Autonomous Anomaly Detection of Orchard Tree Crown Delineations

Author: Edward Baleni *Supervisors*: A/P Patrick Marais, Dr. Sebnem Er

University of Cape Town, Cape Town, South Africa

Highlights

This study focuses on identifying errors in estimated orchard tree crown boundaries. Image analysis techniques capture an appropriate feature set. Unsupervised anomaly detection discriminates outliers within the proposed dataset. By the end of this study, we look to present a set of salient features and a detection algorithm for the automatic and generalizable detection of anomalies in individual orchard tree crown boundaries.

Problem Statement

Anomaly detection discriminates rare observations from a dataset. The main aim of this study is to develop an algorithm that detects anomalies in orchards to focus **phenotyping** efforts. Relevant data sources are provided by Aerobotics, a company striving to preserve the future of sustainable farming through precision agriculture. Using data like that provided, researchers have previously estimated boundaries (**delineations**) around individual orchard tree crowns using computer vision.

However, correct delineations are often accompanied by anomalies, such as:
A) **false positives** (canopy floor mistaken as crowns),
B) **over-segmentation** (one crown split into many),
C) **under-segmentation** (multiple crowns merged into one).

Identifying anomalies can be supervised, semi-supervised, or unsupervised. Due to unreliable reference data, this study uses an **unsupervised approach**.
Along with a best performing algorithm, we hope to:

- develop a salient feature set that clearly distinguishes outliers,
- and ensure that the feature set and algorithm are generalisable and automatic when exposed to new orchards.

This study uses 11 baseline algorithms and proposes the Multivariate Geary's C Statistic[1] to identify outliers.

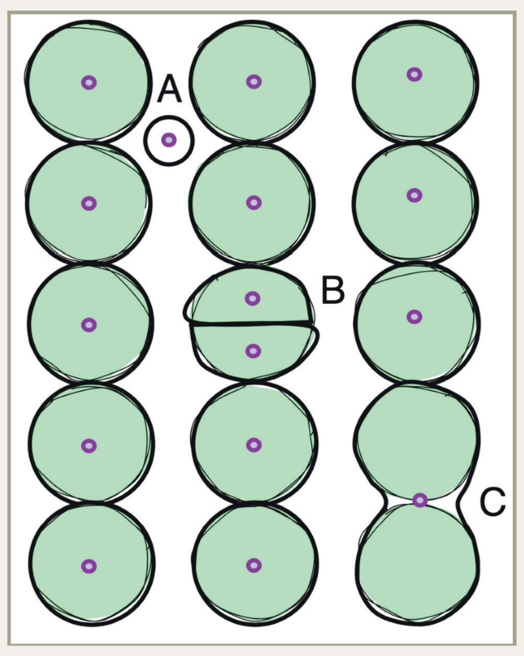


Figure 1a: Illustrating anomaly classifications in the context of this study. A) is a false positive. B) is Over-segmentation. C) is Under-segmentation.

Figure 1b: Example orchard

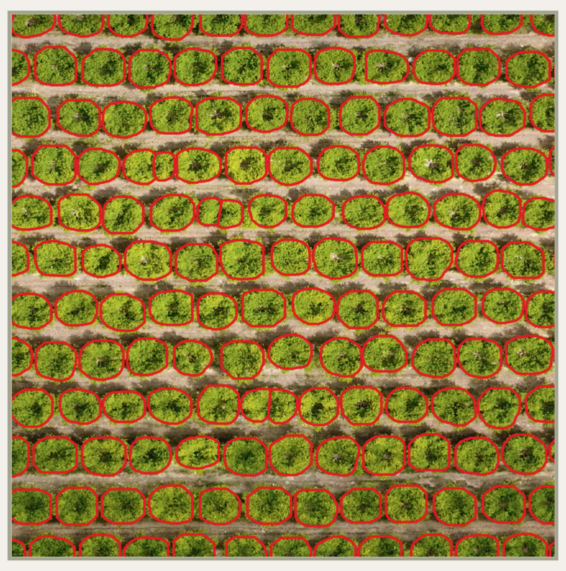
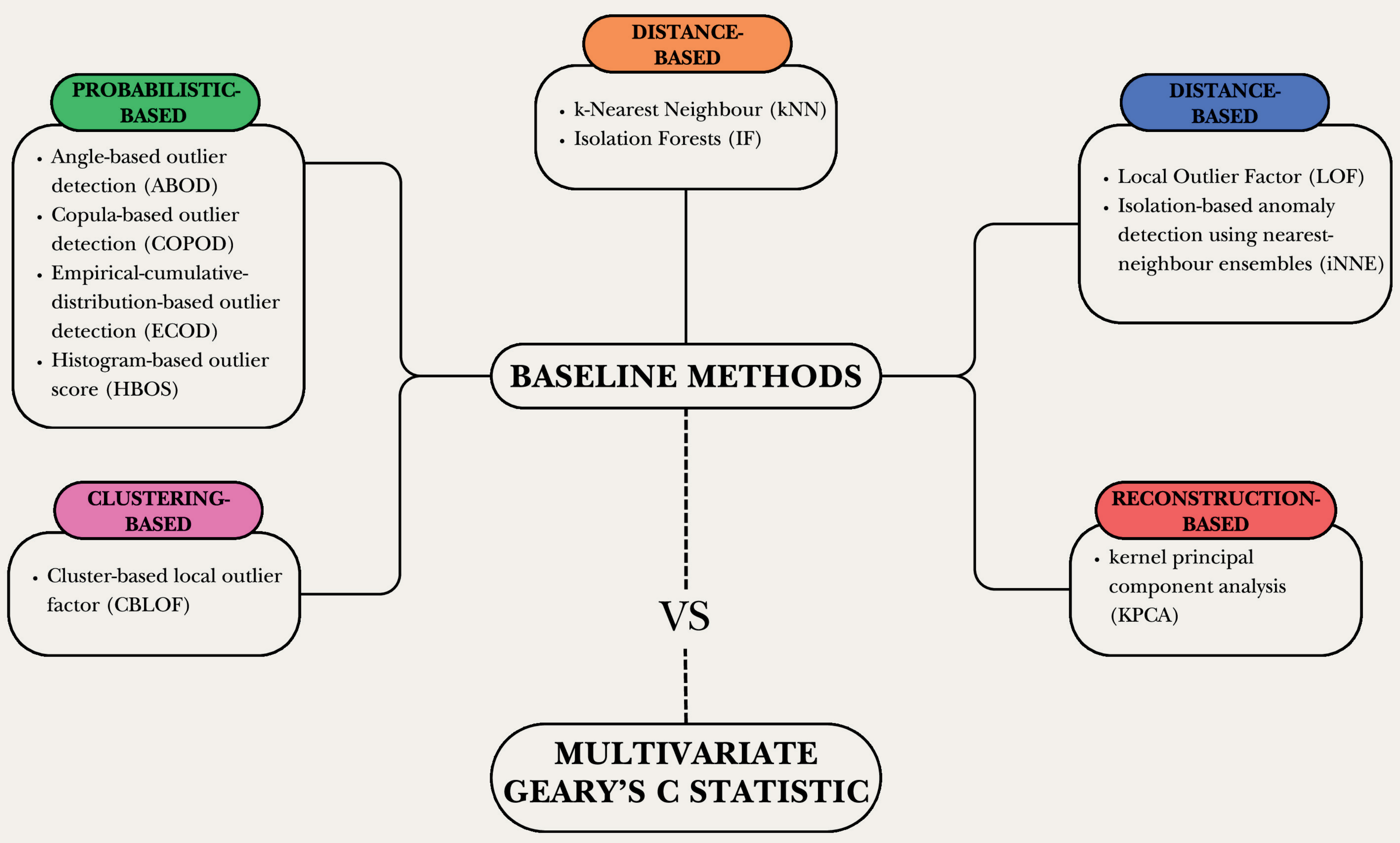


Figure 1c: Individual tree crown delineations

Implemented Methods



Proposed Methodology

- **Local Geary's C Statistic** is popular in univariate studies examining clustering of spatial extents. This is calculated as a weighted square distance between an observation, i, and it's immediate neighbour, j,

$$c_i = \sum_j w_{ij}(x_i - x_j)^2$$

- This statistic is additive in the attribute dimension, allowing for use in the **multivariate context** for K attributes,

$$\boldsymbol{c}_i = \sum_{p=1}^{K} c_{p,i}$$

  - where p denotes the attribute.
- Spatial context is built using **Delaunay triangulation** to construct a spatial weights matrix, w_{ij} ,
- $log(\boldsymbol{c}_i)$ statistic is taken as the measure of outlierness.
- Outlierness is observed via the c statistic, where high values are heterogeneous and considered anomalous.
- Outlierness is a score that depicts how much of an outlier a data instance is.

References

[1] Anselin, L. (2019). A local indicator of multivariate spatial association: Extending geary's c. Geographical Analysis, 51(2):133–150.
[2] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6):610–621.
[3] Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. Pattern Recognition, 37(1):1–19.
[4] Zhao, Y., Nasrullah, Z. and Li, Z., 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. Journal of machine learning research (JMLR), 20(96), pp.1-7.
[5] Han, S., Hu, X., Huang, H., Jiang, M., & Zhao, Y. (2022). Adbench: Anomaly detection benchmark. Advances in Neural Information Processing Systems, 35, 32142-32159.
6] Sergio J. Rey and Luc Anselin. PySAL: A Python Library of Spatial Analytical Methods. The Review of Regional Studies, 37(1):5–27, 2007.
Bradski, G. (2000). The OpenCV library. Dr. Dobb's Journal of Software Tools.
7] Python Software Foundation. Python Language Reference, version 3.11.9
[8] QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation.

Feature Engineering with Image Analysis

- Image analysis techniques have been used to engineer **hand-picked features**.
- Three groups of features are used to characterise anomalies.

1. **Spectral Indices**
   - Used to discriminate biochemical characteristics via spectral absorption levels,
   - To separate vegetation from other matter using vegetative indices,
   - e.g. Near-Infrared Reflectance, Normalized difference vegetation index.
2. **Texture Descriptors**
   - Used to distinguish the texture of the canopy floor from tree crowns,
   - e.g. Gray-Level Co-occurrence matrix statistics (Haralick Features)[2].
3. **Shape Descriptors**
   - Used to characterise the shape of the estimated delineations,
   - e.g. circularity, ellipticity, bending energy, Zernicke moments[3], etc.

Experimental Setup

- In each orchard, the data will be split 60% training and 40% validation using a stratified sample to preserve relative class frequencies. Cross-validation is repeated over 10 independent tests.
- Receiver operating curve (ROC) is obtained over the validation set.
- Average area under the curve of the ROC, is the mean over the repeated experiments.

Results

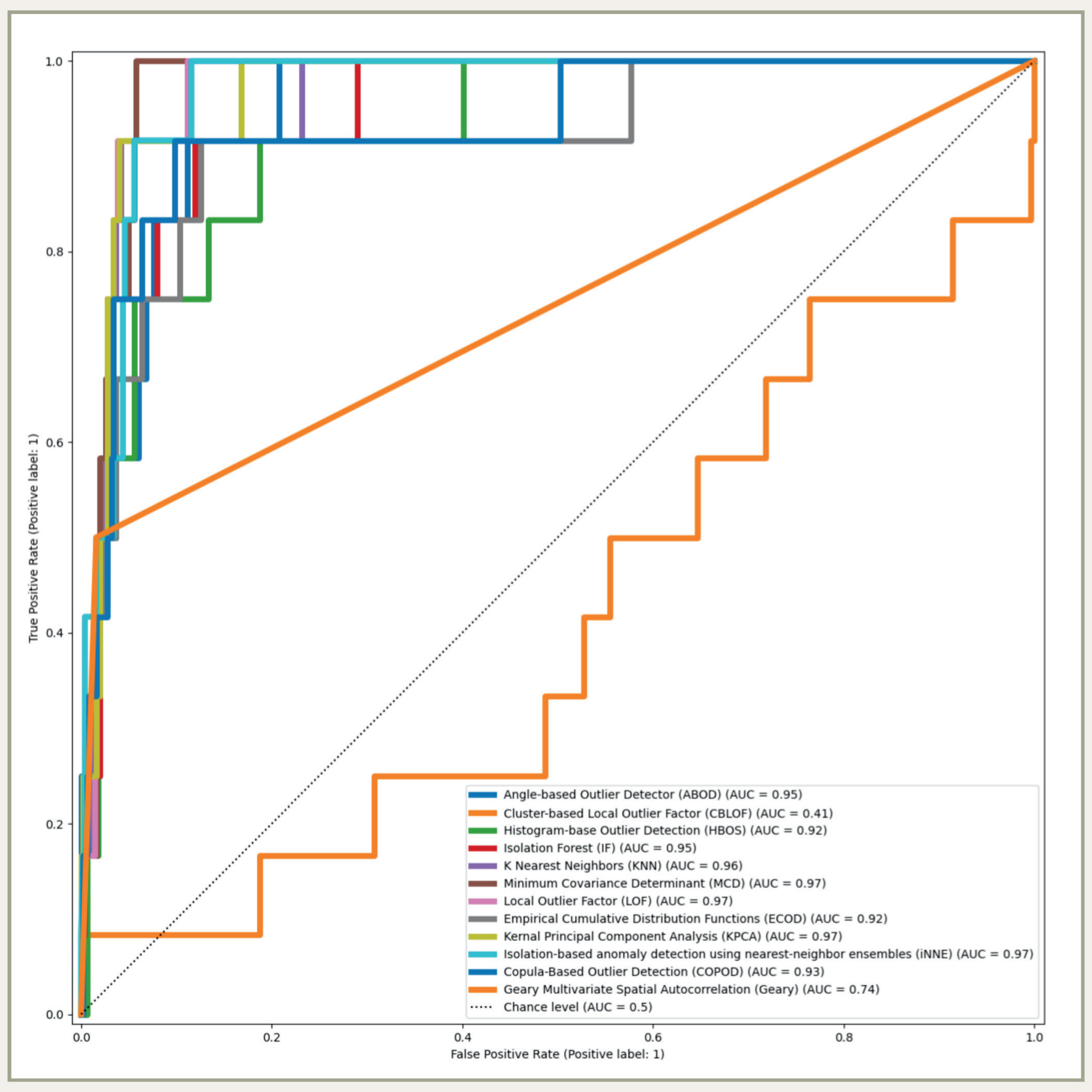


Figure 2a: Receiver operating curve of Orchard 4

Table 1: Average area under the curve of the receiver operating curve, obtained in an inductive setting

Algorithm	Orchard 1	Orchard 2	Orchard 3	Orchard 4	Average
ABOD	0.9812	0.8689	0.8899	0.9253	0.9163
CBLOF	0.9877	0.8619	0.9511	**0.9823 (2)**	0.9457
HBOS	0.9911	0.8795	0.9295	0.9586	0.9397
IF	**0.9908 (2)**	0.8429	0.9360	0.9694	0.9348
KNN	0.9904	0.9020	0.9494	0.9804	0.9555
MCD	0.9855	**0.9623 (1)**	**0.9768 (1)**	0.9808	**0.9763 (1)**
LOF	0.9863	**0.9279 (2)**	0.8996	**0.9839 (1)**	0.9494
ECOD	0.9881	0.8625	0.9242	0.9565	0.9328
KPCA	0.9902	0.9137	**0.9594 (2)**	0.9814	**0.9612 (2)**
iNNE	0.9831	0.8952	0.9512	0.9817	0.9528
COPOD	**0.9930 (1)**	0.8830	0.9064	0.9635	0.9365
Geary	0.9162	0.4900	0.5221	0.7420	0.6676

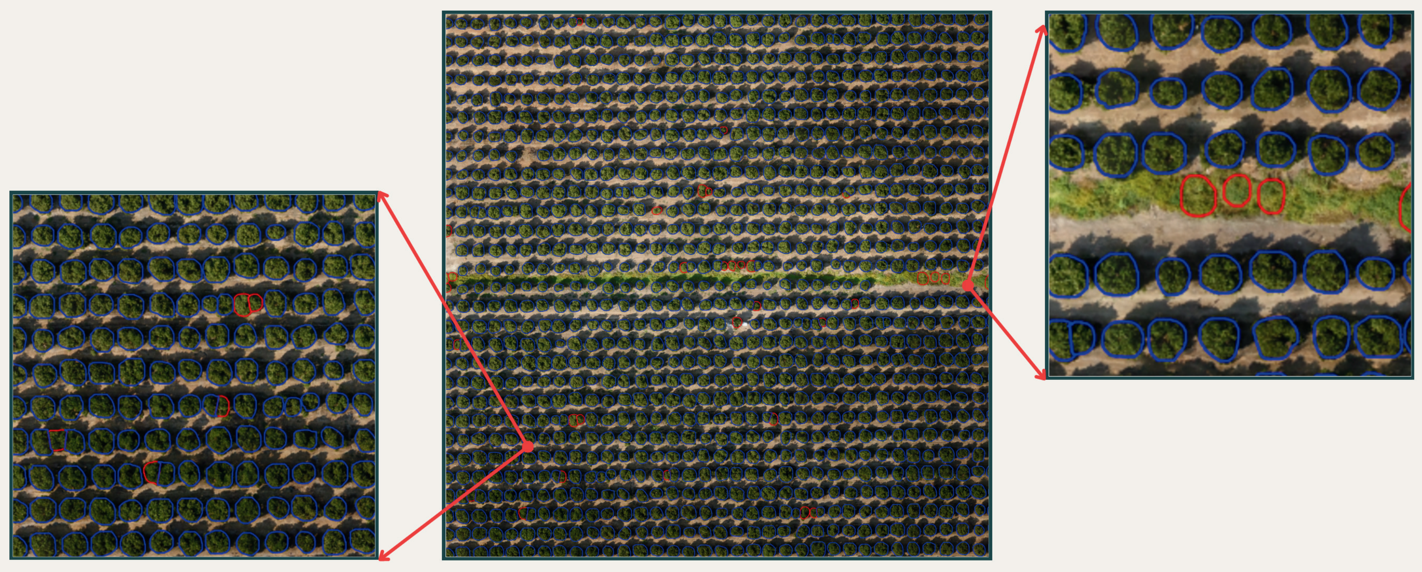


Figure 2b: Detected anomalies in orchard 4 using LOF

Conclusion

- Figure 2a, demonstrates that **CBLOF** are unreliable at default settings.
- Table 1, shows orchard 1 and orchard 3 work well with global methods, orchard 4 works well with local methods and orchard 2 desires a combination.
  - Local methods such as (LOF, CBLOF, iNNE and Geary) find **contextual anomalies**, such as over-segmentation and under-segmentation; e.g. Geary has much better performance in orchard 2 and orchard 4, than the others.
  - Global methods find **point anomalies** such as false-positives.
- **Further work:**
  - Hyperparameter tuning of detection methods.
  - Examination of methods over varying thresholds.
  - Ablation study to assess feature importance.
  - Transductive testing to explore