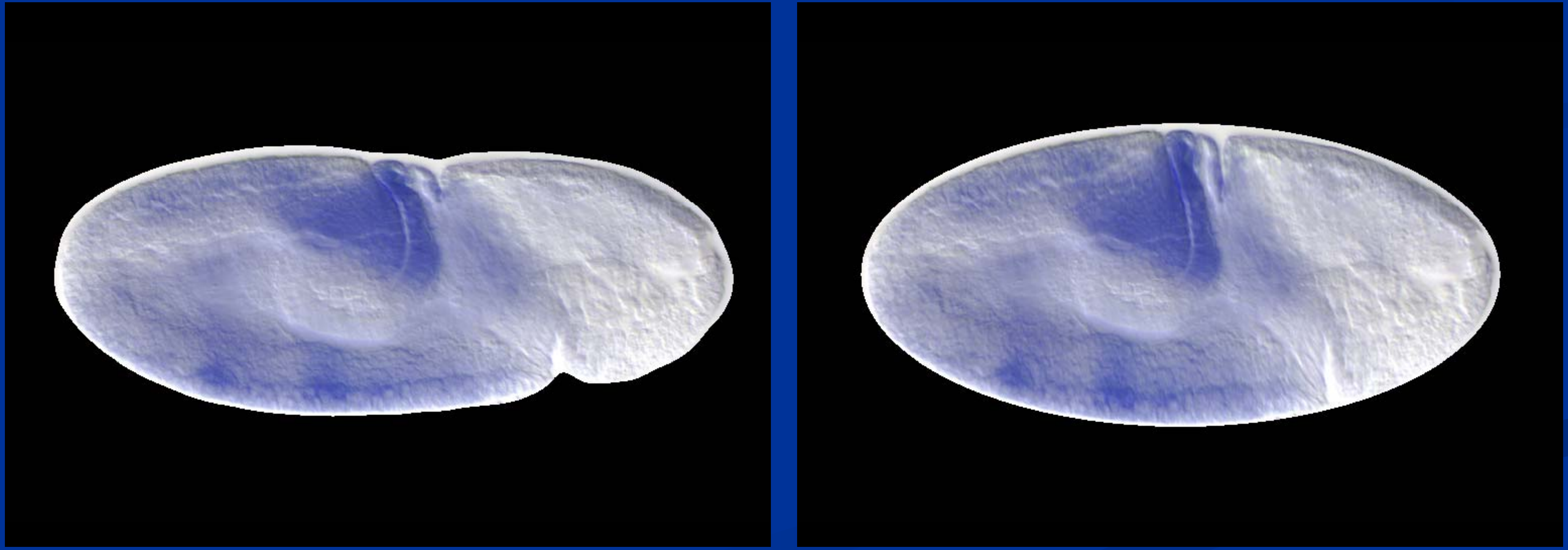
$$minT(u) = min(D(u) + \alpha S^{curv}(u))$$

- With the smoothness term

$$S^{curv} = \frac{1}{2}\int_{\Omega}(\Delta u)^T \cdot (\Delta u)dx$$

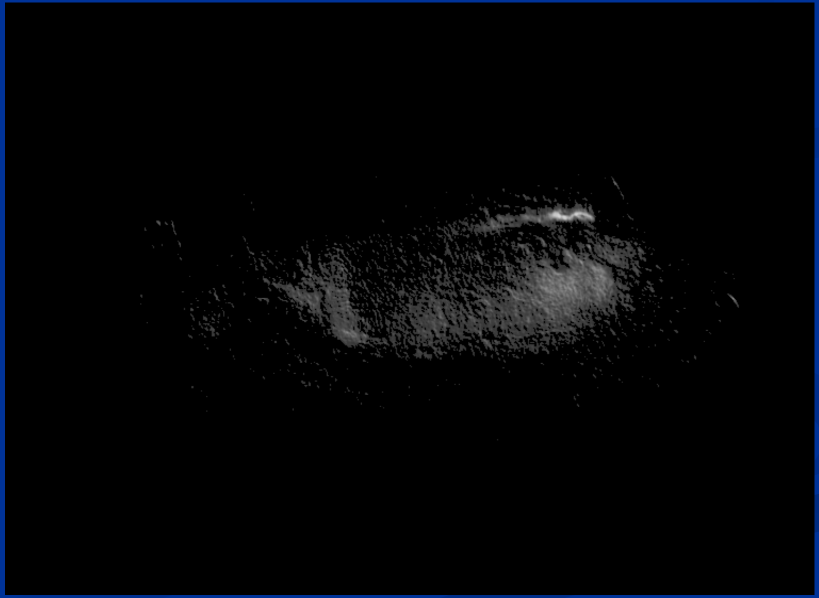
Transformation of Outline to Ellips

- curvature based Nonlinear Registration



Segmentation of the GEP

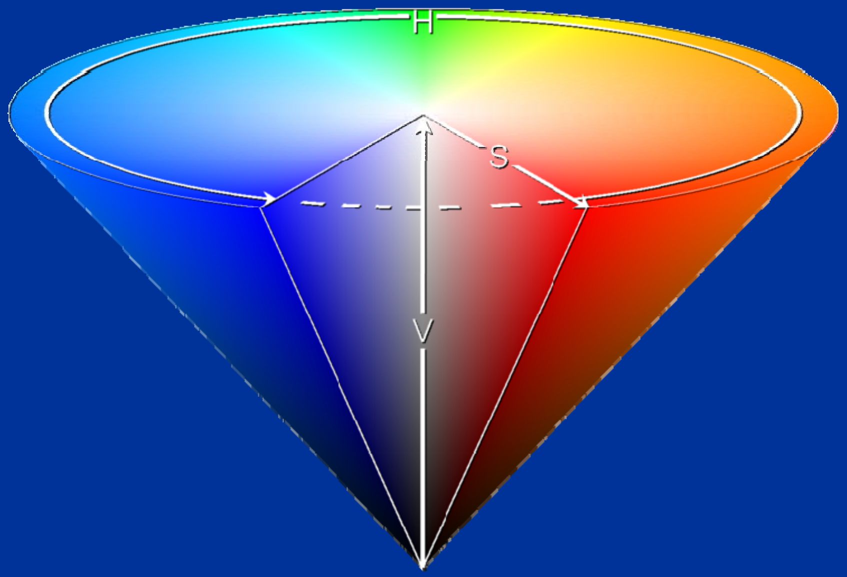
- HSB Colorspace Transformation



S–channel, denoising

The HSB Color Space

- Cone representation of the HSB Color Space



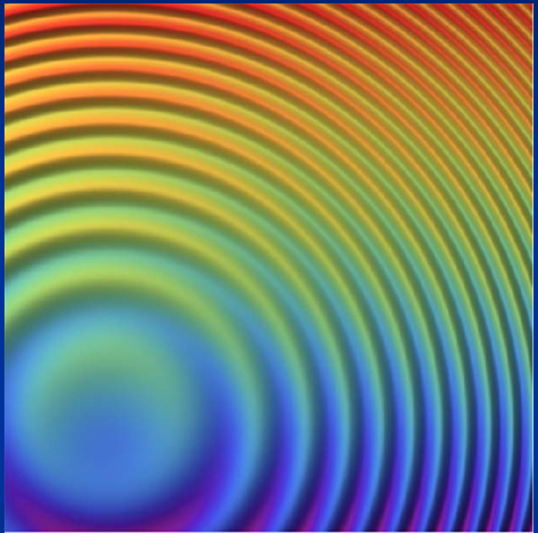
$$V = \frac{(R+G+B)}{3}$$

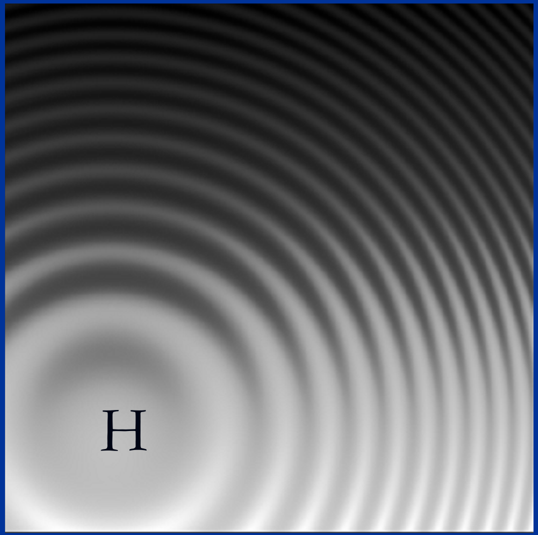
$$S = \begin{cases} \frac{V-\min(R,G,B)}{V}, & V > 0 \\ 0, & V = 0 \end{cases}$$

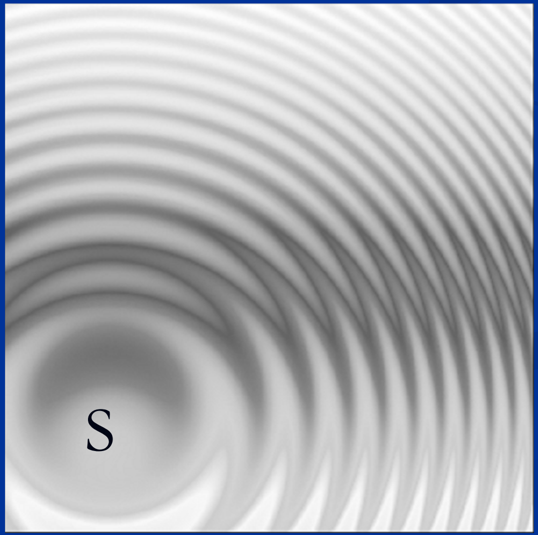
$$H = \begin{cases} 0 + \frac{G-B}{\max - \min}, & R = \max(R,G,B) \\ 2 + \frac{B-R}{\max - \min}, & G = \max(R,G,B) \\ 4 + \frac{R-G}{\max - \min}, & B = \max(R,G,B) \end{cases}$$

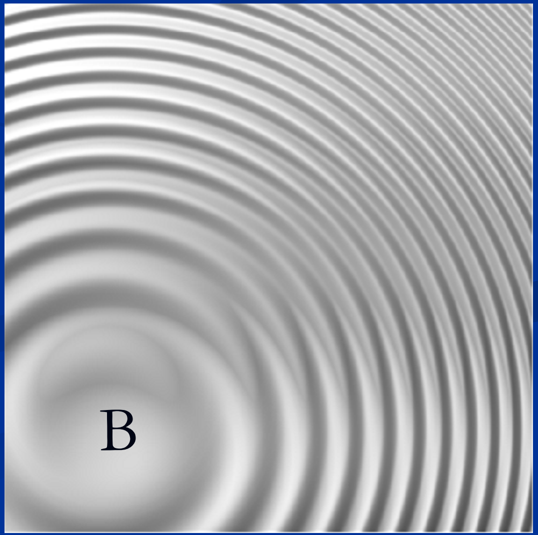
Why HSB Color Space?

- Artificial exaple image transformated to HSB
- H: color Hue (wavelegth)
- S: Color Saturation
- B: Brightness or Illumination









GEP Classification
Fourier Coefficients

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

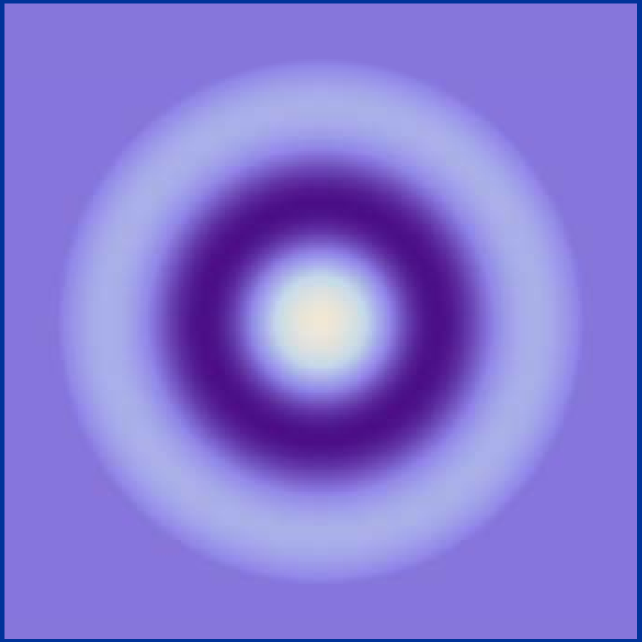
$$\mathcal{P}(r,\phi)=\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}a_{j,k}\,\psi_{j,k}(r,\phi)$$

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

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

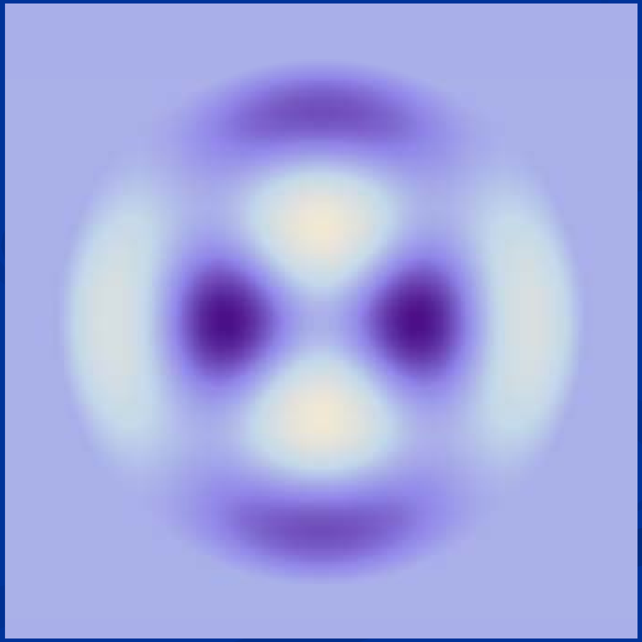
Complete orthonormal system

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



$k=0$

$j=3$

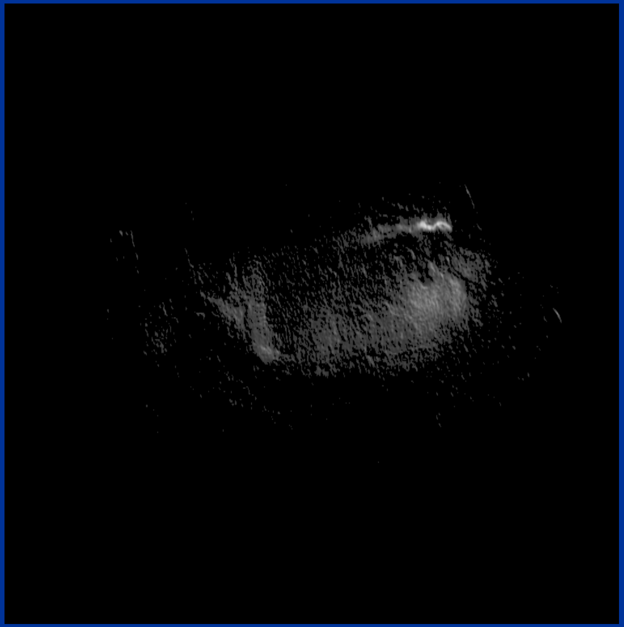


$k=2$

$j=2$

Representation with a set of 420 Fourier coefficients

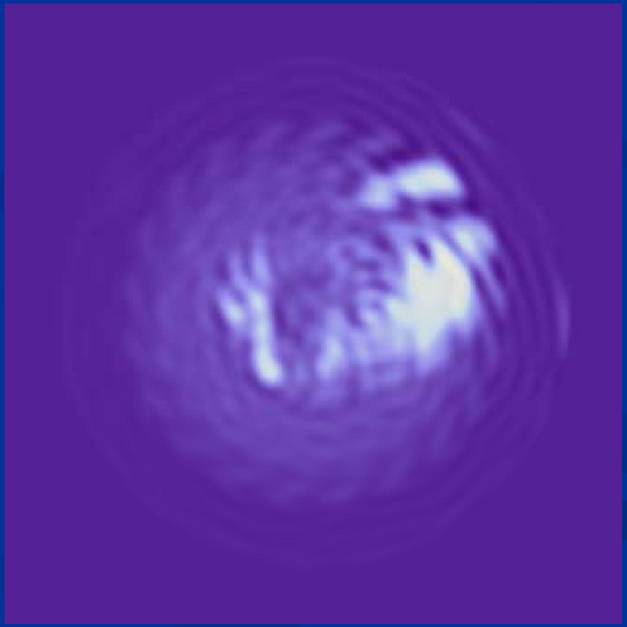
$$a_{j,k} = \int_0^{\ell}\int_0^{2\pi} \psi^*_{j',k'}(r,\phi)\,g(r,\phi)\,r\;\mathrm{d}\phi\mathrm{d}r$$



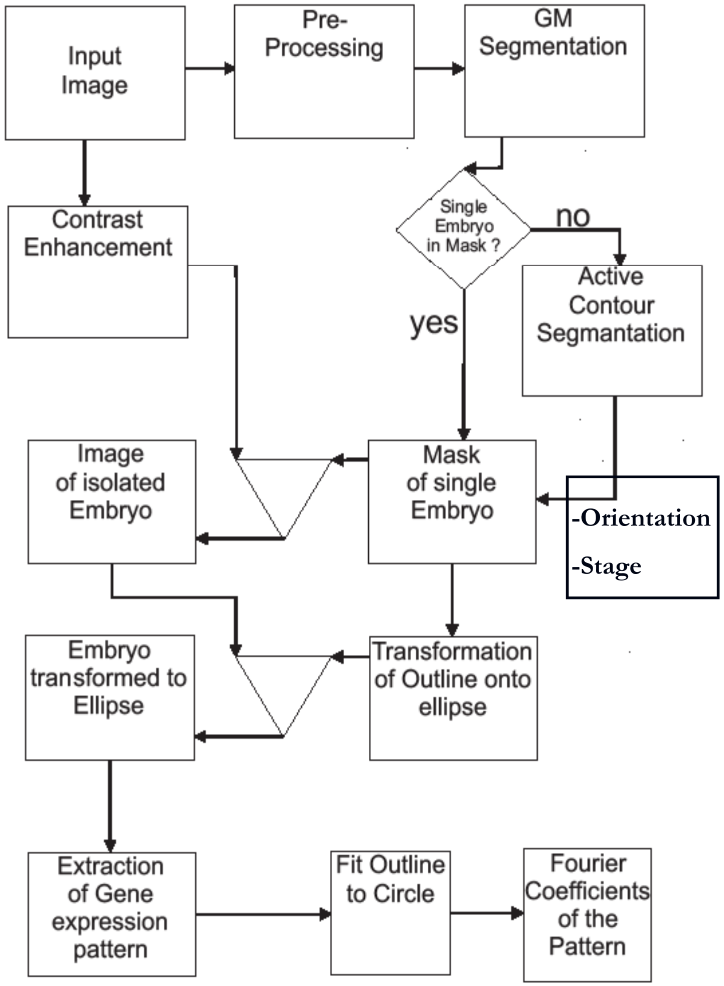
420 Fourier Coefficients

k: [0,...,20]

J: [1,...,20]



Overview Processing Pipeline



Input Image

Pre-Processing

GM Segmentation

Contrast Enhancement

Single Embryo in Mask ?

no

Active Contour Segmantation

yes

Image of isolated Embryo

Mask of single Embryo

-Orientation

-Stage

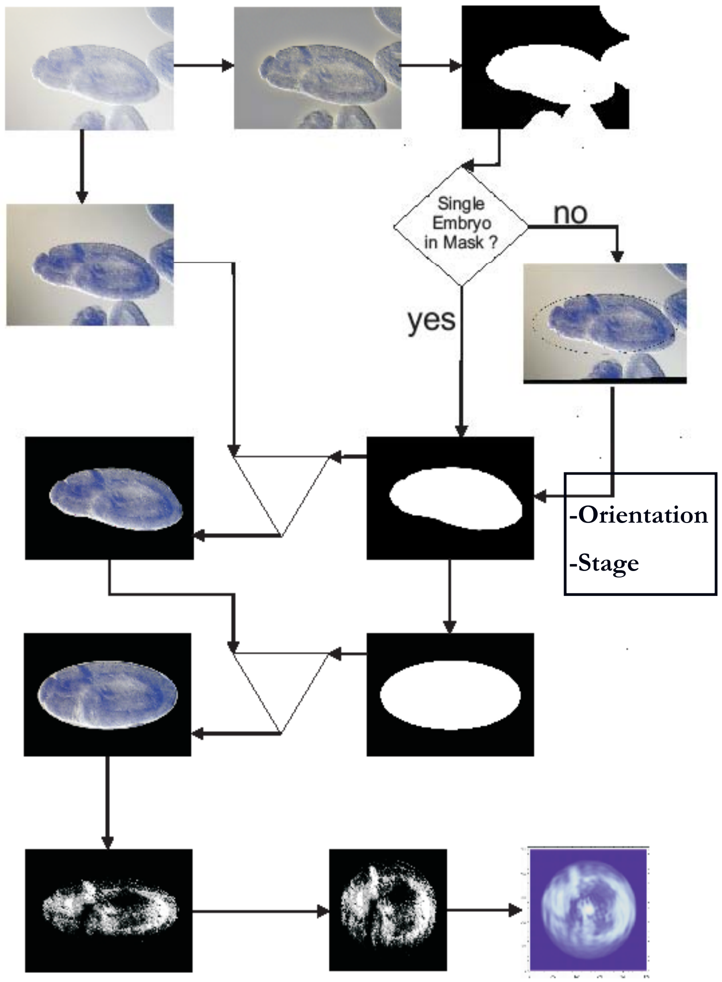
Embryo transformed to Ellipse

Transformation of Outline onto ellipse

Extraction of Gene expression pattern

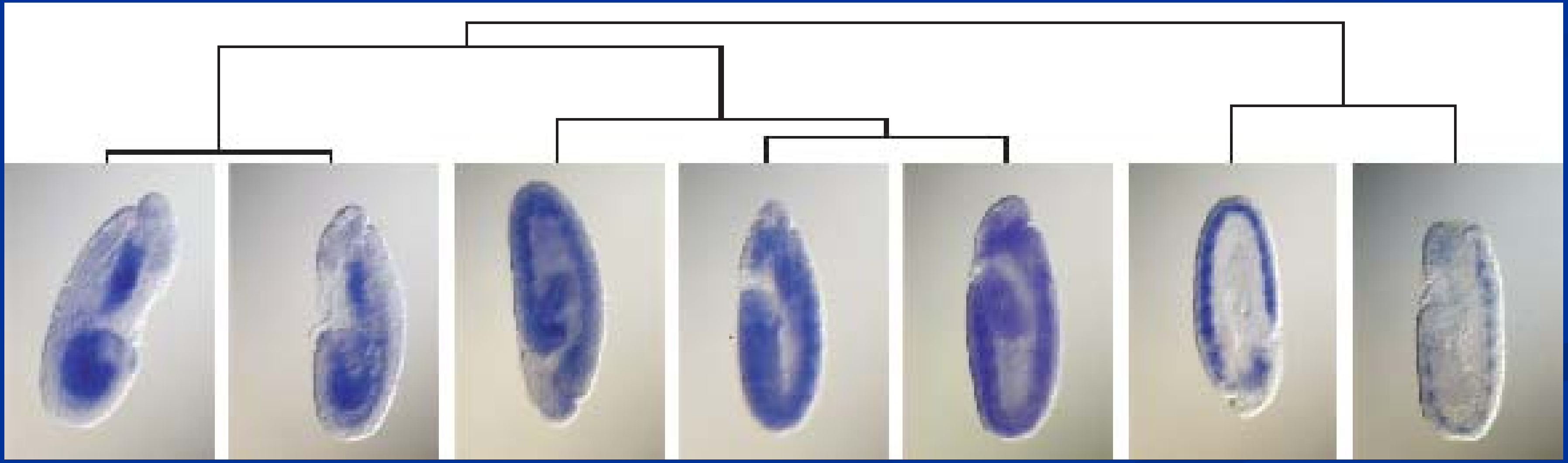
Fit Outline to Circle

Fourier Coefficients of the Pattern

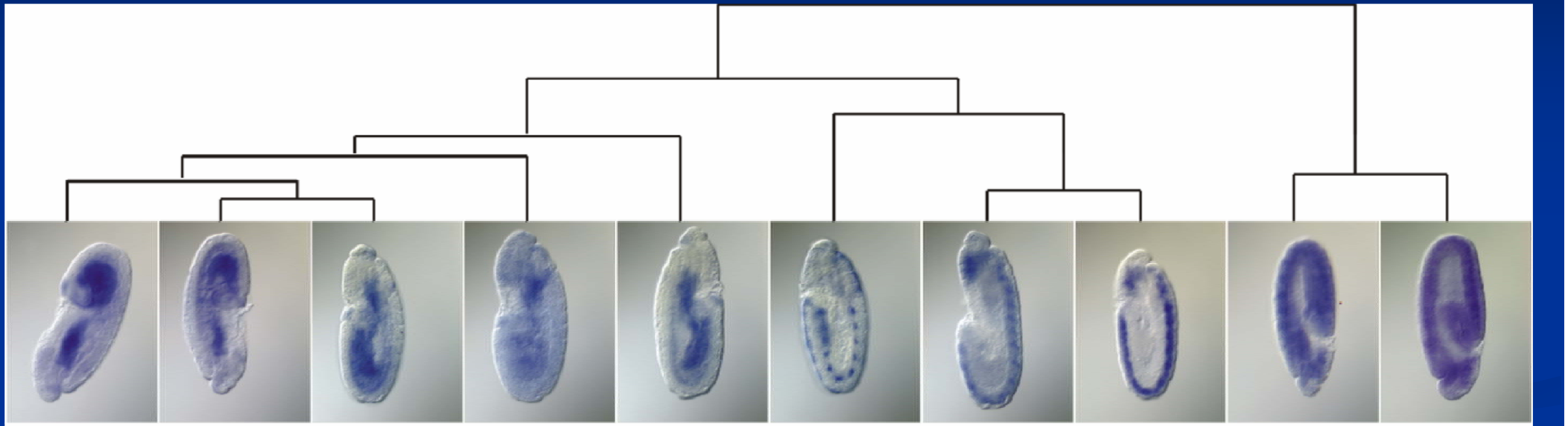


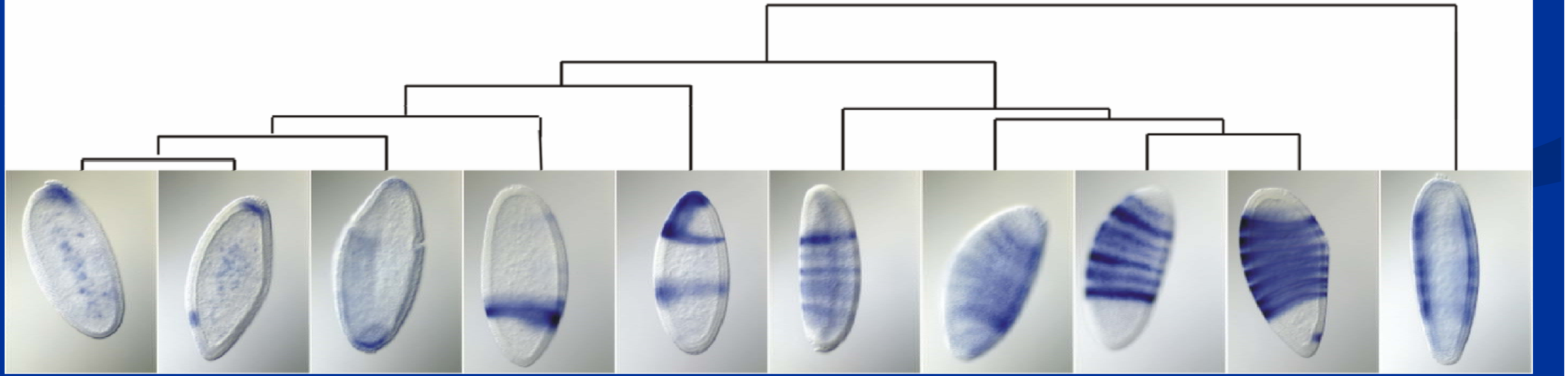
GEP Clustering

- Hierarchical clustering of the absolute values of the coefficient sets using Euclidean norm with an agglomerative algorithm.



Clustering on subsets from dev. stage 4 and 5





That´s it!