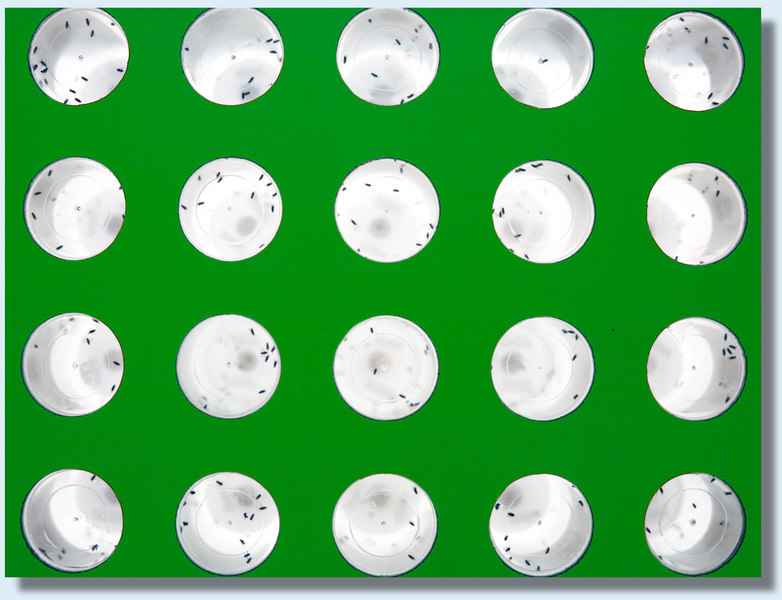
Automatic Object Detection Using DBSCAN for Counting Intoxicated Flies in the FLORIDA Assay

C. Bodenstein*, A. Jansen**, M. Goetz*, H. Scholz**, M. Riedel*

* Juelich Supercomputing Center, Leo-Brandt-Strasse, 52425 Juelich, Germany

** University of Cologne, Department of Biology, Zuelpicher Strasse 47b, 50674 Cologne, Germany



An instrumentation and computer vision pipeline allowing automatic object detection using DBSCAN

- Clustering can be highly efficient for image segmentation
- Example: Counting intoxicated flies on experimentation vial bottom to measure tolerance towards alcohol
- Construction of an autonomous experimentation pipeline consisting out of algorithm, hardware and software
- Automation of the process facilitates reproducibility and consistency, while decreasing manual labor

The FLORIDA Assay

- "**F**ull **L**oss **O**f **R**ighting Reflex **I**n**D**uced by **A**lcohol"
- Aim is to identify genes underlying drug tolerance
- Tested by loss of righting reflex—ability to stand up after falling down and subsequent mechanic stimulation—under alcohol influence
- Ethanol intoxication influences the loss of righting reflex in humans and flies similarly
- The vinegar fly—*drosophila melanogaster*—used as genetic model to analyze underlying behaviors associated to alcoholism [3]
- Exposure of flies to vaporized ethanol in vials
- Counting sedated flies as a measure of intoxication
- Classification of a fly as sedated by its inability to right itself after shaking the vial
- Generation of statistically significant results through multiple repetitions of the experiment

Hardware Setup

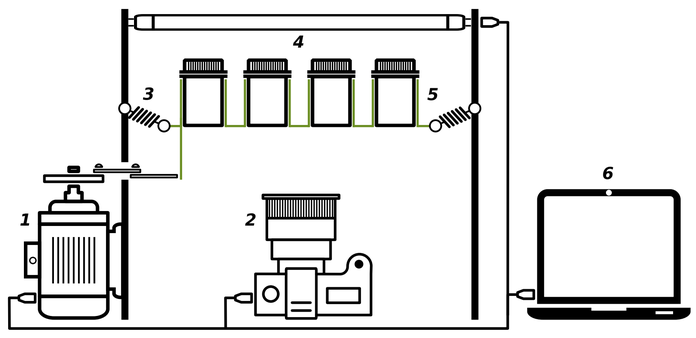


Figure: The automated experimentation setup. It consists of of (1) an electric engine, shaking the vials via a rotary crank, (2) an SLR camera, (3) a spring suspension allowing horizontal motion, (4) a lightning top plate, (5) the experimentation vials containing the flies and (6) a notebook with the analysis software, connected via USB to other devices.

- A single-lens reflex camera (SLR) is used to capture images from multiple experimentation vials
- Electric engine shaking the vials to test loss of righting reflex
- Computer with FLORIDA software controlling the hardware

DBSCAN

- **D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise
- Unsupervised clustering algorithm that is able to find arbitrary shaped clusters
- ϵ = neighborhood search radius
- θ = density threshold
- Recursively expand cluster as long θ points in ϵ
- Utilized parallel HPDBSCAN implementation [1]

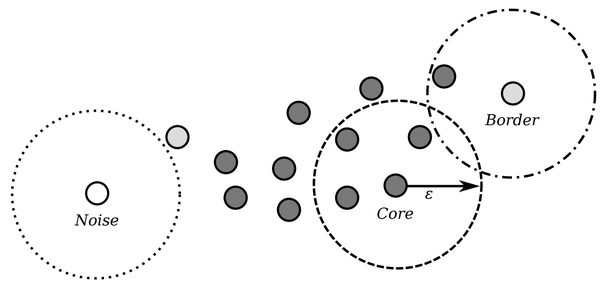


Figure: *DBSCAN* clustering with *minPoints* $\theta = 4$ and search radius ε.

Software

The Algorithm

- Vial segmentation using green alpha key
- Threshold τ binary image to segment flies
- DBSCAN clustering to count flies, each cluster represents one fly
- Split overly large clusters—i.e. dangling flies—by average fly size parameter ρ

Standalone GUI application

- Control and access camera image
- Visualize and supervise vial segmentation, threshold image, clustering
- Adjust and store algorithm parameters
- Run and record experiments

Model Optimization and Results

- Two labeled data sets:
  - 1381 vials in low quality images
  - 880 vials in high quality images
- Evaluation through minimization of the number of miscounted flies
- Optimization of the four model parameters ϵ, θ, τ, and ρ
- Performed parameter grid search on the JURECA HPC system [2]
- Total of 39,900 combinations
- Final results of the evaluation *MSE*: 1.745 and R^2: 0.946
- 95% of the image samples have less than three miscounts

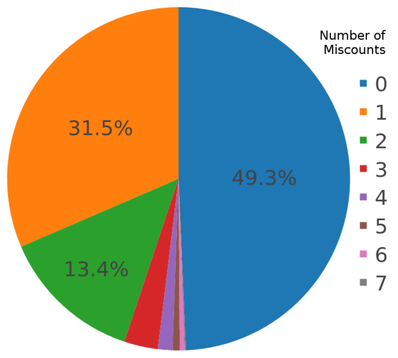


Figure: Absolute miscount distribution of our model using the best grid search parameters

Future Work

Improvements strategies to construct a more robust model:

- Train a convolutional neural network (CNN) on threshold images to determine a better threshold τ
- Dynamically find optimal parameter for each vial picture utilizing a CNN
- Image classification on found clusters to discover multiple flies in one cluster
- Plain neural network prediction:
  - input → vial image
  - output: → fly count

This work will be used in biological experiments at the University of Cologne

References

[1] Markus Götz, Christian Bodenstein, and Morris Riedel. Hpdbscan: highly parallel dbscan. In *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, page 2. ACM, 2015. [2] Dorian Krause and Philipp Thrnig. JURECA: General-purpose supercomputer at Jlich Supercomputing Centre. *Journal of large-scale research facilities*, 2:A62, 2016. [3] Scholz and Mustard. Invertebrate models of alcoholism. In *Behavioral neurobiology of alcohol addiction*, pages 433–457. Springer, 2011.

Contact: c.bodenstein@fz-juelich.de **- Website:** www.fz-juelich.de