Detection and localization of Drosophila egg chambers in microscopy images

Jiri Borovec & Jan Kybic & Rodrigo Nava

Biomedical Imaging Algorithms Group, CMP, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

jiri.borovec@fel.cvut.cz

Abstract
Drosophila melanogaster is a well-known model organism that can be used for studying oogenesis (egg chamber development) including gene expression patterns. Standard analysis methods require manual segmentation of individual egg chambers, which is a difficult and time-consuming task. We present an image-processing pipeline to detect and localize Drosophila egg chambers that consists of the following steps:
1. superpixel-based image segmentation into relevant tissue classes;
2. detection of egg center candidates using label histograms and ray features;
3. clustering of center candidates and;
4. area-based max. likelihood ellipse model fitting.
Our proposal is able to detect 96% of human-expert annotated egg chambers at relevant developmental stages with less than 1% false positive rate and improves the mean adjusted Rand score (ARS) from 0.75 using common watershed technique to 0.86, which is adequate for the further analysis.

Superpixel segmentation
We use superpixel segmentation proposed in [1]. First, BRLEC superpixels are calculated [3] with an initial size of 15 pixels. For each superpixel, color and texture features are computed. Then, the superpixels are assigned to one of four classes (background, follicle cells, cytoplasm, or nurse cells) using a random forest classifier with GraphCut [3] regularization.

Center features
Label histograms. Around a given point, a set of 9 circular regions $R_i$ is defined. For each region, a normalized label (class) pixel histogram within each region is computed.

Ray features, [4]. For each ray $r_i$ we measure the distance $r_i$ to the first background-class point in the given direction. To obtain rotational invariance, the vector is circularly shifted to start with the largest element.

Center classification
The superpixel classifier transmits a tentative central example in green and negative for any examples in red, ignoring the intermediate zone in yellow.

Detection results
Egg detection performance of the detection task by development stages, in terms of false positives, false negatives, and the number of multiple detected eggs before and after post-processing with ellipse fitting.

Ellipse fitting
Maximize the likelihood ($g$ is the entire image, $Y_i$ the ellipse interior)
$$
\max_{\theta, \gamma, y} \prod_i P_Y(Y_i) \prod_i P_B(Y_i),$$

Where $P_Y(Y_i)$ and $P_B(Y_i)$ are image-based foreground/background probabilities for pixel $i$. Taking negative log likelihood $\lambda_i = -\log P_i$ we get
$$
\min \sum_{\theta, \gamma, y} g(Y_i) - \lambda(Y_i),
$$

To obtain a robust fit, ellipses are fitted [6] RANSAC-like to randomly selected subsets of 40% of detected boundary points for each center.

Reference models and post-processing
A grid of models is used, and the top model is selected as the final output. Then, the models are post-processed using a graph-cut approach [3].

Conclusion
We presented a complete pipeline for Drosophila oocyte chamber detection and localization by ellipse-fitting in microscopic images. Our contributions include novel label histogram features, the rotation invariant ray features, and area-based maximum likelihood ellipse fitting. The performance is completely adequate for the desired application – it is important that the number of false positives is small but false negatives are not a problem, as long as a sufficiently high number of egg chambers is detected.

References

Experimental results
Input images (Top), initial segmentation (middle) followed by the detected centers (cluster means) as dots and the fitted ellipses in green (bottom). Expert drawn bounding boxes are shown as red rectangles (not all eggs are annotated). Further examples below.

Center clustering
Detected corresponding to individual eggs are grouped together using density-based spatial clustering (DISCAN) [5]. The distance threshold of DISCAN is set to $3\times$ the superpixel size.