Texture-based Fruit Detection via Images using the Smooth Patterns on the Fruit

Zania S Pothen[1] and Stephen Nuske[2]

***Abstract*— This paper describes a keypoint detection algorithm to accurately detect round fruits in high resolution imagery. The significant challenge associated with round fruits such as grapes and apples is that the surface is smooth and lacks definition and contrasting features, the contours of the fruit may be partially occluded, and the color of the fruit often blends with background foliage. We propose a fruit detection algorithm that utilizes the gradual variation of intensity and gradient orientation on the surface of the fruit. Candidate fruit locations, or "seed points" are tested for both monotonically decreasing intensity and gradient orientation profiles. Candidate fruit locations that pass the initial filter are classified using modified histogram of oriented gradients combined with a pairwise intensity comparison texture descriptor and random forest classifier. We analyse the performance of the fruit detection algorithm on image datasets of grapes and apples using human labeled images as ground truth. Our method to detect candidate fruit locations is scale invariant, robust to partial occlusions and more accurate than existing methods. We achieve overall F1 accuracy score of 0.82 for grapes and 0.80 for apples. We demonstrate our method is more accurate than existing methods.**

I. Introduction

Growers of fruit crops currently have access to very limited information from the current state of the crop in their field. Automated image analysis systems for fruit crops provide growers with high resolution spatial data on their crop yield. This data enables farmers to switch from old inefficient farming "one-size fits all" paradigm to a more cost-effective, site-specific field management system boosting the crop quality and conserving resources on the farm. Such automated imaging systems can also be integrated into fully robotic harvesting or thinning systems.

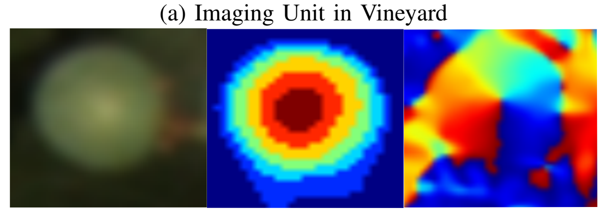
Currently there are no systems available for growers to measure crop yield with high resolution during the growing season. Crop yield is a desirable attribute to be monitored and managed. The current process to estimate yield is by monitoring the farm at harvest time, and recording data during each growing season. However, yield can vary by large amounts from year to year, and using harvest estimates is an extremely coarse approximation of yield. In order to get accurate dense measures of crop yield, the crop needs to be continuously measured during the growing season. The obvious solution would be to exhaustively monitor the fields.

*This work supported by the United States Department of Agriculture (Award number: 2012-67021-19958) and the National Grape and Wine Initiative (info@ngwi.org).

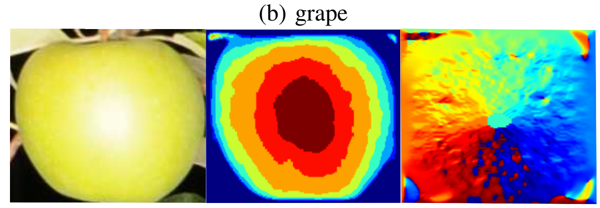
[1]Zania S Pothen, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 zpothen@andrew.cmu.edu

[2]Stephen Nuske, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 nuske@cmu.edu

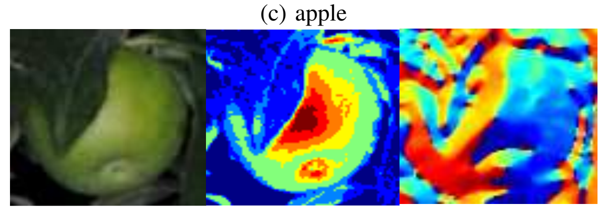
(a) Imaging Unit in Vineyard



(b) grape



(c) apple



(d) partially occluded apple

Fig. 1: An overview of the Yield Mapping System (a) Imaging Unit in Vineyard – consisting of stereo cameras and a pair of flashes for capturing high resolution images. [(b),(c),(d)] Example images of smooth round fruit and the distinct patterns formed on them - (b) a grape, (c) an apple and (d) the visible portion of a partially occluded apple. (Left) RGB Images of the fruit. (Middle) The Instensity pattern on the fruit surface viewed using the jet heatmap of the grayscale image. The deep red tones represent intensity values close to the maximum intensity found at the center of the fruit while the successive color tones represent intensity bands of decreasing strength. (Right) The Orientation pattern viewed using a jet heatmap of the gradient orientation image of the fruit.

While the approach might work well for small sized fields, it becomes economically intractable for larger fields owing to the labour intensive nature of the work. Additionally, the manual counting mechanism is performed just before harvest.

Over the past few years, our research group has focused on developing a vision-based system for automatic fruit-detection and high resolution yield-estimation. Our current system is deployed on a vehicle operating at high velocities (>1.5 m/s), and captures images using a custom hardware configuration and high-powered flash lighting. The images are processed for crop-yield statistics.

The broad steps employed in predicting yield automatically and non-destructively are:

- Collect images of fruit-wall in each row of the field using custom hardware (Figure 1a)
- Detect and count fruit in individual images
- Associate data between fruit locations in the image and physical locations in the real-world
- Generate high resolution yield estimates using the registered fruit information and the vehicle state information

The details of the system and approach is described in Nuske et al. [12]–[14]

We are motivated to increase accuracy of our system and this paper focuses on novel image processing approaches to specifically leverage the unique smooth patterns on the surface of the fruit (Figure 1(b-d)). The smooth texture on the fruit's surface results in a distinct intensity profile and gradient orientation pattern. These patterns can be used to distinguish between fruit and background foliage. We present a novel keypoint detector, called Angular Invariant Maximal Detector, for detecting smooth round fruit such as grapes and apples, and it has the following novel attributes which we see as the contributions of this work:

1) The Angular Invariant Maximal is scale invariant
2) Color-agnostic, operates in most challenging situation of green immature fruit over a cluttered green leaf background
3) Robust to partial occlusions and
4) Detects a variety of round fruit, such as grapes and apples with very high precision and with little need for manual parameter specification.

The rest of the paper is organized as follows: 1) a section on related work on fruit detection, 2) implementation details of the Maximal Orientation Detector, 3) a description of the data-sets and the experimental setup and 4) the results and conclusion.

II. Related Work

Current approaches for detecting fruit in images are based on three different types of visual cues. The three different cues of fruit appearance correspond to color, shape, and texture.

Fruit detection methods based on color are useful only for segmenting fruit that are of different color to the green leaf background. Examples have been shown for the following fruit - mangoes (Payne et al. [15]), apples and grapes (Dunn and Martin [5]).

Shape and contour-based approaches overcome some of the limitations of color-based methods. Bansal et al. [2] describes a method for detecting immature green citrus fruits, by using the symmetrical shape of smooth spherical fruit. Rabatel & Guizard [11] present an approach to detect berry shapes from visible fruit edges using an elliptical contour model. Sengupta and Lee [16] propose a method to detect fruit using shape analysis. These methods give reasonable performance in uncluttered environments where occlusions do not distort the contour of the fruit.

Texture-based methods of fruit detection, are typically associated with an external illumination source. This was exploited by Grossetete et al. [6], who demonstrate that a hand-held device can be used to measure the size of isolated clusters of fruit. Similar approaches detect the shading on apples (Wang et al. [18]) and oranges (Swanson et al. [17]). This paper presents a texture-based approach for curved fruit detection, that unlike existing work is scale invariant, is robust to partial occlusions and requires little parameter adjustment when switching between crops and varieties.

III. Approach

Several challenges are associated with fruit detection in unstructured field environments. They are

1) Large variation in lighting using natural illumination causes inconsistent shading and high contrast regions in the image.
2) Lighting variation causes low contrast between the fruit and background vegetation.

The large variation in natural lighting can be overcome to some extent by imaging with a well-designed flash and camera pair. However, even with controlled lighting, there is still the problem of low-contrast between the fruit and background vegetation. The fruit detection mechanism described in this paper addresses these challenges and is able to successfully count fruit in cluttered vineyard and orchard environments.

A. Fruit Detection Overview

The two approaches for finding fruit locations in images are: a thorough search of each point in the image or identifying a set of points, known as keypoints, that have a high probability of being classified as fruit. Hung et al. [7] and Hung et al. [8] present an approach that performs per-pixel classification using unsupervised multi-scale RGB and IR feature learning for fruit segmentation. Though this approach is effective for a large variety of fruit, it is not possible to run this in real time as it is computationally expensive to classify each and every point in an 4288x2848 image. Instead, by first detecting keypoints with an efficient algorithm, the search space is narrowed down from a million points to less than a few thousand points and hence the computation required is drastically reduced. Nuske et al. [12] presents a technique to identify potential fruit locations by locating radially symmetric points and Nuske et al. [14] presents a method, known as the Invariant Maximal detector, that uses the distinct shading on the fruit caused by flash to detect the keypoints. Both these method provide high recall rate and can be run in less than 0.2s per image. In this paper we present an alternate keypoint detection algorithm that utilizes the patterns in the intensity profile and gradient orientation images to detect fruit with higher recall and precision than the Radial Symmetry and Invariant Maximal detector. The goal of keypoint detection is to quickly find potential fruit locations with a high percentage of the true fruit centers being detected. At this stage there can be number of false positives being detected without detriment to eventual accuracy because these false detections can be filtered out using a feature classification approach. The classification method uses high-dimensional features extracted around each keypoint that describe the visual appearance at the keypoint. Then using a classification algorithm the true-fruit keypoints can be identified as those whose features are similar to fruit appearance model constructed before run-time. Using the approach detailed above, our algorithm for fruit detection is split into the following blocks:

1) Keypoint Detection - detection of potential fruit locations
2) Extract feature descriptors for each keypoint
3) Classify keypoints as fruit / not fruit

B. Keypoint Detection

Here, we describe the Angular Invariant Maximal. The detector locates the fruit center by utilizing the distinct intensity profile and gradient orientation pattern of the fruit surface.

1) Terminology: The notation used in the remainder of this paper is as follows:

- SP - Seedpoints- points corresponding to regional maxima.
- x_m,y_m is the x y coordinate of the image seedpoint.
- r radii from the seed point.
- θ - angle in sector.
- I - 8-bit single channel gray-scale image. $I(x_m, y_m, r, \theta)$ returns the intensity value at pixel location $(x_m + r\cos(\theta), x_m + r\sin(\theta))$.
- S_j - sector corresponding to scan-angle θ.
- $l_{(i)}$ - single scan line in sector S.

2) Angular Invariant Maximal: The effect of specular reflection on the smooth round shaped fruit surface is that the fruit center has the maximum intensity and the intensity decreases monotonically towards the edges of the fruit. The intensity pattern formed on the fruit surface, due to this, is visible as concentric bands of decreasing intensity in the grayscale image I and the gradient orientation $\arctan(\frac{\partial I}{\partial y}, \frac{\partial I}{\partial x})$ at each point on the fruit surface is directed towards the fruit center forming an easily identifiable angular pattern (Figure 1(b-d), 2).

3) Keypoint Detection Steps: Fruit centers surrounded by a consistent pattern of intensity and gradient orientation are identified using the following steps:

To find possible fruit centers, we first locate seed points $SP = (x_m, y_m)$ across the entire image that are the local

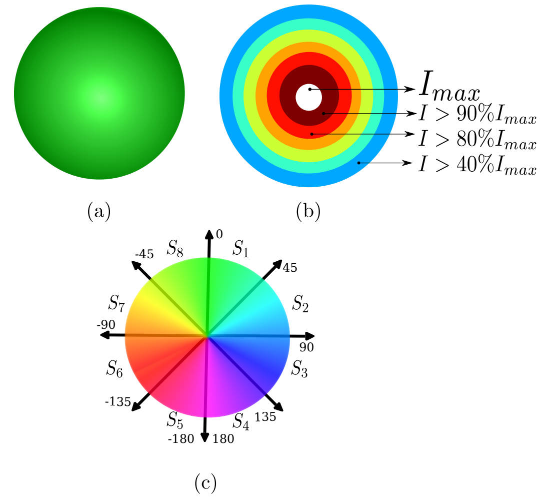


Fig. 2: Synthetic images of round fruit: The maximal intensity, Imax, formed at the center of the fruit surface - is seen as a bright - white spot (a). From this maximal intensity point the intensity drops gradually away from the center towards the edges of the fruit. This results in (b) intensity bands of decreasing strength (each intensity band is depicted in a different color) and in (c) a distinct orientation pattern which varies from $[-180,\ 180]$, represented using gradual variation in color tone, across the surface of the fruit and which can be segmented into angular segments S_1 to S_8.

maxima $\frac{\partial I(x_m, y_m, r, \theta)}{\partial r}|_{(r=0)} = 0$. We then examine the region $I(x_m, y_m, r, \theta)$ around these seed points to determine whether they match the ideal patterns (Figure 2) found on the fruit surface. For this, the region around each seed point is divided into 8 sectors $\{S_1, \ldots, S_8\}$. Each sector is grown radially outward along three scan lines $\{l_0(r,\theta), l_1(r, \theta + \frac{\pi}{12}), l_2(r, \theta + \frac{\pi}{6})\}$. The scan lines are grown only if

- the intensity of the points on the scan line decreases monotonically with r.
- the difference in the intensity of the seed point and the points on the scan line is greater than a threshold η. The conditions are expressed as $I(x_m, y_m, r,\ \theta) > \eta I((x_m, y_m, 0, \theta)$

While growing the scan lines, a record of the number of intensity bands per scan line and the number of points in each sector that have the same gradient orientation as the dominant angle of the sector is maintained. Each band represents a discrete drop in intensity, which we fix at 10%. The sector is grown until one of the following conditions are met

- a majority of its scan lines have already recorded at least a certain number of bands of intensity drop
- a majority of its scan lines have stopped growing

We consider this a valid center if half of the sectors have a minimum number of intensity bands and the sectors have at least half of their gradient orientation matching the dominant angle of the sector.

This approach does not require any further inputs and is effective over different grape and apple varieties of varying sizes.

C. Feature Descriptor

Once keypoints are found we then form feature descriptors of the visual appearance around the keypoint. Color based descriptors that use RGB and L*a*b channels of the image are not suitable for discriminating fruit that have a similar color to its background. Local texture based descriptors like- Surf (Bay et al. [3]), SIFT (Lowe [9]) or FREAK (Alahi et al. [1]) - are invariant to lighting changes and do not rely on fruit having distinguishable color from the background. However, we have found the accuracy of the commonly used SIFT (HOG) and SURF features is restricted because they often over-emphasize the magnitude of gradient intensity which is almost negligible on the surface of the fruit, where the change of intensity is gradual and intensity contrast is very low. We overcome these restrictions by using a feature descriptor that uses the oriented gradients without the gradient magnitude to describe the texture of the fruit.

The feature descriptor - Radial Histogram of Oriented Gradients plus Pairwise Intensity Comparisons descriptor - describes the region around the center of the fruit using a modified histogram of oriented gradient and a pairwise comparison of intensity bands.

1) Radial Histogram of Oriented Gradients (RadHOG): To create the modified histogram of oriented gradients, we extract a circular patch, for each keypoint, from the orientation image. This patch is then subdivided into 16 angular sectors. For each sector, a histogram of 8 orientation bins of $45°$ each is formed to cover the $-180°$ to $180°$ range of orientations. The number of angular sectors and orientation bins have been empirically chosen. This 16x8 gradient histogram forms the first part of the descriptor (128 dimensions).

2) Radial Pairwise Intensity Comparisons (RadPIC) : To create the second part of the descriptor we utilize the intensity pattern formed on the surface of the fruit. The intensity value is at its highest at the center of the fruit and it gradually decreases as the radius from the center is increased. The intensity value of all the points at a given radii is the constant. To create the second part of the descriptor we extract a circular patch of radius R, for each keypoint, from the grayscale image. On which, we construct a set of concentric circles of increasing radii from the keypoint. For each circle, we sum the difference between the intensity value of the keypoint and each pixel that forms the circle. The resulting R (optimal R for our grape datasets is 15 and 35 for our apple datasets) length string is concatenated with the 16x8 gradient histogram to form the feature descriptor.

D. Classification of keypoint features

For classification we use the random forest classifier described in Breiman [4]. We have used random forest classifier for its relative ease in tuning. We use the feature descriptors computed at keypoints for a subset of images known as the training set to build the random forest classifier. The images for the training set are sampled randomly from a dataset and the fruit centers in these images are manually defined. The features corresponding to actual fruit centers serve as positive samples while the features of keypoints that do not align with the manually defined fruit centers are treated as negative samples of fruit appearance for constructing the random forest classifier. The generated random forest classifier is then used to classify the candidate features extracted in each image in the dataset as fruit or not.

(a) Apples

(b) Grapes Varieties

Fig. 3: Images showing the detected keypoints for different varieties of apples (a) and grapes (b) using the Angular Invariant Maximal (AIM) detector. From these images we can see that the AIM detector is able to identify fruits of different sizes and is robust to partial occlusions.

IV. Results

A. Sensor Equipment and Datasets

Our imaging system consisting of a pair of RGB stereo cameras and a pair of flashes (Figure 1) is setup to optimize low motion blur, capture increased depth-of-focus, and uses low illumination power for fast-recycle times permitting high-frame rates. This camera and illumination design maintains high image quality at high vehicle velocities and enables deployment on large scales. The imaging system is mounted onto the side of the farm vehicle facing the fruit wall. Depending on the size of the fruit zone a distance of 0.9 to $1.5m$ is maintained between the imaging system and the fruit zone. The farm vehicle was driven at $1.5m/s$ through each row and the images were captured at $5Hz$. Our datasets consist of images of different wine-grape varieties- Petite Syrah, Pinot Noir and Merlot, table grape variety- Scarlet Royal- and apple varieties- Granny Smith, Red Delicious and HoneyCrisp (Table I). The grape datasets were collected from different vineyards across California, while the apple datasets were collected from orchards in Washington State and Pennsylvania.

B. Algorithm setup for performance evaluation

For evaluating the performance of our detection algorithm we perform leave one out cross validation on the images in the training set. The training set is also used, as described

Dataset	Location	Fruit Attributes	Camera	Image Resolution	Flash	Days to harvest
Granny-Smith	Rock Island, WA	Green	Nikon D300s	1072x712	AlienBees ABR800 ringflash	7
Red-Delicious	Rock Island, WA	Red	Nikon D300s	1072x712	AlienBees ABR800 ringflash	14
HoneyCrisp	Bisleville, PA	Green	Nikon D300s	1072x712	AlienBees ABR800 ringflash	60
Scarlet-Royal	Delano, CA	Green	Pointgrey Grasshopper	4288x2848	Xenon flashlamp (5-10J)	90
Pinot-Noir	Galt, CA	Green	Pointgrey Grasshopper	4288x2848	Xenon flashlamp (5-10J)	90
Merlot	Paso Robles, CA	Green	Pointgrey Grasshopper	4288x2848	Xenon flashlamp (5-10J)	90
Petite-Sirah	Galt, CA	Green	Prosilica GE4000	2800x2200	Einstein 640 mono flashlights	90

TABLE I: Dataset details- fruit variety and fruit attributes, sensor details and field conditions.

earlier, for building the random forest feature classifier. Here we use the training set to also evaluate accuracy. Each training set approximately contains N=10 images that have been randomly selected from a dataset and the fruit centers in these images are manually labeled to serve as ground truth for validation. The keypoints identified as fruit by our algorithm, are true positives (TP) if they are near the ground-truth and false positives (FP) if they are not near the manually defined fruit centers. For leave one out cross validation, we separate an image, defined as the validation image, from the training set, compute its keypoints and features and then classify them using the random forest classifier that has been built using the remaining (N-1=9) images.

C. Keypoint Detection Performance

In Table II we compare the average precision and recall values obtained for the orientation detector with those of Radial Symmetry and the Invariant Maximal detector for each of the datasets. For the keypoint detector the main goal is to identify as many fruit centers as possible while narrowing down the search space from a million or so pixels to few thousand (Figure 3). Therefore, the keypoint detector should have a high recall rate. But it is also important that the detector have a reasonable precision rate (minimum 0.05), as a very low precision rate will result in the classifier having to deal with an extremely imbalanced data set with the non-fruit class forming an overwhelming majority which causes the classifier to under perform. With this in mind we note that the Angular Invariant Maximal detector not only provides a high recall rate (>0.95) for different grape and apple varieties but it also provides a higher precision rate (>0.1) than the other two methods (Table II).

The advantages of having a higher precision is clearly seen in the classification results shown in (Figure 4). We note that even though the recall rate of the Angular Invariant Maximal detector is lower than the Invariant Maximal it out performs the Invariant Maximal in the classification step with F1 scores (>0.82) for grapes and (>0.8) for apples due to its higher precision rate.

V. CONCLUSIONS

In this paper we have presented a novel keypoint dectector - Angular Invariant Maximal (AIM) for detection of smooth round fruit. The AIM keypoint detector utilizes the distinct intensity and gradient orientation pattern formed on the surface of the fruit. We have demonstrated that AIM works for different sizes and different varieties of grapes and apples, it is robust to partial occlusions and it has a higher precision that our previous method. The increase in precision in the keypoint detection step has boosted the overall performance of our fruit detection system to 0.82 F1 score in grapes and 0.8 F1 score for apples. In the future we will investigate other imaging configurations and large pools of data to further increase accuracy of this fruit detection and classification system.

VI. ACKNOWLEDGEMENTS

This work is supported in part by the US Department of Agriculture under grant number 20126702119958 and the National Grape and Wine Initiative. The authors would like to thank the following people who assisted in collecting the data for this paper; Qi Wang, Kyle Wilshusen, Luke Yoder, Karen Lewis, Tara Baugher, Luis Sanchez, Anji Perry and Franka Gabler.

REFERENCES

[1] Alahi, A., Ortiz, R., & Vandergheynst, P. (2012). FREAK: Fast retina keypoint. In IEEE Conference on Computer Vision and Pattern Recognition. 2012: pp. 510-517.

[2] Bansal, R., Lee, W.S., Satish, S. (2013) Green citrus detection using fast Fourier transform (FFT) leakage. Precision Agriculture, 14: pp. 5970

[3] Bay, H., Ess, A., Tuytelaars, T., Gool, L. (2008) Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110: pp. 346-359

[4] Breiman, Leo. (2001). Random forests. Machine learning 45.1: pp 5-32.

[5] Dunn, G., & Martin, S. (2004). Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. Australian Journal of Grape and Wine Research, 10: pp 196198.

[6] Grossetete, M., Berthoumieu, Y., Costa, J. P., Germain, C., Lavialle, O., & Grenier, G. (2012). Early estimation of vineyard yield: Site specific counting of berries by using a smartphone. In International Conference of Agricultural Engineering. Abstract number C1915

[7] Hung, C., Nieto, J., Taylor, Z., Underwood, J., & Sukkarieh, S. (2013a). Orchard fruit segmentation using multi-spectral feature learning. In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference pp. 53145320.

Dataset	Angular Invariant Maximal		Radial Symmetry		Invariant Maximal	
	Recall	Precision	Recall	Precision	Recall	Precision
Granny Smith	0.94	0.26	0.86	0.11	0.99	0.01
Honeycrisp	0.96	0.31	0.91	0.05	0.99	0.01
Red Delicious	0.84	0.32	0.91	0.06	0.99	0.01
Mean	0.91	**0.30**	0.89	0.07	0.99	**0.01**
Pinot-Noir	0.94	0.2	0.89	0.14	0.94	0.12
Petite-Sirah	0.94	0.10	0.92	0.06	0.98	0.08
Merlot	0.96	0.2	0.91	0.09	0.84	0.09
Scarlet-Royal	0.98	0.14	0.84	0.16	0.97	0.09
Mean	0.96	**0.16**	0.89	0.11	0.93	**0.01**

TABLE II: Comparison of keypoint detection performance of the Angular Invariant Maximal, Radial Symmetry and Maximal Detector.

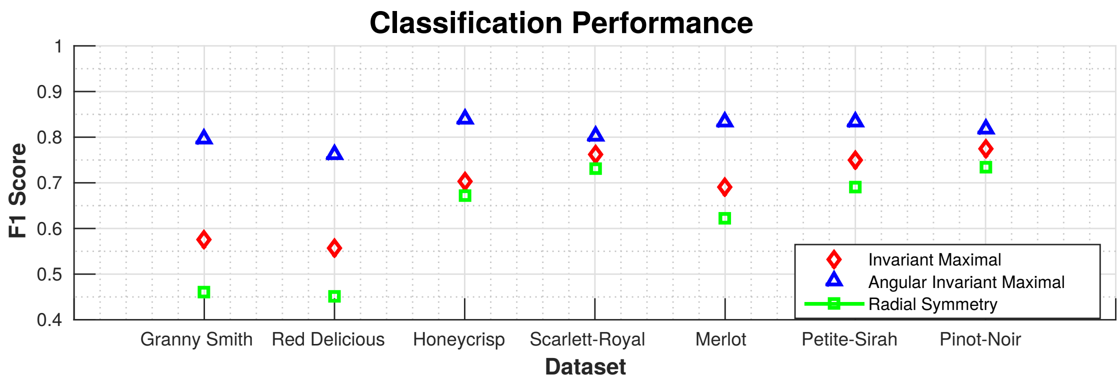


Fig. 4: Comparison of the overall classification performance- F1 score- for each detector (Angular Invariant Maximal, Radial Symmetry and Maximal Detector).

[8] Hung, C., Underwood, J., Nieto, J., & Sukkarieh, S. (2013b). A feature learning based approach for automated fruit yield estimation. In9th International Conference on Field and Service Robotics (FSR).

[9] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60: pp. 91110.

[10] Loy, G., & Zelinsky, A. (2003). Fast radial symmetry for detecting points of interest. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25: pp. 959973.

[11] Rabatel, G, and C Guizard. (2007). Grape berry calibration by computer vision using elliptical model fitting. European Conference on Precision Agriculture, 6: pp. 581- 587.

[12] Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011). Yield estimation in vineyards by visual grape detection. InProceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 2352-2358

[13] Nuske, S., Gupta, K., Narasimhan, S., & Singh, S. (2012). Modeling and calibration visual yield estimates in vineyards. In Proceedings of the International Conference on Field and Service Robotics.

[14] Nuske, S.; Wilshusen, K.; Achar, S.; Yoder, L.; Narasimhan, S.; Singh, S. (2014). Automated Visual Yield Estimation in Vineyards. Journal of Field Robotics, 31: pp. 837860

[15] Payne, A., Walsh, K., Subedi, P., Jarvis, D. (2013) Estimation of mango crop yield using image analysis-segmentation method. Computers and Electronics in Agriculture, 91: pp. 57-64

[16] Sengupta S., Lee W.S. (2012) Identification and Determination of the Number of Green Citrus Fruit under Different Ambient Light Conditions. Presented at International Conference of Agricultural Engineering CIGR

[17] Swanson, M., Dima, C., & Stentz, A. (2010). A multimodal system for yield prediction in citrus trees. In ASABE Annual International Meeting, Pittsburgh, PA.

[18] Wang, Q., Nuske, S., Bergerman, M., & Singh, S. (2012). Automated crop yield estimation for apple orchards. In Proceedings of the International Symposium on Experimental Robotics.