

A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation

Raphael Linker¹

Published online: 8 August 2016
© Springer Science+Business Media New York 2016

Abstract A procedure for estimating the number of mature apples in orchard images captured at night-time with artificial illumination was developed and its potential for estimating yield was investigated. The procedure was tested using four datasets totaling more than 800 images taken with cameras positioned at three heights. The procedure for detecting apples was based on the observation that the light distribution on apples follows a simple pattern in which the perceived light intensity decreases with the distance from a local maximum due to specular reflection. Accordingly, apple detection was achieved by detecting concentric circles (or parts of circles) in binary images obtained via threshold operations. For each dataset, after calibration of the procedure using 12 images, the estimates of the number of apples were within a few percent of the number of apples counted by visual inspection. Yield estimations were obtained via multi-linear models that used between two and six images per tree. The results obtained using all three cameras were only slightly better than those obtained using only two cameras. Using images from only one side of the tree did not worsen the results significantly. Overall, the yield estimated by the best models was within $\pm 10\%$ of the actual yield. However, the standard deviation of the yield estimation errors corresponded to $\sim 26\text{--}37\%$ of the average tree yield, indicating that improvements are still needed in order to achieve accurate yield estimation at the single-tree level.

Keywords Artificial vision · Image analysis · Fruit detection

Introduction

Estimation of the expected yield is still a major challenge in orchards. Such information would be valuable to growers and companies that provide harvesting and post-harvesting services as it would help them manage equipment and labor force more effectively. To

✉ Raphael Linker
linkerr@tx.technion.ac.il

¹ Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, 32000 Haifa, Israel

date, yield estimation is based on visual inspection of a limited number of trees by the grower, which is both time consuming and prone to errors. At first glance, it may seem that this task could be easily automated using computer vision. However, the unstructured and cluttered environment of the tree canopy makes this task very challenging, especially when the fruits are featureless and with a color similar to that of the canopy. Jiménez et al. (2000), Kapach et al. (2012) and Gongal et al. (2015) compiled reviews on the general topic of fruit detection and location in tree images, although Kapach et al. (2012) focused on vision-based fruit location for harvesting robots, which imposes additional constraints on the system (such as computing efficiency and accurate location). These review papers also discussed the pros and cons of the various types of imaging system (e.g. color, depth/time-of-flight, hyperspectral, thermal imaging). Standard color cameras offer obvious advantages in terms of cost and ease of operation, and a large number of studies have focused on the use of such cameras (e.g. Pla et al. 1993; Zhao et al. 2005; Kurtulmus et al. 2011; Zhou et al. 2012; Silwal et al. 2014). However, the images captured with such cameras are very sensitive to the illumination conditions (e.g. Linker et al. 2012; Payne et al. 2013). Night-time imaging using artificial illumination minimizes this issue and may constitute an advantageous alternative (e.g. Sites and Delwiche 1988; Payne et al. 2014; Cohen et al. 2014; Linker and Kelman 2015). The recent studies of Cohen et al. (2014) and Linker and Kelman (2015), which focused on the detection of green mature apples in night-time images, reported encouraging results with a strong linear relationship between the number of objects identified as apples and the number of apples counted by visual inspection of the images. However, as mentioned by Wang et al. (2012), the performance of vision-based yield estimation depends on two factors: (1) how accurately apples are detected by the image processing procedure and (2) how accurately the number of apples visible in the tree image(s) correlates with yield. This second factor was not investigated in the studies of Cohen et al. (2014) and Linker and Kelman (2015). Vision-based estimation of green apple yields has been reported by Wang et al. (2012), who used night-time color images, and Stajanko et al. (2004), who used day-time thermal images. Both studies reported R^2 values between the estimated and actual number of fruit per tree in the range 0.8–0.9. However, these studies were conducted with small trees characterized by low fruit loads and a sparse canopy: Although Wang et al. (2012) did not provide detailed results at the tree level, it can be inferred from the results that the average tree yield was ~ 35 apples; Stajanko et al. (2004) reported yields ranging from 50 to 120 apples per trees. While such “small trees/high planting density” orchards are increasingly popular in Europe and the US, Israeli growers still favor older and larger trees, mostly due to Jewish Laws which prevent consuming the harvest of the first 3 years after planting. As a result, Israeli mature orchards such as the one used in the present study have a much higher fruit load (target yield of ~ 300 – 350 fruit per tree). The trees are also much larger and the canopy is very dense (Fig. 1). The dense canopy makes the detection of fruit especially challenging and the significant depth of the canopy means that a lower percentage of the overall fruit is visible to the camera.

The objective of the present study was 2-fold. First, to develop and test a novel procedure for detecting and locating apples in night-time orchard images. Second, to determine whether tree yield could be inferred from such images. Special emphasis was attached to using calibration sets with as few images/trees as possible since the user needs to provide ground-truth information for these images/trees. The present work should be viewed as an extension of previous works of Linker and co-authors dealing with fruit location and yield estimation: Linker et al. (2012) suggested an approach based on color and texture analysis, which was initially applied to day-time images and later to night-time



Fig. 1 Typical images of the orchards in which the present study was conducted

images (Cohen et al. 2014). Linker and Kelman (2015) used light distribution around local maxima to locate apple in night-time images. These two approaches were compared by Linker et al. (2015), who also reported for the first time preliminary results relative to yield estimation. Besides presenting a novel approach for locating apples, which is somewhat similar to the one described by Linker and Kelman (2015) but simpler to calibrate and implement, this paper investigates more thoroughly the possibility of using night-time images covering only part of a tree for estimating tree yield.

Materials and methods

Datasets

The work was conducted with four datasets of images captured in “Golden Delicious” orchards at the Matityahu Research Station in Northern Israel in 2013, 2014 and 2015. The imaging system and the 2013 dataset (156 images of 26 trees) have been previously described in Linker and Kelman (2015). The same imaging system was used to capture images on July 17, 2014, August 5, 2014, and August 13, 2015. In 2014, images were captured from both sides of the trees (258 images of 43 trees, same trees on both nights) while in 2015 images were captured only from one side of the trees (156 images of 52 trees). In all 3 years, each tree was picked individually at the end of the season to provide ground-truth yield data but, due to a technical mistake, the yield of three trees was not recorded properly in 2013. In all three datasets, the number of fruit visible in each image was determined by visual inspection.

Algorithm

Concept

The development of the image analysis procedure was based on the observation that since apples are basically spherical objects with uniform texture, in night-time images captured with artificial illumination, the light distribution on apples is expected to follow a simple pattern in which the perceived light intensity decreases with the distance from a local maximum due to specular reflection. Figure 2a–c shows that this is indeed the case for

Fig. 2 Two typical sub-images. *Frames a* and *d* show original sub-images (300 by 300 pixels and 500 by 500 pixels, respectively). *Frames b* and *e* show gray-level versions of frames *a* and *d*. *Frames c* and *f* show binary images obtained by applying the thresholding operation $B(i, j) = (\tau_1 \leq I(i, j) \leq \tau_1 + 10 | \tau_2 \leq I(i, j) \leq \tau_2 + 10)$. In frame *c* the inner and outer rings on apples correspond to $\tau_1 = 170$ and $\tau_2 = 200$, respectively. In frame *f* the inner and outer rings on the apple correspond to $\tau_1 = 120$ and $\tau_2 = 170$, respectively. In both cases the results obtained with additional values of τ , which led to additional concentric circles on the apples, are not shown for clarity

ideal “fully visible” apples. In particular, the binary image displayed in Fig. 2c (which was obtained by applying simple thresholds to Fig. 2b) shows that pixels with similar intensities form rings. Furthermore, Fig. 2d–f show that even when an apple is partially occluded, the light distribution on the region of the apple surface which is visible follows a similar circular pattern. In other words, if one considers a set of pixels which have a similar intensity, the “apple pixels” will form part of a ring. Figure 2c and f also shows that the rings corresponding to different intensities are (approximately) concentric. Accordingly, the procedure described below is based on identifying rings (not necessarily complete) and estimating their centers. Apples are detected where a sufficient number of concentric rings are found.

Detailed description of the procedure

The flowchart of the procedure is presented in Fig. 3. The original color image was converted into a gray-scale image according to

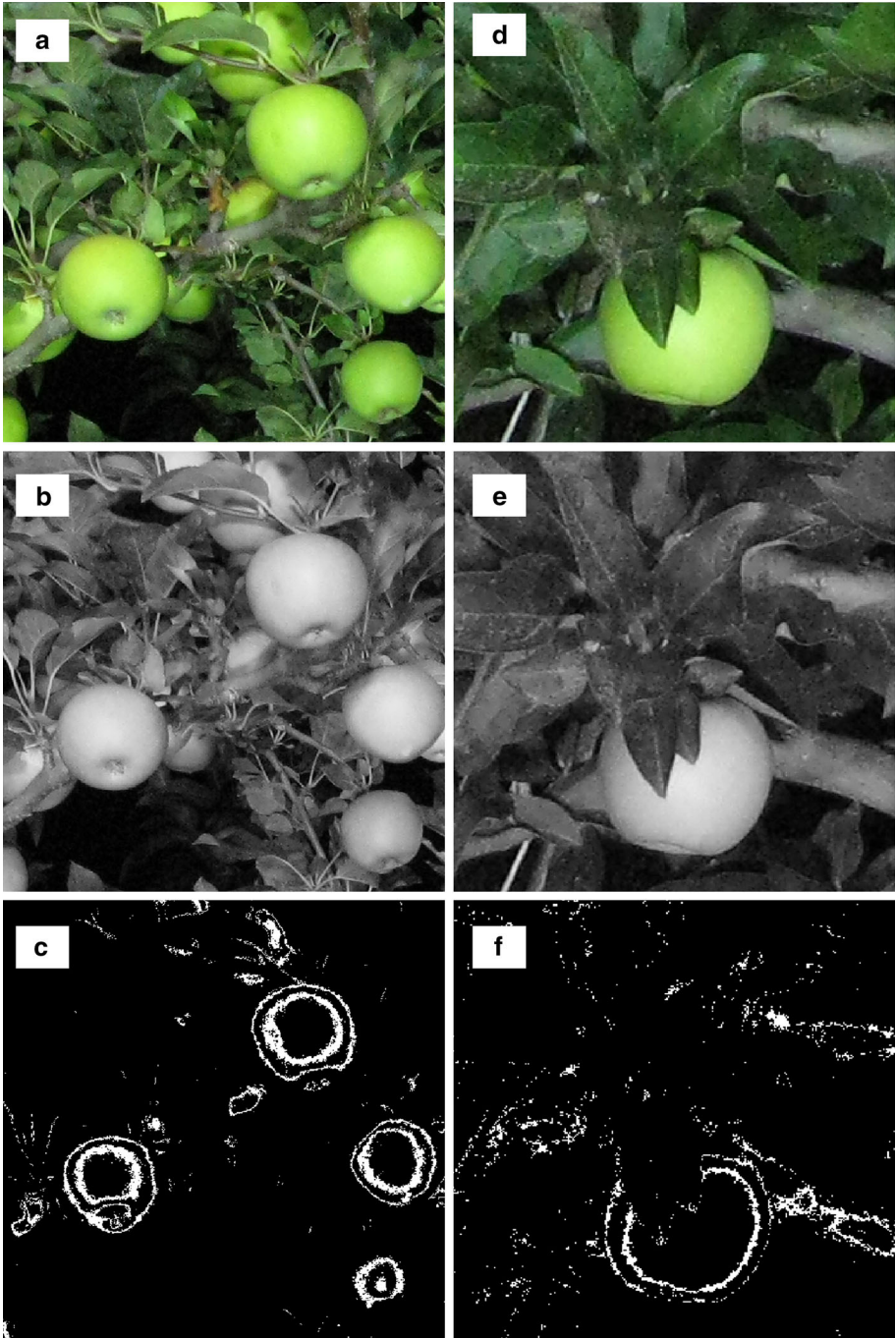
$$I = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B \quad (1)$$

where I denotes the gray level and R , G and B denote the original values of the Red, Green and Blue channels, respectively.

Following this conversion, a binary image was obtained by simple thresholding operation, keeping only the pixels with intensity between τ_{\min} and $\tau_{\min} + 10$. The purpose of this operation was to extract all the pixels within a given intensity range. The Circle Hough Transform (CHT) (in its phase-coding formulation developed by Atherton and Kerbyson (1999)) was used to detect circles within a given radius range in this binary image. The CHT returns not only the estimated center and radius of each circle but also the so-called “accumulator value” associated with each circle. This accumulator value is a measure of the certainty of the circle detection. The results (centers, radii and accumulator values) were stored in a dynamic stack and the procedure was repeated after incrementing the threshold τ until the maximum value τ_{\max} was reached. The stack containing all the detections was then scanned recursively and centers which were within D_{\max} pixels of each other were united, i.e. considered to correspond to concentric circles. Finally, the score (mean accumulator value) associated with each set of concentric circles was calculated:

$$S_i = \frac{\sum_{k=1}^n A_{i,k}}{n_i} \quad (2)$$

where $A_{i,k}$ is the value of the accumulator of the k^{th} circle included in the i^{th} set of concentric circles and n_i is the number of circles in this set. The sets of concentric circles containing a sufficient number of circles and having a high enough score were considered to correspond to apple objects.



Calibration

The whole procedure required four kinds of parameters:

- The minimum and maximum values (τ_{\min} and τ_{\max}) for the iterative thresholding operation
- The search for circles using CHT required a radius range and a minimum value of the accumulator for circle detection.
- Uniting circles required a value for D_{\max} , the maximum distance between centers for considering circles as concentric
- Apple detection required a minimum value for the score of concentric circles. Rather than considering a single parameter, the required score was related to the number of concentric circles as follows:

$$\begin{aligned} \text{Apple detected if } & S_i \geq \theta(n_i) \\ \text{Apple rejected if } & S_i < \theta(n_i) \end{aligned} \quad (3)$$

where S_i is the score associated with the i^{th} set of concentric circles, n_i is the number of concentric circles in this set and $\theta(n)$ is the minimum score required.

The values of τ_{\min} and τ_{\max} can be readily set by inspecting the dynamic range of the images. For the datasets used in this study, such inspection led to selecting $\tau_{\min} = 100$ and $\tau_{\max} = 200$.

With respect to the CHT analysis, preliminary analysis with five images led to performing CHT with two radius ranges with different accumulator thresholds, and combining the results

- Radius range 10–30 pixels, accumulator threshold α_1
- Radius range 31–60 pixels, accumulator threshold α_2

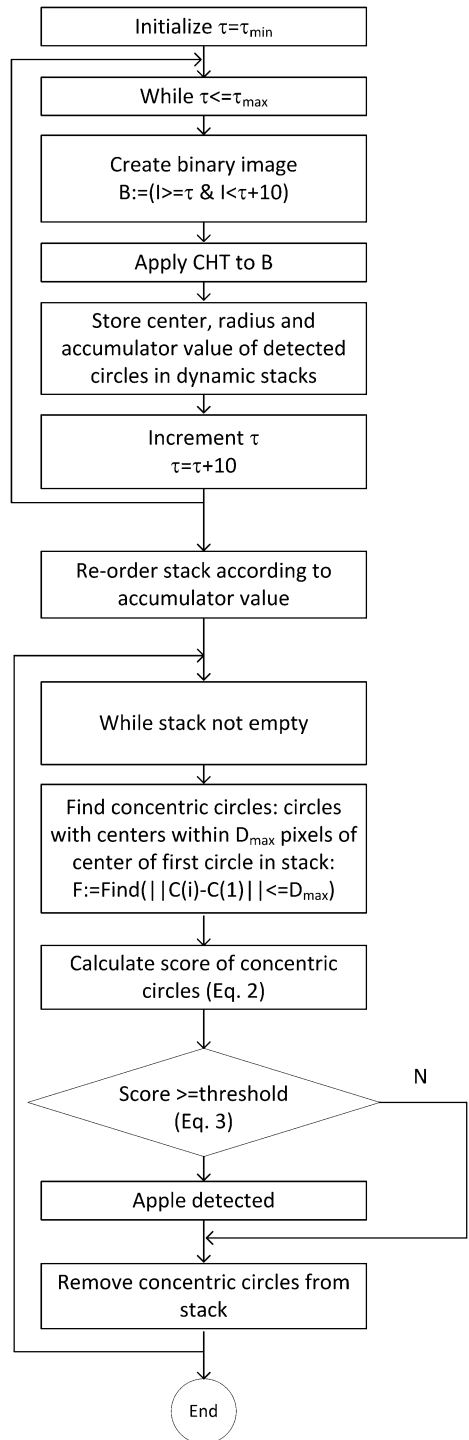
In order to reduce the computation burden, the parameters α_1 and α_2 were restricted to values ranging from 0.02 to 0.20 at 0.03 increments (7 values), leading to 49 possible combinations of (α_1, α_2) . The range for the accumulator threshold was based on the accumulator values observed for typical situations, such as shown in Fig. 2. For instance, the accumulator values for the “apple” inner and outer rings shown in Figs. 2c and f were in the range 0.16–0.20 and 0.08–0.10, respectively.

Calibration of the parameters was performed using a genetic algorithm with the following parameters: Uniform cross-over function, cross-over fraction: 0.8, migration factor: 0.2, parent selection by stochastic universal sampling, mutation using Gaussian distribution, population size: 20, number of generations: 30. The procedure was calibrated separately for each dataset. In each case, 12 images were selected randomly and the genetic algorithm was used to determine the set of parameters which led to the highest coefficient of correlation (R^2) between the number of detections and the actual number of apples.

Comparison with other approaches

The 2013 dataset has been previously used in the studies of Cohen et al. (2014) and Linker and Kelman (2015) who used color-and-texture analysis (described in Linker et al. 2012) and spatial light distribution, respectively. These two approaches were applied to the July 2014 dataset. Since the datasets of 2013 and July 2014 were similar in terms of apple size (apple radius approx. 60 pixels), the procedure used by Linker and Kelman (2015) was applied “as-is”, i.e. without recalibration. The procedure of Cohen et al. (2014) had to be

Fig. 3 Flowchart of the procedure for apple detection. *CHT* Circular Hough Transform



recalibrated. This was achieved using the same 12 calibration images as above, in which close to 350 apples were labeled manually. A genetic algorithm was used to determine the set of parameters which led to the highest coefficient of correlation (R^2) between the number of detections and the actual number of apples.

Yield estimation

Various regression models for estimating tree yield using the information extracted from the images were developed. As detailed in Linker and Kelman (2015), the imaging system captured three images at fixed heights. In 2013 and 2014, images were captured from both sides of the tree while in 2015 images were captured only from one side. Different yield models were investigated, based on different sets of images: Images from all heights (from one side only or both sides); bottom and middle images only (from one side only or both sides); middle and top images only (from one side only or both sides); and bottom and top images only (from one side only or both sides). The general form of the models based on images from both tree sides was

$$Y_k = \alpha(B_{1,k} + B_{2,k}) + \beta(M_{1,k} + M_{2,k}) + \gamma(T_{1,k} + T_{2,k}) + \lambda \quad (4)$$

where $B_{i,k}$, $M_{i,k}$ and $T_{i,k}$ are the number of detections in the bottom, middle and top image of the side i of the k^{th} tree, respectively. Similarly, the general form of the models based on images of one side only was

$$Y_k = \alpha B_{1,k} + \beta M_{1,k} + \gamma T_{1,k} + \lambda \quad (5)$$

In both cases, the coefficients corresponding to images not used in the model were set to 0. Altogether eight types of models were developed for the 2013 and 2014 datasets, and four types of models for the 2015 dataset. Due to the small size of the 2013 dataset, it was not possible to split this dataset into two distinct calibration and validation sub-sets. Therefore models were created using all the available data and the objective was merely to establish the feasibility of developing such models. For the 2014 and 2015 datasets, calibration of the models was performed using the images of 20 trees selected randomly and the remaining trees were used to validate the models. The sensitivity of the results to the choice of the calibration trees was investigated by repeating the procedure ten times using different calibration trees. This Monte-Carlo cross-validation approach was preferred over the more common k-fold cross-validation approach since k-fold cross-validation with $k = 10$ would have resulted in relatively large calibration sets (47 trees) while one of the objectives of this study was to rely on as few calibration trees as possible. In addition, the Monte-Carlo approach made it possible to use the same number of calibration trees for both years despite the different sizes of the datasets.

Results

Estimation of the number of apples in images

Preliminary analysis with eight of the calibration images (two from each dataset, selected randomly) indicated that most apples were associated with at least three concentric circles, which led to considering the following score threshold $\theta(n)$:

$$\theta(n) = \begin{cases} \theta_1 & n = 1 \\ \theta_2 & n = 2 \\ \theta_3 & n = 3 \\ \theta_4 & n \geq 4 \end{cases} \tag{6}$$

The corresponding procedure required calibration of six parameters: (D_{\max} , A , θ_1 , θ_2 , θ_3 , θ_4), where A denotes the identifier of the (α_1 , α_2) combination (integer variable). The results of the calibration are presented in Table 1. For all four datasets, a correlation coefficient higher than 0.90 was obtained between the number of detections and the actual number of apples according to the manual count. However, the relationship between the two varied between the datasets. Also, the fitted values of most of the parameters varied between dataset, with the noticeable exception of α_1 and α_2 (the accumulator thresholds of the CHT procedure). Increasing the size of θ (i.e. considering different score thresholds for 4, 5, or more circles) improved the results only marginally (details not shown).

The results of the calibration are also presented graphically in the left frames of Fig. 4. For each dataset, the calibrated model was used to estimate the number of apples in all the images, leading to the results presented in the Fig. 4e–h and Table 2. Overall, the error between the actual and estimated number of apples in the images corresponded to less than 1 % of the actual number of apples for three of the datasets and 6 % for the last dataset. There was no consistent relationship between the type of image (i.e. camera position) and the estimation error. Although the standard deviations of the estimation errors appear to be much higher in 2014 and 2015, this is partly due to the fact that the number of apples per image was much lower in 2013: on average 43 apples per image in 2013, compared to 69, 75 and 77 apples per image in the dataset of July 2014, August 2014 and 2015, respectively.

Although the main objective of this study was to obtain a high correlation between the number of objects detected and the actual number of apples, it was still of interest to estimate the accuracy of the procedure by considering the number of false positive and missed detections. These were evaluated for the calibration images and 12 additional images chosen randomly. The results are summarized in Table 3. There was a large difference between the results obtained for the 2013 dataset and the other datasets. For the 2013 dataset, the number of false positive was 1–2 % and the number of missed detections was around 10 % while, for the other datasets, the number of false positive was around 7 % and the number of missed detections was around 33 %. Since the same imaging system was used on all years and the images were captured at a similar developmental stage, these large differences appear to be related to the number of fruit visible in the images, which, as noted above, was much lower in 2013.

Table 1 Results of the calibration procedure

Dataset	Fitted parameter value							R ²	Ratio between number of apples and detections
	α_1	α_2	θ_1	θ_2	θ_3	θ_4	D_{\max}		
2013	0.14	0.05	0.865	0.970	0.145	0.042	0.487	0.99	1.09
July 2014	0.14	0.05	0.668	0.746	0.331	0.083	0.915	0.91	1.44
Aug 2014	0.14	0.05	0.334	0.243	0.121	0.055	0.624	0.98	1.27
2015	0.14	0.05	0.845	0.583	0.271	0.064	0.638	0.93	1.44

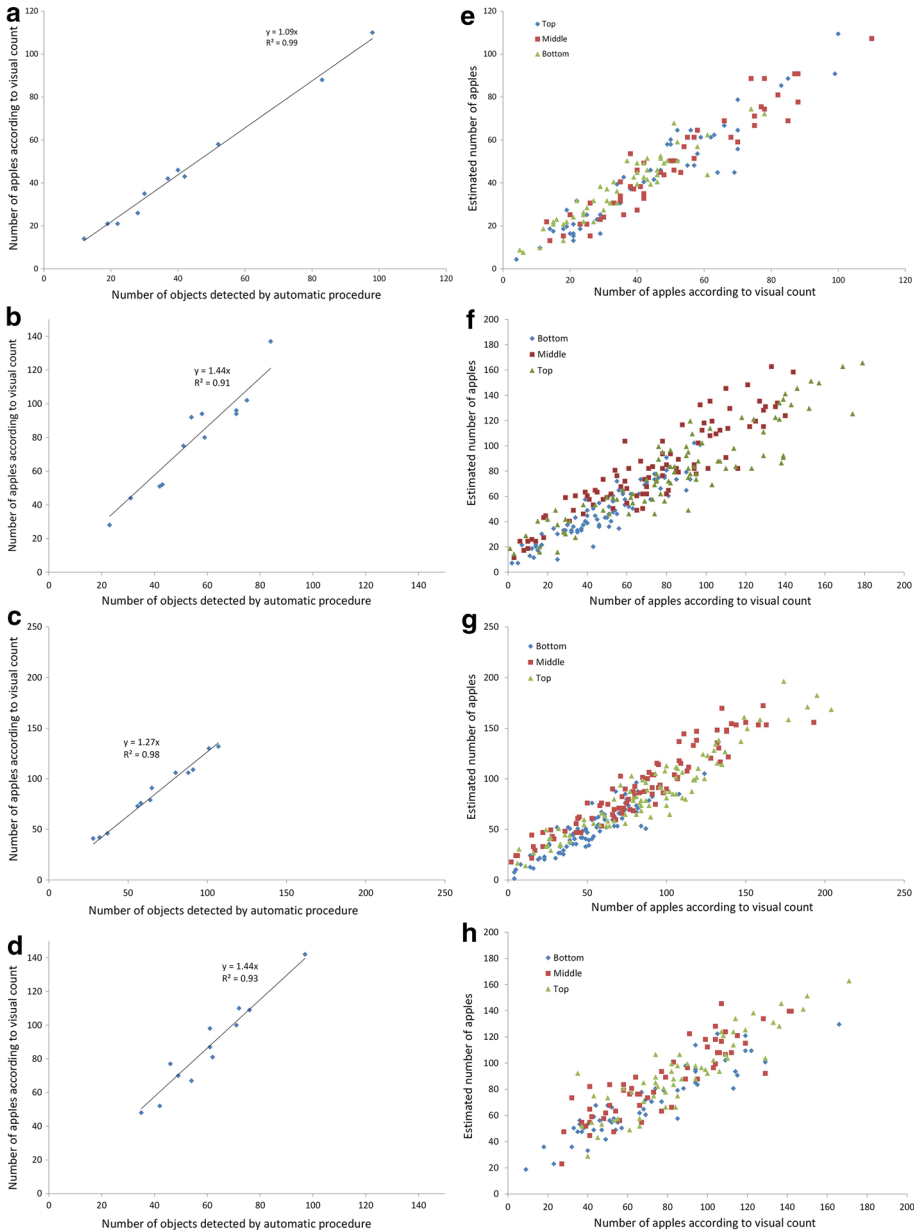


Fig. 4 *Left frames*: Relationship between the number of objects detected and the actual number of apples in the 12 calibration images of the datasets of 2013, July 2014, August 2014 and 2015, respectively. *Right frames*: Actual Versus estimated number of apples in each image of the datasets of 2013, July 2014, August 2014 and 2015, respectively. The *symbols* indicate the camera position

Yield estimation

As mentioned above, multi-linear models were calibrated for each dataset using images from two or three cameras, either from only one side or from both sides of the row. The

Table 2 Results of the apple detection procedure

Images		Dataset			
		2013 (156 images)	July 2014 (258 images)	Aug 2014 (258 images)	2015 (156 images)
Bottom camera	Number of apples according to visual inspection	1832	4264	4569	3945
	Relative estimation error, %	1.61	-1.25	-4.24	1.43
	Std of estimation error, apples	7.31	8.9	10.18	13.75
Middle camera	Number of apples according to visual inspection	2605	6551	7404	4490
	Relative estimation error, %	-2.86	10.39	5.42	11.77
	Std of estimation error, apples	6.69	15.25	17.86	14.35
Top camera	Number of apples according to visual inspection	2276	7473	7885	5062
	Relative estimation error, %	7.4	-6.55	-3.87	5.43
	Std of estimation error, apples	6.32	17.76	14.44	13.93
Total	Number of apples according to visual inspection	6713	18288	19798	13497
	Relative estimation error, %	0.35	0.79	0.56	6.37
	Std of estimation error, apples	6.96	15.39	13.43	14.31

Std standard deviation

Table 3 Rate of false positive and missed detections for the 12 calibration images and 12 additional images

Dataset	Calibration images		Validation images	
	False positive, (%)	Missed apples, (%)	False positive, (%)	Missed apples, (%)
2013	2	12	1	8
July 2014	4	35	7	35
Aug 2014	6	31	9	36
2015	3	34	7	31

results for the dataset of 2013 are presented in Table 4. The standard deviation of the yield estimation error ranged from 78 to 105 fruit/tree, which corresponds to ~20–27 % of the average tree yield (390 fruit/tree). The best results were obtained when using either all three cameras or only the middle and top cameras. Using images from only one side did not affect the results dramatically.

The results for the dataset of July 2014 are presented in Tables 5 and 6. As mentioned above, for each type of model, ten models were calibrated using different subs-sets of the data. The results in Tables 5 and 6 correspond only to the 23 trees not used to calibrate the respective model. Overall, as with the 2013 dataset, very similar results were obtained using images from only one side or both sides of the row (comparison between corresponding columns in Tables 5 and 6). The results obtained using all three cameras were only slightly better than those obtained using only either the bottom and top or the middle

Table 4 Standard deviation of the yield estimation error using different combinations of images

Model based on	Standard deviation of yield estimation error (fruit per tree)			
	All three cameras	Bottom & middle cameras only	Bottom & top cameras only	Middle & top cameras only
One side only	78	97	95	80
Two sides	78	86	105	79

Results for 2013 dataset (23 trees)

and top cameras. The relative error of yield estimation depended on the (random) choice of the trees used for calibration. However, the estimated yield was almost always within $\pm 10\%$ of the actual yield. The standard deviation of the yield estimation error of the models based on all three cameras ranged from 132 to 185 fruit/tree, which corresponds to $\sim 26\text{--}37\%$ of the average tree yield (~ 500 fruit per tree). The estimation error at the tree level can be appreciated in Fig. 5 which shows the results of Model 3 based on all images from both sides of the row.

The results for the dataset of August 2014 and 2015 are presented in Tables 7, 8, 9. As with the July 2014 dataset, using all three cameras or only the bottom and top cameras provided the best results, and using images from both sides did not improve the results appreciably for the August 2014 dataset. Overall, in terms of yield estimation, the results are slightly worse than those obtained for the July 2014 dataset. Still, the yield estimated by the large majority of models is within $\pm 10\%$ of the actual yield. In terms of yield estimation at the level of individual trees, for both datasets, the standard deviation of the yield estimation error ranged from ~ 26 to 35% of the average tree yield (~ 640 fruit per tree in 2015).

Discussion

The performance of vision-based yield estimation must be judged based on two criteria: (1) how accurately apples are detected by the image processing algorithm and (2) how accurately the yield can be estimated using the number of objects detected in the images.

In terms of apple detection, the procedure is far from excellent, with a correct detection rate ranging from ~ 65 to 85% . Furthermore, the number of missed apples appears to depend on the number of apples in the image. Nevertheless, within a given dataset, the detection rate was very consistent, so that it was possible to use a small number of images (12) to establish a relationship between the number of detected objects and the actual number of apples. For each dataset, such a relationship led to an estimate of the total number of fruit which was within a few percent of the number of fruit counted by visual inspection. A comparative investigation with the procedures described by Cohen et al. (2014) and Linker and Kelman (2015) was conducted with the datasets of 2013 and July 2014. Table 10 shows that all three approaches led to similar results. The advantage of the present approach is its simplicity and the fact that it requires calibration of only a small number of parameters. Also, by comparison to approaches that include a supervised classifier (such as Cohen et al. 2014), the present method does not require time-consuming labeling of objects in the calibration images.

Table 5 Mean and standard deviation of the yield estimation error using different combinations of images captured from both sides of the tree rows

Model	Actual yield	All three cameras		Bottom & middle cameras only		Bottom & top cameras only		Middle & top cameras only	
		Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error
1	10678	7	138	7	168	8	139	10	151
2	9858	6	134	18	174	8	135	1	160
3	10712	-8	158	-2	198	-8	165	-8	170
4	10828	2	147	9	180	4	151	5	160
5	11078	1	154	-4	181	0	157	2	167
6	12066	-12	132	-17	163	-13	135	-13	146
7	11327	-1	180	0	189	2	175	5	157
8	11895	8	160	10	196	10	164	9	153
9	11742	-3	176	-3	200	-2	169	-2	175
10	10409	0	160	-6	181	0	162	0	164

Results for the July 2014 dataset (43 trees). Each row corresponds to the validation results of a model calibrated with a different subset of 20 trees. The highest errors for each configuration of the imaging system are indicated in bold

Table 6 Standard deviation of the yield estimation error using different combinations of images captured from only one side of the tree rows

Model	Actual yield	All three cameras		Bottom & middle cameras only		Bottom & top cameras only		Middle & top cameras only	
		Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error
1	10678	5	164	4	189	5	158	11	149
2	9858	7	146	12	180	8	145	5	165
3	10712	-8	171	-3	193	-7	170	-7	185
4	10828	0	166	5	178	4	159	7	170
5	11078	1	169	-4	175	-1	164	-1	182
6	12066	-10	137	-13	162	-10	133	-11	153
7	11327	-8	181	-2	176	-5	171	-5	172
8	11895	4	185	7	216	4	184	8	185
9	11742	-11	168	-9	184	-9	158	-8	175
10	10409	3	153	-1	174	4	147	2	168

Results for the July 2014 dataset (43 trees). Each row corresponds to the validation results of a model calibrated with a different subset of 20 trees. The highest errors for each configuration of the imaging system are indicated in bold

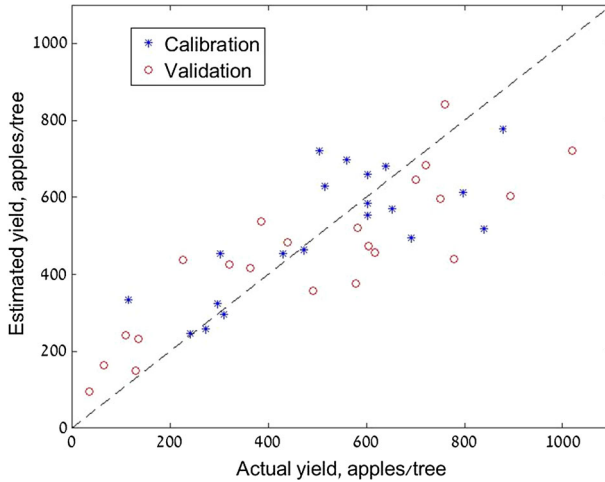


Fig. 5 Actual yield Versus yield estimated using all images from both sides of the row (July 2014 dataset, Model 3 in Table 5)

Table 7 Mean and standard deviation of the yield estimation error using different combinations of images captured from both sides of the tree rows

Model	Actual yield	All three cameras		Bottom & middle cameras only		Bottom & top cameras only		Middle & top cameras only	
		Relative error yield %	Std of yield error	Relative error yield %	Std of yield error	Relative error yield %	Std of yield error	Relative error yield %	Std of yield error
1	10678	10	130	10	141	10	130	14	140
2	9858	9	137	17	157	9	136	7	155
3	10712	-8	171	-5	194	-6	167	-3	179
4	10828	10	159	10	169	10	159	9	173
5	11078	-1	160	-6	181	0	159	-1	173
6	12066	-15	143	-21	160	-13	140	-17	151
7	11327	4	179	2	191	3	177	-2	166
8	11895	7	190	7	195	4	178	8	187
9	11742	-5	156	-4	183	-5	156	-1	165
10	10409	-5	159	-6	165	-5	157	-5	171

Results for the August 2014 dataset (43 trees). Each row corresponds to the validation results of a model calibrated with a different subset of 20 trees. The highest errors for each configuration of the imaging system are indicated in bold

In terms of yield prediction, the results are encouraging but further improvements are required to reduce the estimation error at the single-tree level. Two factors contribute to the relatively large errors at the single-tree level: the simplicity of the models used and the significant overlap between neighboring trees. In terms of modeling, the multi-linear models used in this work are over-simplistic and more accurate models could probably be

Table 8 Mean and standard deviation of the yield estimation error using different combinations of images captured from only one side of the tree rows

Model	Actual yield	All three cameras		Bottom & middle cameras only		Bottom & top cameras only		Middle & top cameras only	
		Relative error yield %	Std of yield error	Relative error yield %	Std of yield error	Relative error yield %	Std of yield error	Relative error yield %	Std of yield error
1	10678	6	164	5	175	5	164	13	143
2	9858	9	151	13	174	9	149	7	178
3	10712	-3	170	-3	186	-3	168	1	189
4	10828	9	157	11	162	9	157	8	182
5	11078	-1	159	-4	184	0	153	-4	200
6	12066	-15	135	-19	159	-13	132	-20	158
7	11327	-6	178	-5	172	-7	172	-10	193
8	11895	-1	185	-1	189	-2	175	2	174
9	11742	-20	146	-21	165	-16	149	-13	178
10	10409	2	156	3	170	2	156	1	171

Results for the August 2014 dataset (43 trees). Each row corresponds to the validation results of a model calibrated with a different subset of 20 trees. The highest errors for each configuration of the imaging system are indicated in bold

Table 9 Mean and standard deviation of the yield estimation error using different combinations of images (captured from only one side of the tree rows). Results for the 2015 dataset (32 trees)

Model	Actual yield	All three cameras		Bottom & middle cameras only		Bottom & top cameras only		Middle & top cameras only	
		Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error	Relative error yield (%)	Std of yield error
1	20851	-5	207	-5	210	-9	224	-4	197
2	20970	-3	218	-2	227	-3	219	-6	192
3	19797	14	193	9	170	15	201	12	179
4	20619	-6	191	-6	200	-8	195	-6	191
5	20674	-12	226	-11	218	-11	221	-9	170
6	19787	-9	223	-4	188	-6	219	-7	223
7	20101	-3	198	-3	205	-3	199	-4	176
8	21111	-13	182	-13	190	-13	185	-12	179
9	21050	-4	200	-4	207	-4	203	-9	184
10	19927	2	191	7	195	0	193	0	189

Each row corresponds to the validation results of a model calibrated with a different subset of 20 trees. The highest errors for each configuration of the imaging system are indicated in bold

obtained using more flexible, non-linear models such as neural networks. However, such models would require a much larger dataset for calibration. In its present state, the method requires that the grower provides the actual yield of 20 trees, which does not seem

Table 10 Comparison between the results obtained with the current procedure and the procedures described in Cohen et al. (2014) and Linker and Kelman (2015)

		Dataset	
		2013	July 2014
Number of apples according to visual inspection		6713	18288
Current procedure	Relative estimation error, %	0.35	0.79
	Std of estimation error, apples	6.96	15.39
Cohen et al. 2014	Relative estimation error, %	0.45	1.21
	Std of estimation error, apples	10.52	16.8
Linker and Kelman 2015	Relative estimation error, %	0.00	−0.86
	Std of estimation error, apples	10.78	17.02

Std standard deviation

excessive. The use of more flexible models (prone to overfitting) would require that the grower provides the actual yield for a much larger number of trees, which would not be realistic. However, the limited performance of the multi-linear models is also probably partly due to the significant overlap that exists between neighboring trees. While such overlap is virtually non-existent in images acquired in young orchards with small trees, such as used by Wang et al. (2012), in mature orchards such overlap is unavoidable and further work will be required to deal adequately with this issue. Finally, since the images captured by neighboring cameras overlap, performing registration of the images and analyzing a single composite image for each side of the tree would probably improve the results. Due to the three-dimensional architecture of the trees and the lack of clear “anchor points” for registration (corners, lines, etc.) image registration is far from trivial and was not attempted in the present study.

Conclusion

Yield estimation using color imaging remains a challenging task. The approach presented in this work is simple and requires ground truth data for only 20 trees. Although the actual yield estimation error depended on the choice of the calibration trees, in the present study, the total yield estimates were almost always within $\pm 10\%$ of the actual yield. However, the standard deviation of the yield estimation errors corresponded to more than 25 % of the average tree yield, indicating that improvements are still needed in order to achieve accurate yield estimation at the single-tree level. Furthermore, the relationship between the number of visible (or located) apples and tree yield was specific to each dataset, meaning that in practice the grower would be required to provide ground-truth data for the calibration trees. For practical applications, unless one were able to devise a one-fits-all relationship, one of the main challenges will be to develop models with enhanced performance while maintaining the calibration set reasonably small.

Acknowledgments This work was supported by the Israel Plants Production & Marketing Board—Apple Division. The author wishes to thank the team at the Matityahu Research Station.

References

- Atherton, T. J., & Kerbyson, D. J. (1999). Size invariant circle detection. *Image and Vision Computing*, *17*, 795–803.
- Cohen, O., Linker, R., Payne, A. & Walsh, K. (2014). Detection of fruit in canopy night-time images: Two case studies with apple and mango. In: Proceedings of the 12th International Conference on Precision Agriculture. ISPA, Monticello, IL, USA, Paper 1420. <https://www.ispag.org/icpa/Proceedings>
- Gongal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: a review. *Computers and Electronics in Agriculture*, *116*, 8–19.
- Jiménez, A. R., Ceres, R., & Pons, J. L. (2000). A survey of computer vision methods for locating fruit on trees. *Transactions of the ASAE*, *43*, 1911–1920.
- Kapach, K., Barnea, E., Mairon, R., Edan, Y., & Ben Shahaar, O. (2012). Computer vision for fruit harvesting robots—state of the art and challenges ahead. *International Journal of Computer Vision and Robotics*, *3*, 4–34.
- Kurtulmus, F., Lee, W. S., & Vardar, A. (2011). Green citrus detection using ‘eigenfruit’, color and circular gabor texture features under natural outdoor conditions. *Computer and Electronics in Agriculture*, *78*, 140–149.
- Linker, R., Cohen, O., & Naor, A. (2012). Determination of the number of green apples in RGB images recorded in orchards. *Computers and Electronics in Agriculture*, *81*, 45–57.
- Linker, R., & Kelman, E. (2015). Apple detection in nighttime tree images using the geometry of light patches around highlights. *Computers and Electronics in Agriculture*, *114*, 154–162.
- Linker, R., Kelman, E., & Cohen, O. (2015). Estimation of apple orchard yield using night time imaging. In *Proceedings of the 10th European Conference on Precision Agriculture, 12–16 July 2015, Beit Dagan, Israel*.
- Payne, A., Walsh, K. B., Subedi, P., & Jarvis, D. (2013). Estimation of mango crop yield using image analysis–segmentation method. *Computers and Electronics in Agriculture*, *91*, 57–64.
- Payne, A., Walsh, K. B., Subedi, P., & Jarvis, D. (2014). Estimating mango crop yield using image analysis using fruit at ‘stone hardening’ stage and night time imaging. *Computers and Electronics in Agriculture*, *100*, 160–167.
- Pla, F., Juste, F., & Ferri, F. (1993). Feature extraction of spherical objects in image analysis: an application to robotic citrus harvesting. *Computers and Electronics in Agriculture*, *8*, 57–72.
- Silwal, A., Gongal, A., & Karkee, M. (2014). Apple identification in field environment with over-the-row machine vision system. *Agriculture Engineering International: CIGR Journal*, *6*, 66–75.
- Sites, A. & Delwiche, M.J. (1988). Computer vision to locate fruit on a tree. Paper No. 85-3039, ASAE, St Joseph, MI, USA.
- Stajanko, D., Lakota, K., & Hocevar, M. (2004). Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. *Computers and Electronics in Agriculture*, *42*, 31–42.
- Wang, Q., Nuske, S., Bergerman, M. & Singh, S. (2012). Automated crop yield estimation for apple orchards. In: Proceedings of the 13th International Symposium on Experimental Robotics. Quebec City, Canada. In: Desai, J.P., Dudek, G., Khatib, O., Kumar, V(Eds). pp: 745–758
- Zhao, J., Tow, J. & Katupitiya, J. (2005). On-tree fruit recognition using texture properties and color data. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, (pp. 263–268).
- Zhou, R., Damerow, L., Sun, Y., & Blanke, M. M. (2012). Using colour features of cv. ‘Gala’ apple fruits in an orchard in image processing to predict yield. *Precision Agriculture*, *13*, 568–580.