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*
Digital images are represented on regular grids.

Pixel values: \{x\}

8-bit Range: [0,...,255]
Digital images are represented on regular grids.

Pixel values: \( \{x, y, z\} \)

8-bit Range: \([0, \ldots, 255]\)
Image Processing Basics

- Digital images can be represented with discrete functions

\[ f(x, y) = \begin{pmatrix} f(0, 0) & f(0, 1) & \cdots & f(0, N - 1) \\ f(1, 0) & f(1, 1) & \cdots & f(1, N - 1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M - 1, 0) & f(M - 1, 1) & \cdots & f(M - 1, N - 1) \end{pmatrix} \]

- 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

<table>
<thead>
<tr>
<th>image matrix</th>
<th>convolution matrix</th>
<th>convolution result for one pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 40 41 45 50</td>
<td>0 1 0</td>
<td>42</td>
</tr>
<tr>
<td>40 40 42 46 52</td>
<td>0 0 0</td>
<td></td>
</tr>
<tr>
<td>42 46 50 55 55</td>
<td>0 0 0</td>
<td></td>
</tr>
<tr>
<td>48 52 56 58 60</td>
<td>0 0 0</td>
<td></td>
</tr>
<tr>
<td>56 60 65 70 75</td>
<td>0 0 0</td>
<td></td>
</tr>
</tbody>
</table>
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
Requirements on the Processing pipeline

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

- **Shading correction**
  \[ S_{cor}(x, y) - S(x, y) - S_{smooth}(x, y) \]

- **Contrast optimization**
  \[ P_{cont}(x, y) = (S(x, y) - \text{Min}(S)) \times \frac{\text{Max}(S) - \text{Min}(S)}{255} \]
Shape Segmentation

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

Mean
Feature: Standard Deviation
Feature: Gradient Magnitude
Gaussian Mixture Model

- Gaussian parameters are estimated with the Expectation Maximisation 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
Minimizing the energy cost function of the snake:

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

with

\[ E_{\text{int}} = (\alpha_1 |x_3(s)|^2 + \alpha_2 |x_{33}(s)|^2)/2 \]

\[ E_{\text{image}} = (g(|\nabla I((x(s))|)|)^2 \]
A snake that minimizes $E$ must satisfy the Euler equation

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

Interpreted as a force balance problem this would mean

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

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:

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

Registration: \( \min(T(u)) = \min(D(u) \mid \alpha S(u)) \)
Alligning the Shapes

- **Rigid Registration**
  
  **Distance measure:**
  \[ D(u) = \frac{1}{2} \int_{\Omega} (R(x) - T(x - u(x)))^2 \, dx \]

  **Registration:**
  \[ \min(T(u)) = \min(D(u) \mid \alpha S(u)) \]
The outline is used to annotate the stage of the embryo

- Stage 1
- Stage 6
The orientation is corrected by rigid registration to a stage specific standard shape.
Transformation of Outline to Ellips

- curvature based Nonlinear Registration
Curvature Based Nonlinear Registration

- Minimal difference between the template and reference shape

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

- With the smoothness term

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

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

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

\[ H = \begin{cases} 0 + \frac{G-B}{\max(G,B) - \min(G,B)}, & R = \max(R,G,B) \\ 2 + \frac{B-R}{\max(B,R) - \min(B,R)}, & G = \max(R,G,B) \\ 4 + \frac{R-G}{\max(R,G) - \min(R,G)}, & B = \max(R,G,B) \end{cases} \]
Why HSB Color Space?

- Artificial example image transformed to HSB
- H: color Hue (wavelength)
- S: Color Saturation
- B: Brightness or Illumination
GEP Classification
Fourier Coefficients

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

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

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

\[ \psi_{j,k}(r, \phi) = N_{j,k} e^{ik\phi} J_k \left( \frac{r j_k}{\ell} \right) \]
Complete orthonormal system

\[ \psi_{j,k}(r, \phi) \equiv N_{j,k} e^{i k \phi} J_k \left( \frac{r j k, j}{\ell} \right) \]
Representation with a set of 420 Fourier coefficients

\[ a_{j,k} = \int_0^\ell \int_0^{2\pi} \psi_{j',k'}^*(r, \phi) g(r, \phi) r \; d\phi \, dr \]

420 Fourier Coefficients

\( k: [0, \ldots, 20] \)

\( J: [1, \ldots, 20] \)
Overview Processing Pipeline

1. **Input Image**
   - Pre-Processing
   - GM Segmentation

2. **Orientation**
   - Single Embryo in Mask?
     - no
     - Active Contour Segmentation
     - yes
     - Mask of single Embryo

3. **Image of isolated Embryo**
   - Embryo transformed to Ellipse
   - Transformation of Outline onto ellipse
   - Extraction of Gene expression pattern
   - Fit Outline to Circle
   - Fourier Coefficients of the Pattern

4. **Stage**
   - Orientation - Stage
   - Single Embryo in Mask?
     - no
     - yes

5. **Final Output**
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!