Automated Visual Yield Estimation in Vineyards

Stephen Nuske, Kyle Wilshusen, Supreeth Achar, Luke Yoder, Srinivasa Narasimhan, and Sanjiv Singh
Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213
e-mail: nuske@cmu.edu

Received 5 September 2013; accepted 30 June 2014

We present a vision system that automatically predicts yield in vineyards accurately and with high resolution. Yield estimation traditionally requires tedious hand measurement, which is destructive, sparse in sampling, and inaccurate. Our method is efficient, high-resolution, and it is the first such system evaluated in realistic experimentation over several years and hundreds of vines spread over several acres of different vineyards. Other existing research is limited to small test sets of 10 vines or less, or just isolated grape clusters, with tightly controlled image acquisition and with artificially induced yield distributions. The system incorporates cameras and illumination mounted on a vehicle driving through the vineyard. We process images by exploiting the three prominent visual cues of texture, color, and shape into a strong classifier that detects berries even when they are of similar color to the vine leaves. We introduce methods to maximize the spatial and the overall accuracy of the yield estimates by optimizing the relationship between image measurements and yield. Our experimentation is conducted over four growing seasons in several wine and table-grape vineyards. These are the first such results from experimentation that is sufficiently sized for fair evaluation against true yield variation and real-world imaging conditions from a moving vehicle. Analysis of the results demonstrates yield estimates that capture up to 75% of spatial yield variance and with an average error between 3% and 11% of total yield.

1. INTRODUCTION

Yield predictions in vineyards are important for managing vines to optimize growth and eventual fruit quality. For instance, if an overly large crop is forecast, fruit may be removed during the season to achieve certain fruit quality goals. This practice of crop thinning is much more effective when based on an accurate yield estimate. Yield forecasts also prepare a grower for the harvest operation, for shipping their crop, storing their crop, and also selling their crop on the market. Typical yield predictions are performed using knowledge of historical yields and weather patterns along with measurements manually taken in the field. The current industry practice for predicting harvest yield is labor-intensive, expensive, inaccurate, spatially sparse, destructive, and riddled with subjective inputs. Typically, the process for yield prediction is for workers to sample a certain percentage of the vineyard and extrapolate these measurements to the entire vineyard. Agronomic studies have established that large spatial variability in vineyards across multiple countries and growing conditions exists (Taylor, Tisseyre, Bramley, Reid, & Stafford, 2005). However, the sample size is often too small in comparison to the spatial variability across a vineyard, and as a result the yield predictions are inaccurate and spatially coarse.

[Author list amended on 23 September 2014 after error in first online publication on 11 August 2014: The following authors have been added: Srinivasa Narasimhan, Robotics Institute, Carnegie Mellon University.]

There is a gap between the methods available to predict the yield in a vineyard and the needs of vineyard managers to make informed decisions for their vineyard operations. Using carefully designed illumination and a camera system (see Figure 1) paired with novel algorithms that automatically detect the fruit within the imagery, our method can make dense predictions of harvest yield efficiently and nondestructively. We overcome difficulties in imaging and design our camera and illumination setup to optimize for low motion blur, increased depth-of-focus, and low illumination power for fast-recycle times permitting high-frame rates. This design maintains high image quality at high vehicle velocities and enables deployment at large scales, features overlooked in many existing visual yield estimation studies. Our results demonstrate that we can automatically detect and count grapes to forecast yield efficiently with high precision and accuracy.

We also overcome the challenges distinguishing the grape berries from similarly colored leaves, in nonuniform lighting, and variable scales due to variable berry size or variable distance of the fruit from the camera. Our approach can be distinguished from a number of existing studies on the detection of grapes, which are based upon one of three different types of visual cues of grape berries appearance: either color (Diago et al., 2012; Dunn & Martin, 2004), shape (Rabatel & Guizard, 2007), or texture (Grossetete et al., 2012). Each of the three cues has advantages in certain conditions and, all have some limitations in others. Color on its own is not suitable for distinguishing green grapes on a

background of green leaves. The grape shape can be difficult to identify in cluttered images with leaves and other spurious contours. Grape texture can be less distinguishing under certain illumination conditions.

We take a different approach. Rather than focusing on a single cue, we develop a robust and versatile algorithm to detect grape berries under a variety of conditions by exploiting all three main visual cues that grape berries have. Our approach is to detect candidate hypotheses of where grapes may be located in images using a shape transform, and then classify these candidate locations using texture and color descriptions within a robust classification framework that exploits all the visual cues.

We develop a model that relates image measurements to yield and optimize the model for two key goals:

1. accurate estimation of the spatial distribution of yield,
2. accurate estimation of the overall yield.

Preliminary results of our approach were reported in Nuske, Achar, Bates, Narasimhan, & Singh (2011) and Nuske, Gupta, Narasihman, & Singh (2012). We extend our prior work in the following ways:

- Present a study of three different visual texture descriptions and evaluate each on a variety of datasets (Sections 3.2 and 5.3.2).
- Introduce a novel algorithm for berry keypoint detection that is invariant to berry size (Section 3.1.2).
- Expand our model relating image measurements to yield predictions and demonstrate how to optimize for accuracy (Section 4).
- Report on field experiments now spanning four growing seasons of multiple varieties and growing systems, totaling tens of acres of vines, all validated with true yield measurements (Section 5).

The experimental results are the first visual vineyard yield estimation results that are collected over hundreds of vines and also encompassing multiple growing seasons. This is a substantially larger scale than other existing work—by comparison, the next largest study in terms of dataset size was a study using a static camera of 10 vines (Diago et al., 2012). Their study artificially increased the size of the dataset by defruiting vines sequentially, which increases the distribution of yields but does not capture the true variance in leaf and vine occlusions and cluster-to-cluster occlusions. Our system stands apart from existing approaches in the complete design that solves all of the true imaging and visual detection challenges. Our method is also the first to present a way to eliminate the double-counting problem from overlapping imagery and also the challenge of geometrically referencing the measurements by estimating camera position along the row. Unlike other work, our experiments are conducted from a moving vehicle in a completely un-isolated fashion such that we replicate all issues that would be faced by a real deployment. No other study so far has evaluated all the challenges and variables that must be considered in a rigorous study, including variance in visual detection performance, differences in berry-size, detection of green juvenile berries, the variable distance of the camera to the fruit zone, estimating location along the vineyard rows, and the problems of leaf/vine occlusion and cluster-on-cluster occlusion.

2. RELATED WORK

Viticultural studies (Clingeleffer, Dunn, Krstic, & Martin, 2001) have revealed that current industry practices to forecast yield are inaccurate because of sampling approaches that tend to adjust toward historical yields and include subjective inputs. The calculation of final cluster weight from weights at véraison (i.e., onset of color development) uses fixed multipliers from historic measurements (Wolpert & Vilas, 1992). The multipliers are biased toward healthier vines, thus discriminating against missing or weak vines, and multipliers for cluster weights vary widely by vineyard, season, and variety. In general, the approaches have required manual samples of mean berry size, mean cluster count, or mean cluster size (Serrano, Roussel, Gontier, Dufourcq, 2005), or a combination of the three, and then an extrapolation to yield across a vineyard block.

Sensor-based technologies using trellis tension monitors, multispectral sensors, terahertz-wave imaging, and visible-light image processing have all been proposed for yield estimation in vineyards. A dynamic yield estimation system based on trellis tension monitors has been demonstrated (Blom & Tarara, 2009), but it requires a permanent infrastructure to be installed. Further, variation in yield is not the only cause of variations in trellis tension. Ambient temperature and vine size also affect trellis tension, resulting in a loss of accuracy when relating tension to yield. Information obtained from multispectral images has been used to forecast yields with good results, but it is limited to vineyards with uniformity requirements (Martinez-Casasnovas & Bordes, 2005). Related multispectral imagine work has been applied to detecting almonds with a combination from RGB and ir images (Hung, Nieto, Taylor, Underwood, & Sukkarieh, 2013a). A proof of concept study by Federici, Wample, Rodriguez, & Mukherjee (2009) has shown that terahertz imaging can detect the curved surfaces of grapes, and it also has the potential to detect through an occluding thin canopy. The challenge for this approach is to achieve fast scan rates to be able to deploy the scanner on a mobile platform while also having sufficient wavelength/resolution to both detect as well as penetrate.

Conventional RGB camera imagery has been proposed in a number of works as a means for fruit detection and yield estimation. Jimenez et al. (Hung, Underwood, Nieto, & Sukkarieh, 2000) provide a summary of fruit detection work. Singh et al. (2010), Zhou, Damerow, Sun, and Blanke (2012), Hung, Underwood, Nieto, & Sukkarieh (2013b), and

Tabb, Peterson, and Park (2006) present methods for detecting and classifying fruit pixels in imagery collected in apple orchards. Yang, Dickinson, Wu, and Lang (2007) demonstrated a similar approach to recognize tomatoes for harvesting. A method presented by Berenstein, Shahar, Shapiro, and Edan (2010) used a combination of color and edge features for a coarse estimate of where grape clusters and leaves reside in the images for the purpose of precision spraying. A new method that uses 3D image reconstructions for fruit, leaf, and stem detection, demonstrated by Dey, Mummert, and Sukthankar (2012), could classify fruit, leaves, and shoots based on 3D reconstructions generated from image sequences, which, unlike our work, is sensitive to slight wind while imaging.

The most closely related works to ours in vineyard yield estimation with image analysis can be divided by which type of visual feature is used to identify the grapes within images: color, shape, or texture.

Color discrimination has been demonstration by Dunn & Martin (2004). This approach was attempted on Shiraz post-véraison (red-grapes with full color development) in short row segments. A similar color-based method that detects fruit close to harvest after the onset of color-development is presented by Diago et al. (2012) and Marden, Liu, & Whitty (2013). These simple color-based methods are not applicable for the majority of real-world examples where the fruit appears over a background of similarly colored leaves, as is the case in green grape varieties and in all varieties before véraison. The contour and shape of the berry is another method used to detect the fruit (Rabatel & Guizard, 2007). The contour is more broadly applicable to detect green grapes on green leaves, although partially visible grapes and cluttered images filled with a variety of contours on leaves and stems make it difficult to completely distinguish the grapes using contour and shape alone. The visual texture on the surface of the grape, especially when imaged with an illumination source, is also a viable cue for detection as exploited by Grossetete et al. (2012), who demonstrate a hand-held device that can be used to measure the size of isolated clusters. Similar approaches detect the shading on apples (Wang, Nuske, Bergerman, & Singh, 2012) and oranges (Swanson et al., 2010). All three types of cues—color, shape, or texture—have seen some success in detecting grapes, although none is convincing as a sole cue for detecting the fruit. In our work, we demonstrate how to exploit all three of these cues to create a strong classification algorithm that separates berries from leaves and surpasses the ability that can be achieved with one of these visual cues alone.

Finally, in performing a comparison of our work to all prior visual yield estimation studies, the most significant differences are in terms of the scale of deployment and also rigorous evaluation of all real-world considerations. The only other works performed in which experiments included measurements evaluated against total vine yield are Diago et al. (2012) and Dunn & Martin (2004). These examples used limited sets of vines in their experimentation—10 vines and 1 vine, respectively. Both of these examples used defruiting to artificially increase the size of the datasets, which caused their experimental dataset to not be representative of the true variance in vine occlusions and cluster-to-cluster occlusions. Also, they did not consider individual berries, they only considered raw pixel count, which is sensitive to the berry size, the distance from the vine, the focal length, image resolution, as well as berry count (the desired measurement). Also, they used a white sheet or white wall backdrop to block out background disruptions. Further, to draw more differences, our work is unique in that it is a system that can be deployed from a moving vehicle though the use of automatic image registration to deal with double counting or undercounting of fruit. We also consider detection of green juvenile fruit on a green leaf background. There are other works that do look at detecting green berries among green leaves, such as Grossetete et al. (2012) and Rabatel & Guizard (2007), but their experiments are restricted to a small set of isolated clusters and do not evaluate total vine yield. Our work is the first system and study of all these real-world considerations, and evaluation is performed over hundreds of vines and four growing seasons.

3. BERRY DETECTION

We deploy a sideways-facing camera on a small vineyard utility vehicle; see an illustration in Figure 6. The images capture the vines and are processed with our algorithm to detect and estimate the crop yield.

Yield in a vineyard is a combination of the following crop components: the number of clusters, the number of berries per cluster, and the berry weight. Given the variance in yield of a set of vines, this variance can be broken down into the three yield components as follows (Clingeleffer et al., 2001):

1. Variance in number of clusters per vine—contributes approximately 60% of the total yield variance.
2. Variance in number of berries per cluster—contributes approximately 30% of the total yield variance.
3. Variance in berry size—contributes approximately 10% of the total yield variance.

These three yield components combine to describe all 100% of variance in harvest yield. Current practice is to take samples of each of these components to compute an average and compute the final yield. We take an approach to estimate the first two of these items together in one measurement—that of the number of berries per vine. The reason for this is that it is difficult, especially late in the season, to delineate the boundaries of clusters within images. However, it is possible to count the total number of berries seen, hence combining the two components—number of clusters per

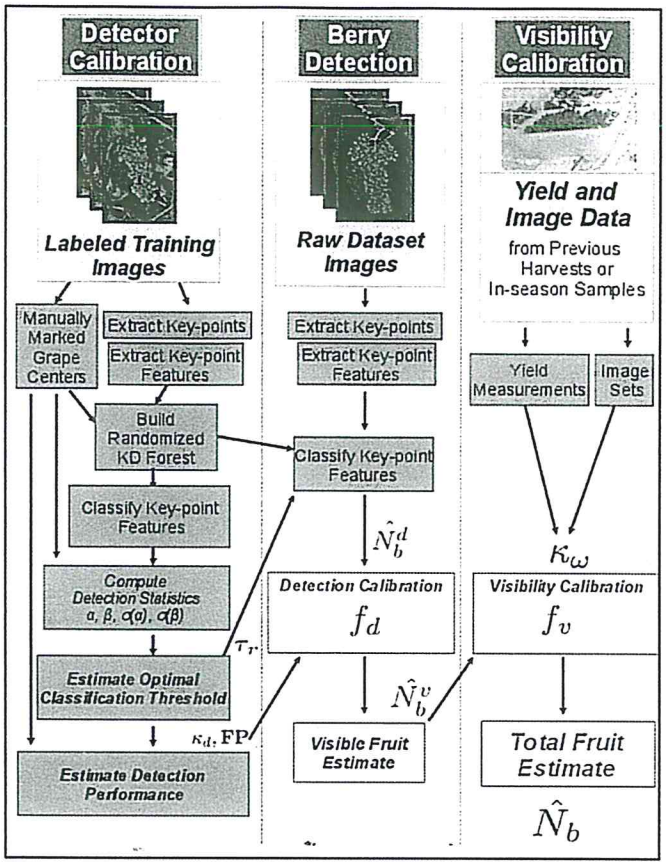


Figure 2. Our yield estimation process is divided into three facets: detector-calibration to the left, berry-detection shown in the center, and visibility-calibration to the right. The first facet is detector calibration, which uses a set of training images (we use a set of 10 images), where grape locations are manually marked. The training images are processed to find grape keypoints (Section 3.1), features are extracted at these keypoints (Section 3.2), and features are labeled as grape or nongrape features based on their proximity to manually marked locations. The labeled feature-descriptors are then stored in a randomized KD-forest. Detection performance is then characterized using the training set; this allows for the identification of the optimal voting ratio (Section 4.2) and also the computation of a detection calibration function to account for trends in detection performance (Section 4.2.1). After training, the core part of our approach is berry detection. Here, grape keypoints are found in each image, and then features are extracted at the keypoints and classified as grape and nongrape using the randomized KD-forest with the optimal voting ratio [Eq. (1)]. After detections are computed, the output is corrected using the function in Eq. (10) to estimate visible fruit. The final part of our process is visibility calibration, which uses yield data (either from prior harvest or sparse in-season samples) to form a calibration function [Eq. (17)] to estimate total yield from our dense image measurements.

vine and berries per cluster—into one measurement: berries per vine.

Our approach does not attempt to measure berry weight. However, we account for 90% of the harvest yield variation with berries per vine (Clingeleffer et al., 2001). Furthermore, instead of taking a small sample and extrapolating, we aim to estimate nondestructively the specific yield at high resolution across the entire vineyard. Hence, we will not introduce sampling errors into the process.

The overall process is presented in Figure 2, which is split into three phases: berry-detection, detector-calibration, and visibility-calibration.

Our algorithm, as first proposed in Nuske et al. (2011) and expanded here, is split into the following stages:

(1) Detect potential berry locations (Section 3.1).
(2) Classify and group the potential locations that have similar appearance to grape berries (Section 3.2).

3.1. Detecting Image Keypoints—Potential Berry Locations

The first step of our algorithm is to find image keypoints that are potential grape berry locations; see Figure 3(b). These keypoints will later be classified as either *berry* or *not-berry*.

There are two reasons why we detect keypoints. The first is to identify distinct grape-berries, which is important to extract measurements that are invariant to the stage of the berry development. Other work, such as Dunn & Martin (2004) and Diago et al. (2012), detect grape pixels only. However, the number of pixels that belong to fruit will change as the grapes get larger, and therefore the measurements will not be invariant to berry development. The second reason is to reduce computation by only considering a set of keypoints; this reduces the amount of the image that is subsequently processed.

We present two keypoint detection algorithms. The first is a radial symmetry algorithm [introduced by Loy & Zelinsky (2003)] that uses the circular shape of the berry as a cue for detection. The second is a novel maximal point detection algorithm that searches for the maximal point of shading in the center of grapes that have been illuminated by a flash. The two algorithms are introduced in the following two subsections and evaluated later in Section 5.3.1.

3.1.1. Radial Symmetry Transform

One approach we use is to search for points that have peaks in a radial symmetry transform from Loy & Zelinsky (2003). The radial symmetry transform requires us to know the radii of the berries as seen in the image ahead of time. The berry radii (in pixels) are dependent on the focal length of the camera, the actual berry size, and the distance from the camera. The focal length is kept fixed in our tests and the vehicle maintains a relatively constant distance from the vines. There is still variation in the radius in which the berries appear in the image from differing berry sizes and also some variation in location within the vine. We account for this variation by searching for radially symmetric points over a range of possible radii, finding points that exhibit the most radial symmetry.

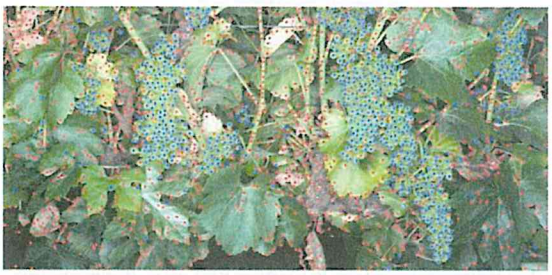
3.1.2. Invariant Maximal Detector

We have developed an alternative keypoint detection algorithm that searches for peaks in intensity corresponding to grape centers. These peaks are found by finding local maxima in three-by-three kernels, and the peaks are validated by an iterative growing procedure that identifies peaks with symmetrical shading created by a flash. We continue in this section to describe this procedure in more detail.

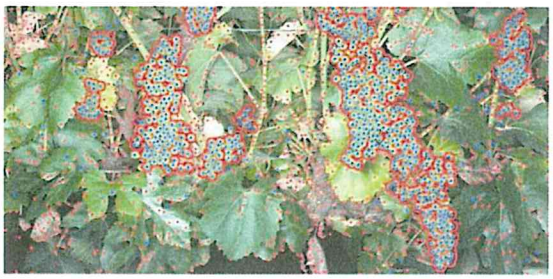
Lighting upon grapes is controlled, as flashes are positioned parallel to the camera's optical axis to illuminate the grapes (discussed in Section 5.1). This leads to grapes that have a strong specular reflectance at the center of the

(a) Input image

(b) Detect Potential Berry Key-points

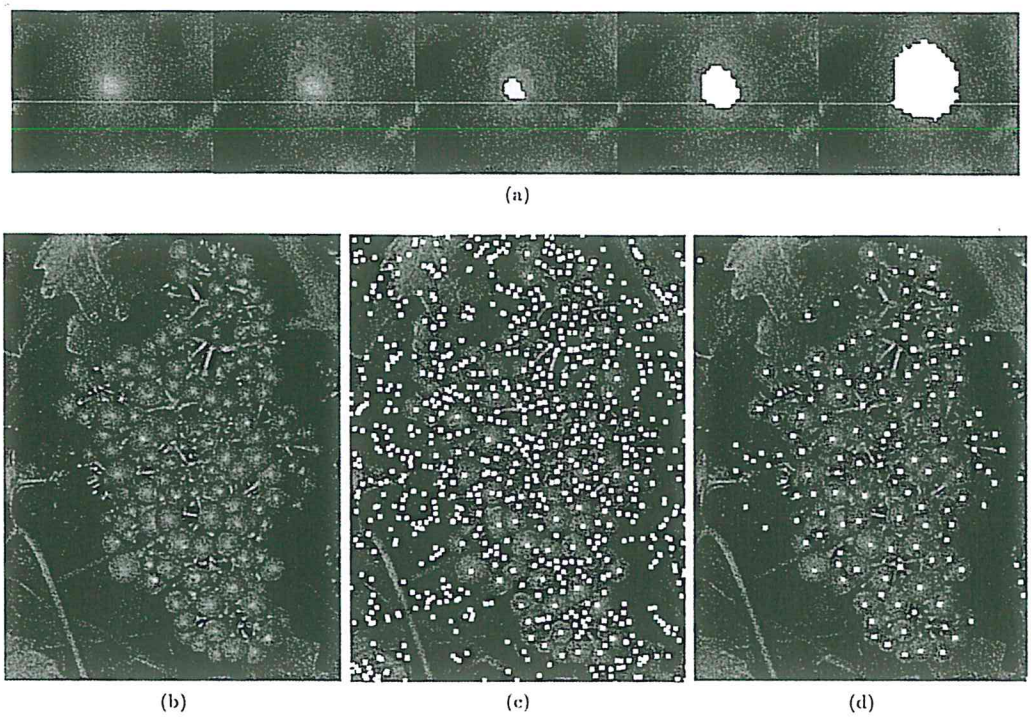


(c) Identify Locations with Similar Appearance to Grape Berries



(d) Group Neighboring Berries

Figure 3. Example images showing the functioning of our visual berry detection algorithm. Input image is seen in (a). Keypoints, shown in (b), are the potential berry locations. In (c), points marked in blue have been classified as having an appearance similar to a berry, and in (d) berries that neighbor other berries are clustered together.



(a)

(b) (c) (d)

Figure 4. Examples demonstrating the invariant maximal initial keypoint detector. Part (a) demonstrates the seeded growth into regions of decreasing intensity over many iterations. Parts (b), (c), and (d) show the input image, the initial maximal points, and the final keypoints that pass the seeded growing algorithm, respectively.

berry. From this point of specular reflectance, pixel intensity decreases steadily toward the edges of the grape. If we can find this point of specular reflectance, we can find a set of keypoints that are potential berries.

There are other algorithms in the literature that detect the shading on fruit produced by a flash, such as Grossetete et al. (2012), Swanson et al. (2010), and our previous work (Wang et al., 2012). However, these approaches do not consider fruit of substantially different sizes, and they would require some reconfiguration to function on fruit of different size, or images with different resolution.

We present an approach that does not need a size parameter as input and can function on berries of substantially different sizes, from 10 pixels in diameter up to well over 100 pixels in diameter.

The identification of potential grape centers in images is as follows. Image noise is first eliminated through a Gaussian pyramid downsampling operation. Then, an initial set of local regional max within their immediate neighborhood is identified; see Figure 4(c).

Filtering of this set is then achieved through an iterative region growing method, operating on each regional maxima. Regional maxima serve as the seed points to region growing. Through an iterative process, shown in Figure 4, these seed points are grown and evaluated for symmetry and decreasing intensity profile.

At each iteration, pixels of decreasing intensity adjacent to the regional max are included in the region. This adds a *ring* of pixels darker than the maxima but brighter than a fraction of the maxima intensity (we use 85% as a threshold for each ring). Once the new "ring" of darker pixels has finished growing, a new iteration begins by growing a new ring of pixels, which includes darker pixels than the previous ring but brighter than a fraction of the previous ring's intensity. Each image point can only belong to one grown region. This ensures that after an initial region is grown, an adjacent region cannot contain the same pixels, eliminating overlapping maxima. This process repeats until a set number of rings has been added (three intensity levels) or one of the following shape criteria has failed:

- First, the region's centroid should be within a range of the region's maxima point.
- Second, the shape of the region should not become elongated.

A region that has three levels of decreasing intensity and passes these two shape criteria is considered a possible grape keypoint. Finally, a mask image of pixels that belong to regions is maintained, such that each image point can only belong to one grown region. This ensures that after an initial region is grown, an adjacent region cannot contain the same pixels, eliminating overlapping maxima.

3.2. Classifying Appearance of Candidate Keypoints as Berry/Not-berry

The next stage in our algorithm is to classify the detected keypoints into *grapes* or *not-grapes*; see Figure 3(c). We take an image patch around each detected center and compute a combination of both color and texture features from that image patch. For the color features, we form a six-dimensional vector from the three RGB channels and the three L*a*b color channels.

Our initial work in Nuske et al. (2011) presented the use of Gabor filters to classify the grape's texture. In this paper, we evaluate three different texture features from three broad classes of features, and we will study the applicability of each type of feature to a set of different imaging conditions. In particular, we choose three of the broad categories:

1. Filter banks: used as far back as the 1980s (Laws, 1980) in image texture classification tasks. We use Gabor filters with four scales and six orientations and combine to form a 24-dimensional feature vector, as per the original work (Nuske et al., 2011).
2. Local texture features: proposed in the late 1990s, they describe histograms of gradients (HoG), or similar variations, in a support region around an image location. Here, we use a SIFT descriptor (Lowe, 2004) computed in a support region centered at our candidate centers. Since we are using our specific berry keypoint detection algorithms as presented in the previous section, we do not use the SIFT keypoint detection step. The raw descriptors total 128 dimensions, which we decimate into its most discriminative axes using PCA into a 32-dimensional vector.
3. Binary relations: a recent class of image features that compile sets of pairwise binary intensity relations between pixels or regions surrounding an image location. We use the Fast Retinal Descriptor (FREAK) (Alahi, Ortiz, & Vandergheynst, 2012), which is designed to mimic the human retinal construction and is noted to achieve higher performance rates than other recent binary features [BRIEF (Calonder, Lepetit, Strecha, & Fua, 2010)]. Similar to the process used for SIFT descriptors, we perform dimensionality reduction from the 200 raw dimensions into its most discriminative axes using PCA into a 32-dimensional vector.

We study these three types of features in the context of three different illumination conditions with which we have collected image data in vineyards: natural illumination, flash illumination, and cross-polarized flash illumination. Results of the study are presented in Section 5.

We concatenate the color and texture features to form a set of candidate image features, I_c, computed at the berry keypoint candidates. To classify the candidate features, we use an *a priori* constructed randomized KD-forest (Lepetit, Lagger, & Fua, 2005) from a set of training samples, $\mathbf{T}$, extracted from a subset of images. We manually define berry centers in the training images that correspond to the positive examples of the appearance of berries, $\mathbf{T}_p$. For negative samples, $\mathbf{T}_n$, we compute features at our detected keypoints that do not align with a manually defined berry center.

Then for each candidate feature $I_c(i)$ we extract a set of nearest nodes $\mathbf{T}_c(\mathbf{i})$, from the randomized KD-forest. We then vote on the class of each interest point by computing a ratio of the number of positively labeled features, $T_r = \frac{|\mathbf{T}_c(\mathbf{i}) \in \mathbf{T}_p|}{|\mathbf{T}_c(\mathbf{i})|}$, to the total size of the candidate feature set. We define a threshold, τ_r, applied to the voting ratio, which is adjusted to control how conservative or liberal the classification is. Our resulting set of classified berries is $b_d \in I_c$, where

$$T_r > \tau_r. \tag{1}$$

We group these detected berries into clusters less than a threshold in distance from each other. We remove clusters of detections that are smaller than an area threshold, which removes spurious detections; see Figure 3(d).

3.3. Combining Berry Detections in Image Sequences

Since we are collecting data from a moving platform and wish to produce yield estimates that are assigned to specific locations in the vineyard, it is not sufficient merely to be able to detect grapes in an image. First, we need to have state information for the data collection vehicle to be able to back-project the grape detections from each image to a position along the vineyard row. Second, we need to use the registered grape locations to avoid double counting of grapes between consecutive images.

To gain state estimation for the data collection vehicle, we use a second camera facing backward from the vehicle angled at approximately 45° down toward the ground for positioning. The downward-facing camera is a stereo pair and we use a visual odometry algorithm (Kitt, Geiger, & Lategahn, 2010) to track the position of the vehicle as it moves along the row. With state information data for the vehicle, individual grapes can be registered to locations on the local fruit wall. The problem is that there can be significant overlap between consecutive images. This overlap needs to be accounted for during grape registration to prevent a single grape from being counted multiple times.

The process to identify individual grapes is as follows. First, the detections from each image from a vineyard row are back-projected onto the local fruit wall. The entire length of the row is then partitioned into short segments (around 0.5 m in length). A segment that contains projected grape detections from more than one image is a region in which camera views overlapped. The detections from the image with the most detections lying in the segment are retained, and detections in the segment from the other images are discarded. This heuristic is used to choose the image where the effect of occlusions was minimum. Taking the maximum image measurement rather than attempting to merge all image measurements together avoids the issue of attempting to do fine registration while fruit is moving from wind or the vehicle pulling at the vine. Each segment now has a visible berry count associated with it, and this can be aggregated over rows, vines, or sets of vines.

4. RELATING IMAGE-BASED MEASUREMENTS TO YIELD PREDICTIONS

The previous section describes how to detect berries within images. This section describes how to take these image measurements and form an estimate of fruit yield; Figure 2 explains the procedure.

Viticulturists have long studied the process of predicting the size of the harvest yield, and they have developed models of the various yield components (Clingeleffer et al., 2001). In the most basic form, the weight of the harvest (W_h) can be expressed as a product of the number of berries (N_b) and the mean weight of the berries (W_b),

$$W_h = N_b W_b. \tag{2}$$

Our approach to predict the yield at harvest time is focused on measuring the number of berries (N_b). The number of berries is stable shortly after berry set (once the period known as *shatter* is passed) and it accounts for 90% of the variation in yield (Clingeleffer et al., 2001); varying berry weight is responsible for the remaining 10%. Our berry count forms a part of a yield forecasting function, $f(\cdot)$, which outputs an estimate ($\hat{N}_b$) of the actual berry count:

$$\hat{N}_b = f(N_b^d). \tag{3}$$

4.1. Visible Berries and Estimating Self-occlusions

In our prior work (Nuske et al., 2011) we used the visible berry count as a prediction of the cluster size, assuming the detected berry count is proportional to the total berry count:

$$N_b \propto N_b^d. \tag{4}$$

In the Results section (Section 5.4), we study the visible berry count in controlled experiments, and also, in an attempt to improve the measurement of the occluded berries in a cluster, we propose two potential modifications.

The first alternative measurement we propose is to take the convex hull formed by all the visible berries in the cluster. Assuming the cluster has uniform density and an average thickness of the grape cluster to be D, we multiply the area A to this fixed cluster depth, and we normalize with the average berry radius R_b,

$$N_b \propto D\frac{A}{R_b^2}. \tag{5}$$

The second alternative is to predict the size of a cluster using a 3D ellipsoid model. A grape cluster's volume can be approximated with an ellipsoid cutting off the image plane as an ellipse. We find the best-fit ellipse for the berry center locations with same normalized second central moments. Given the semiaxes of the ellipse in pixels R_1 and R_2, with $R_1 \geq R_2$, the volume of the corresponding ellipsoid would be proportional to the volume occupied by the berries (B_v) in the cluster. Using the average berry radius (R_b) of the cluster, we can calculate the total number of berries occupied by the cluster:

$$\begin{aligned} V_c &\propto \frac{4}{3}\pi r_1 r_2^2, \\ N_b &= V_c \Big/ \left(\frac{4}{3}\pi r_b^3\right). \end{aligned} \tag{6}$$

We study these three approaches to measuring grape cluster size in controlled laboratory tests in the Results section (Section 5.4).

4.2. Optimizing for Yield Estimate Accuracy

Ultimately, the goals are to achieve accurate yield estimates. In this section, we isolate the metrics of interest and present a means to optimize our algorithm to improve spatial yield estimation accuracy.

There are two requisites for $\hat{N}_b$ in terms of accuracy:

1. Accurate estimate of overall yield. Minimizing the error term: e_v.
2. Accurate estimate of spatial yield. Minimizing the error term: e_s.

$$e_v = \frac{\sum(\hat{N}_b^i) - \sum(N_b^i)}{\sum(N_b^i)}, \tag{7}$$

$$e_s = \frac{\Sigma(\hat{N}_b^i - N_b^i)^2}{\Sigma[N_b^i - \mu(N_b^i)]^2}, \tag{8}$$

where N_b^i is the number of berries on the ith vine.

We will revisit details regarding e_v and e_s later, but we begin by introducing a function f that captures the relationship between the detected fruit in the camera image N_b^d to the total fruit N_b.

4.2.1. Detection-Calibration

There are errors in the visual detection process that must be modeled. In Nuske et al. (2011), the performance of the detection algorithm was analyzed to find that the algorithm does not detect some berries visible to the camera, and to a lesser extent there are some occasions when the algorithm falsely reports a berry where there was not one.

We form a measurement of detection using a sample of images where the berry centers have been manually marked, enabling us to measure three metrics: true positives (TP)—the number of berries detected that were actual berries; false positives (FP)—the number of false berry detections; and false negatives (FN)—the number of berries visible in the image that were not detected. Our measure, κ_d, exactly defines the rate of visible berries that are detected by the system,

$$\kappa_d = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{9}$$

We take κ_d, which is the true positive rate, together with the false positives (FP) expressed as a number per linear length of vineyard, and it can be applied to the level of detected fruit N_b^d to estimate the amount of visible fruit:

$$\hat{N}_b^v = f_d(\hat{N}_b^d, \kappa_d, \text{FP}) = \frac{(N_b^d - \text{FP})}{\kappa_d}. \tag{10}$$

Importantly, unlike κ_d, the number of false positives is decoupled from the total number of visible fruit. This decoupled behavior was discovered empirically through our analysis and readily understood and explained, since vines with little or no fruit will induce a similar amount of false positives as vines with larger amounts of fruit.

We use this function to correct for variable detection performance arising from the differing appearance of the vines and variable imaging and lighting configurations.

Referring back to our error, we wish to minimize, e_s, in Eq. (8), which is in fact directly related to the correlation measure $r^2 = 1 - e_s$; the portion of the error term we care most about is the squared error term of the predicted yield compared to the true yield: $\Sigma(\hat{N}_b^i - N_b^i)^2$.

Our model of detected to visible to total fruit is a linear system, and the terms can all be calibrated in closed form regression. Assuming there is sufficiently accurate calibration data, the squared error term will only be nonzero under the presence of some unmodeled variance in our model. It is expected that there is some nonzero variance in our occlusion term κ_ω, where we use the notation for the the standard deviation of this term as σ_{κ_ω}, which reforms the equation as

$$\hat{N}_b = f_v(\hat{N}_b^v, \kappa_\omega, \sigma_{\kappa_\omega}) = \frac{\hat{N}_b^v}{\kappa_\omega + g(\sigma_{\kappa_\omega})}, \tag{11}$$

where $g()$ is a zero-mean Gaussian function. At present, in the absence of a known method to estimate this variance, $\sigma(\kappa_\omega)$, we leave this as an unmodeled error.

However, there is another source of variance, which is the deviation within our detection function $\sigma(f_d(\cdot))$. This variance in the detection function is possible to minimize, which in turn will optimize e_s. Using hand-labeled training images to analyze our berry detection system, we find that there are two different behaviors in the detection function, one in which the mean of the true positive berries output by our system is linearly related to the visible fruit, and another in which we find false positives to be modeled as a variable (β) unrelated to the visible fruit:

$$\text{TP} = N_b^v[\alpha + g(\sigma_{\kappa_d})], \tag{12}$$

$$\text{FP} = \beta[1 + g(\sigma_\beta)]. \tag{13}$$

Thus, we can rewrite our relationship to visible fruit as follows:

$$\hat{N}_b^v = \frac{N_b^d}{\alpha + g(\sigma_\alpha) + \beta[1 + g(\sigma_\beta)]}. \tag{14}$$

$g(\sigma_\alpha)$ and $g(\sigma_\beta)$ are the two sources of variance in the detection function, and in a well-calibrated system they will be directly proportional to the squared error term:

$$\Sigma(\hat{N}_b^i - N_b^i)^2 \propto \frac{\mu(N_b^v)g(\sigma_\alpha)}{\mu(\text{TP})} + \frac{\mu(\text{FP})g(\sigma_\beta)}{\mu(N_b^v)}. \tag{15}$$

We have now identified the terms responsible for spatial accuracy, and we can develop an optimization strategy. The key variable in our berry classification algorithm is the threshold on the randomized KD-forest voting ratio, τ_r from Eq. (1). We use this as the variable to optimize against computing the minimum of our error term:

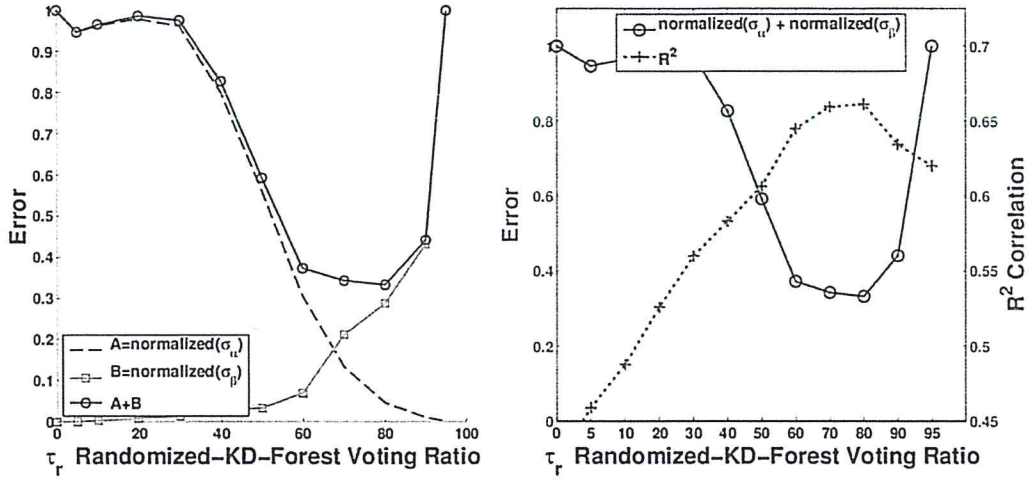
$$\arg\min_{\tau_r} \left(\frac{\mu(N_b^v)g(\sigma_\alpha)}{\mu(\mathbf{TP})} + \frac{\mu(\mathbf{FP})g(\sigma_\beta)}{\mu(N_b^v)} \right). \tag{16}$$

The error term and its components are graphed together in Figure 5. The τ_r parameter is adjusted to balance the variance in the false positives in the system and the variance in true positives, and the optimal point is captured by the error term. The minimal point on the error term curve is used to ultimately reflect the maximal R^2 value against a vineyard dataset. Later in the Results section, this optimization is studied for different texture feature descriptors (Section 5.5).

4.2.2. Visibility Calibration

Once a corrected estimate of visible fruit is derived, we form a second function, $f_v()$, relating the visible estimate of fruit to the total fruit as follows:

$$\hat{N}_b = f_v(\hat{N}_b^v, \kappa_\omega) = \frac{1}{\kappa_\omega}\hat{N}_b^v. \tag{17}$$



(a) Detection metrics (b) Optimize SIFT against R^2

Figure 5. Spatial yield accuracy, R^2, can be maximized by minimizing the detection system's error metric. Specifically, the randomized KD-forest voting threshold τ_r is optimized against normalized σ_α and normalized σ_β from the error term from Eq. (16).

This relationship requires yield data to form an estimate of visibility; κ_v. As discussed in detail in Nuske et al. (2012), the yield data can be collected from previous harvest data or from in-season samples. Although $f_d()$ varies by dataset and imaging type, we find the visibility function $f_v()$ is consistent over seasons for a particular vineyard.

Together the two functions $f_d()$ and $f_v()$ form the two key relationships—from image measurement to visible fruit and from visible fruit to yield prediction; for an overview of the procedure, see Figure 2.

5. RESULTS

Experiments are conducted at four different vineyards over four growing seasons, including the following varietals: Traminette, Riesling, Flame Seedless, Chardonnay. In each vineyard and season, the image-measurements are evaluated against carefully collected harvest yield measurements. The yield estimation is studied for both spatial accuracy and overall accuracy, and we demonstrate how to optimize the algorithm to maximize spatial yield accuracy and also to analyze the accuracy for different methods to calibrate the relation between image measurements and yield.

We analyze the berry detection algorithm over the different grape varieties and illumination conditions, assessing the performance of different texture feature descriptions by quantifying which descriptor performs better in which lighting condition and which generalizes the best across datasets without the need for retraining and tuning.

5.1. Equipment Setup

There are significant challenges in designing an imaging system that adheres to a number of considerations, including the *lighting power*, *lighting distribution*, and *specular reflections* from the vine leaves, the *depth-of-focus* appropriate to the vineyard fruiting-zone size plus the variable position of the camera with respect to the fruiting zone, the *frame-rate* of the camera with respect to desired operating velocity, *motion blur*, which is related to the exposure duration, the vehicle velocity, and vibrations from the engine and the terrain undulations, the *recycle time* of the flashes, *heat accumulation* in the flashes, *position-estimation and image-registration*, *image-resolution* to enable detections of small grape berries early in the season, and *image-quality* to enable detection of the fruit.

The camera is mounted about 0.9 and 1.5 m from the fruiting zone, depending on the size of the fruiting zone for the particular vineyard. Illumination is placed directly to the side of the camera to reduce shadowing.

The most current equipment design was deployed to collect data starting in 2013; see Figure 6. To image the fruit, a 24MM F/2.8D AF NIKKOR lens is used in conjunction with a Prosilica GE4000 camera, which are mounted facing sideways on the vehicle viewing the fruit. Additionally, two Einstein 640 mono flashlights are mounted on both sides of the camera. The stereo camera used for visual odometry estimation is a PointGrey BumbleBee2 mounted to the vehicle pointed down the row used within the visual odometry algorithm to estimate the position of the vehicle along the row. Both cameras are triggered by external pulses to maintain synchronization. The images are captured at 5 Hz and the vehicle is driven at 5.4 km/h, which as a rough comparison

(a)

(b)

(c)

Figure 6. Photos of the equipment used to collect the 2013 datasets. Part (b) shows the Prosilica GE 4000 camera with a red-case mounted in the middle of a vibration-damped mounting plate, attached to a custom reconfigurable aluminum frame with an adjustable ball joint at the head. Beside the camera are the two Einstein 640 flash units with reflectors and a custom diffusion filter attached. There is a Hokoyu UTM30LX laser scanner also seen mounted above the camera on the plate, however it is not used in the experiments presented in this work; its placement is designed to measure the size of the vine's canopy [see Grocholsky, Nuske, Aasted, Achar, & Bates (2011)]. In (c) the Bumblebee2 stereo camera can be seen pointed along the direction of travel to be used in a visual odometry algorithm to measure the position of the vehicle along the row.

to other vineyard vehicles is approximately the same speed as a pesticide spraying tractor and faster than a machine harvester.

In experiments before 2013, different cameras, camera lenses, and illumination sources were used. In 2010, a Canon SX200IS camera was used to image the fruit, and halogen lamps were mounted facing sideways toward the fruit. In 2011 and 2012, we used a Nikon D300s camera to capture images of the grape fruit, and an AlienBees ARB800 ring flash mounted around the camera lens to provide even lighting to the scene. In 2011, a 24MM F/2.8D AF NIKKOR lens was used in conjunction with the Nikon camera, whereas in 2012 a Nikon 18–55 mm f/3.5-5.6G ED II AF-S DX zoom lens was used.

5.2. Datasets

The datasets analyzed consist of wine-grape varieties—Traminette, Riesling, Chardonnay, and Petite Syrah—and a table-grape variety called Flame Seedless. We demonstrate our method at a variety of stages during the growing season, from just after the fruit begins setting right up until just before harvest. Over this time span, the berries range from one-tenth their final size to almost fully grown. See Table I for details of the different datasets. See Figure 7 for image examples from the different datasets.

In each of the datasets, we collect harvest weights of the fruit to evaluate against our image measurements.

Table I. Dataset location, time, and details.

Variety	Location	Date	Days to harvest	Trellis type	Lighting	No. Vines
Traminette	Fredonia, NY	Sep 2010	10	Vert. Shoot Pos.	Day w/ lamp	98
Riesling	Fredonia, NY	Sep 2010	10	Vert. Shoot Pos.	Day w/ lamp	128
Chardonnay	Modesto, CA	Jun 2011	90	Vert. Sprawl	Night w/ flash	636
Chardonnay	Modesto, CA	Jun 2013	75	Vert. Sprawl	Night w/ flash	24
Petite Syrah	Galt, CA	Jun 2013	75	Quad. w/ Vert. Shoot	Night w/ flash	30
Pinot Noir	Galt, CA	Jun 2013	75	Quad. w/ Vert. Shoot	Night w/ flash	32
Flame Seedless	Delano, CA	Jun 2011	40	Split-V	Night w/ flash	88
Flame Seedless	Delano, CA	Jul 2012	1	Split-V	Night w/ flash	88
Flame Seedless	Delano, CA	Jun 2013	7	Split-V	Night w/ flash	88

5.2.1. Wine-grape Vineyards

The Chardonnay dataset was collected in Modesto, California in a sprawling vineyard that has guide wires that are lifted to tuck pendant shoots up such that it could be considered a semivertical shoot positioned vineyard. We collected images on 636 vines on six rows of this vineyard. In 2011, the images were collected just after the berries began to set at 12 weeks before harvest. At this stage, the berries are very small, between 3 and 5 mm in diameter and one-tenth of their final weight. In 2013, the images were collected several weeks later in the relative growing season when the berries were larger at 0.8 g.

The Petite Syrah and Pinot Noir datasets were collected in Galt, California, both in a split quadrilateral trained cordon trellis system. The pruning practice and the trellis system both promote vertical growth of the shoots up and over top guide wires such that the canopy grows up and over the wires in a curtain. The fruit is mainly located close to the cordon below the wires. The camera position in this vineyard must be underneath the canopy looking up at the fruit-zone.

The Riesling and Traminette datasets were collected from an approximately one-acre plot in Fredonia, New York. We used four rows of Traminette vines and four rows of Riesling varieties, consisting of 224 vines total. The Traminette were positioned at 8 ft spacing and the Riesling were positioned at 6 ft spacing, which totaled 450 m of vines. Similarly to the Chardonnay vines sampled, the Riesling and Traminette vines were vertical shoot positioned, allowing fruit to be seen underneath the lifted vine canopy.

Basal leaf removal was performed for all of the Chardonnay, Petite Syrah, Riesling, and Traminette vines as per each vineyard's standard operation. For all but the Riesling dataset, the leaf removal was performed only on the north side, and this is the side that was imaged. The practice is commonly performed by vineyard owners to expose the fruit to the sun to change the flavor characteristics of the grapes (Bergqvist, Dokoozlian, & Ebisuda, 2001; Crippen Jr & Morrison, 1986). Basal leaf removal also makes yield estimation feasible after fruit-set at the end of the growing season because the occluding canopy is removed from the fruit-zone.

5.2.2. Table-grape Vineyards

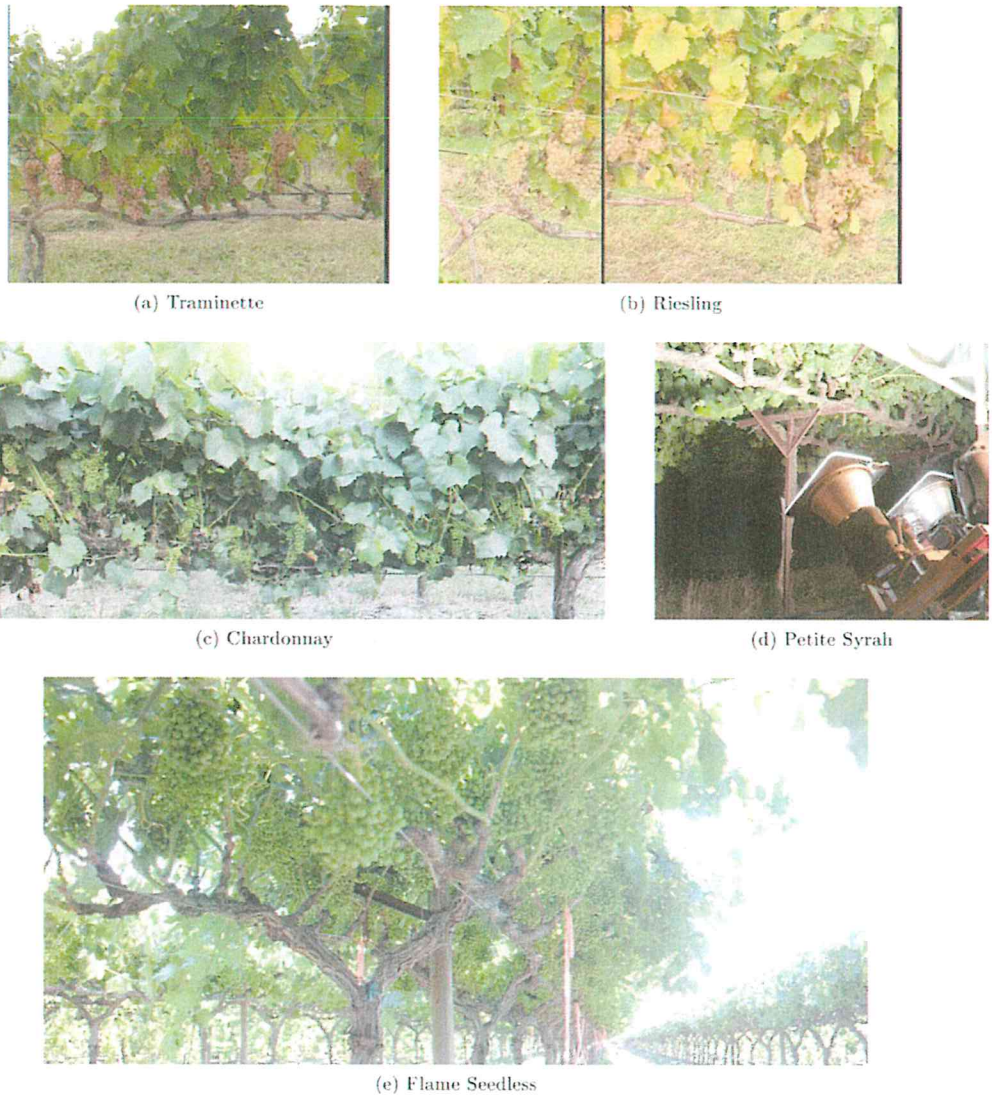
The Flame Seedless datasets were collected over three years—2011, 2012, and 2013—where each occasion consisted of a dataset of 88 vines of a split-V gable vineyard, in Delano, California. The vines are trained in a split system in which each vine is trained into two cordons and the shoots grow over a V-gable trellis such that the fruit hangs in two distinct sections at an angle on each side of the row. The canopy grows out over the V-gable trellis such that to view the clusters, either the canopy must be trimmed back (usually done close to harvest time to enable harvest crews to access fruit) or a cane-lifting device [see Figure 1(b)] is required to reveal the fruit. As per convention in table-grape cultivation, fruit-thinning, shoot-thinning, and leaf-pulling are performed quite rigorously to promote separation of the grape clusters from each other and from leaves. The result of this separation of the fruit is that once the exterior canopy is removed or moved from between the camera and the fruit, the cluster-to-cluster and vine occlusions are noticeably less for this type of vineyard than the wine grape varieties (which is reflected later in Figure 13).

5.3. Berry Detection Performance

We evaluate the performance of our berry detection algorithm by first analyzing the keypoint detection algorithms and then studying the feature classification.

5.3.1. Evaluation of keypoint detection

We take both the radial-symmetry detection algorithm and the invariant-maximal detection algorithm and evaluate based on how many grape centers are detected by each algorithm, as shown in Table II. The table illustrates that the two types of keypoint detection behave differently under different conditions.



(a) Traminette (b) Riesling

(c) Chardonnay (d) Petite Syrah

(e) Flame Seedless

Figure 7. Example images of the different varietals from our yield prediction experiments.

Notable discrepancies between detectors are seen in the 2010 datasets collected in natural illumination. The invariant-maximal detector does not perform well, due to the lack of shading on the surface of the fruit. Also, the same performance is registered in the Flame Seedless 2011 dataset when a cross-polarized flash is utilized, also removing the shading on the surface of the fruit.

In most of the 2013 datasets, the invariant-maximal detector outperforms the radial-symmetry algorithm where the flash produces a peak in the middle of the fruit and a gradual decrease in intensity from the curved surface. The exception is the Chardonnay dataset, where it is believed that a pesticide spray applied just before imaging caused a residue to be left on the surface of the fruit, causing a reduction in the shading produced by the flash.

In general, the results indicate that multiple keypoint detection algorithms are necessary to achieve high performance in multiple different conditions. Further, we also note that the actual location of the keypoints differs between algorithms and the location affects the feature

Table II. Keypoint detection.

Dataset	Radial-symmetry (Recall)	Invariant-maximal (Recall)
Riesling 2010	0.66	0.12
Traminette 2010	0.89	0.09
Chardonnay 2011	0.68	0.91
Chardonnay 2013	0.86	0.77
Flame Seedless 2011	0.89	0.30
Flame Seedless 2012	0.73	0.54
Flame Seedless 2013	0.81	0.85
Petite Syrah 2013	0.87	0.96
Pinot Noir 2013	0.75	0.91

descriptions. This was discovered since performance dropped when training the randomized KD-forest with one detector type and testing with the other keypoint detector type.

5.3.2. Evaluation of feature classification

As mentioned earlier in Section 3.2, we presented a set of three different texture feature descriptors that we evaluate within our algorithm from three broad feature types:

1. Filter banks: Gabor filters
2. Local texture descriptors: SIFT
3. Binary relations: FREAK

We evaluate their relative performance against each other on a set of datasets collected in a variety of illumination conditions; natural illumination, flash illumination, and cross-polarized flash illumination. We use training sets of 20 random images from each of the three types of illumination conditions and manually mark the grape berries in each of the images for ground-truth. Points identified by the algorithm as berries that are near our manually defined ground-truth are considered true positives, and they are false positives if they are not near any manually defined berries. Figure 9 presents the true positive rates and false positive rates through varying values for τ_r (see Section 3.2) from 0 to 1.

We see in Figure 9(a) that a bank of Gabor filters performs substantially better than the SIFT and FREAK features, with color-only having the second best performance. This naturally illuminated dataset was collected after vérason (onset of coloring), and even though this was a green grape variety (Traminette), there was noticeable yellowing of the fruit in comparison to the leaves, indicating why the color features do well. The images are noticeably less clear and noisy than the other types of illumination conditions, and as expected the Gabor filter responses that are placed to encompass each image patch are more robust to these lower quality images.

In Figure 9(b) we see that FREAK features perform best at classifying the berries under conventional flash illumination. We suspect the curved surface of the grape is highlighted by shading from the flash and is captured well in the description of the binary intensity relations of FREAK. This dataset was captured before vérason, and the fruit is much more similar in color to the foliage on the vine. Therefore, we see, as expected, that color alone does not perform well.

Figure 9(c) presents the classification performance with the cross-polarized flash. This type of illumination will remove the glare from the flash, which reduces the glare from the surface of leaves and also removes any glare and shading from the grape surface. In this case, SIFT performs the best, since the shading is not prominent on the grape surface. We hypothesize that the gradient orientation histogram of SIFT is suited to describing the curved contour of the grapes, which is the key visible cue in the absence of shading or distinguishing color.

Finally, we evaluate which descriptor generalizes the best over different datasets and illumination conditions. We test on a dataset using training data from the other datasets. We find that SIFT is best at generalizing across conditions, whereas the Gabor and FREAK descriptions are less able to generalize. In this test, we did not use color in the features, as we see here that color-only performs poorly, since color varies a lot between datasets and illumination conditions. This shows a contrast, as color was seen to be a valuable cue when a classifier was trained and then utilized over the same dataset. In conclusion, it is possible to operate under new conditions without retraining the classifier, but color should not be included in the feature description, and the results indicate that SIFT is the texture descriptor that is best at generalizing.

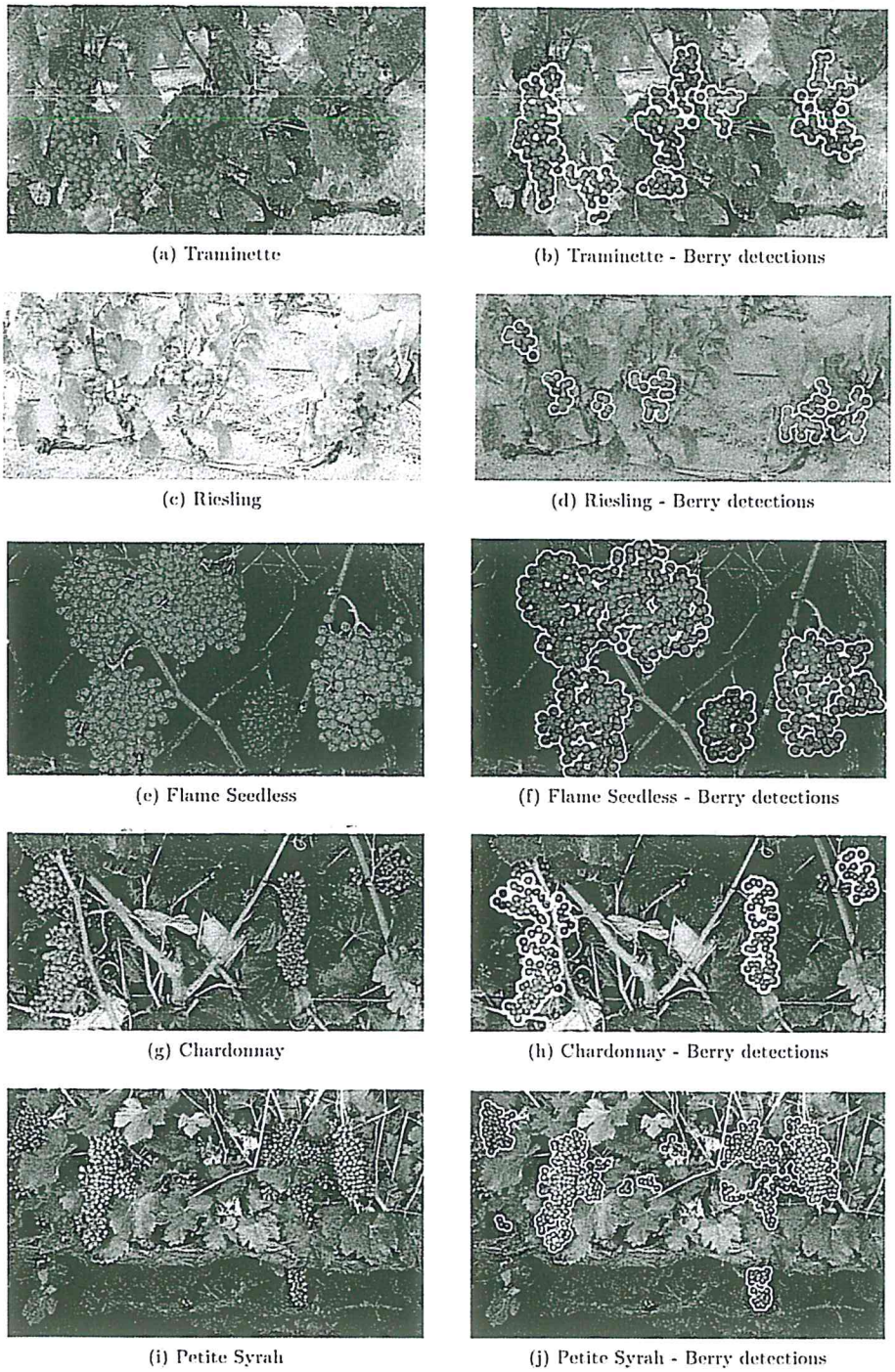
5.3.3. Timings of Berry Detection Algorithm

Evaluating our algorithm on an Intel i5-2500K Quad core 3.3 GHz CPU, with 16GiB of RAM, produces the timing breakdown in Table III.

There has not been great effort into optimizing algorithms for speed. In particular, the Gabor filters use an inefficient implementation. However, the efficiency of the system is currently appropriate for our existing deployments, given that the time before harvest when the images were captured is far greater than the time to process all images in a dataset. Further effort into optimization of algorithms could bring the system close to real-time performance.

5.4. Evaluating Visible Berries and Self-occlusion

First, we evaluate the occlusion of berries within a cluster by the outer layer of clusters (k_l) and study some approaches to potentially improve the estimate of the number of hidden



(a) Traminette (b) Traminette - Berry detections

(c) Riesling (d) Riesling - Berry detections

(e) Flame Seedless (f) Flame Seedless - Berry detections

(g) Chardonnay (h) Chardonnay - Berry detections

(i) Petite Syrah (j) Petite Syrah - Berry detections

Figure 8. Example images demonstrating berry detection in different varietals.

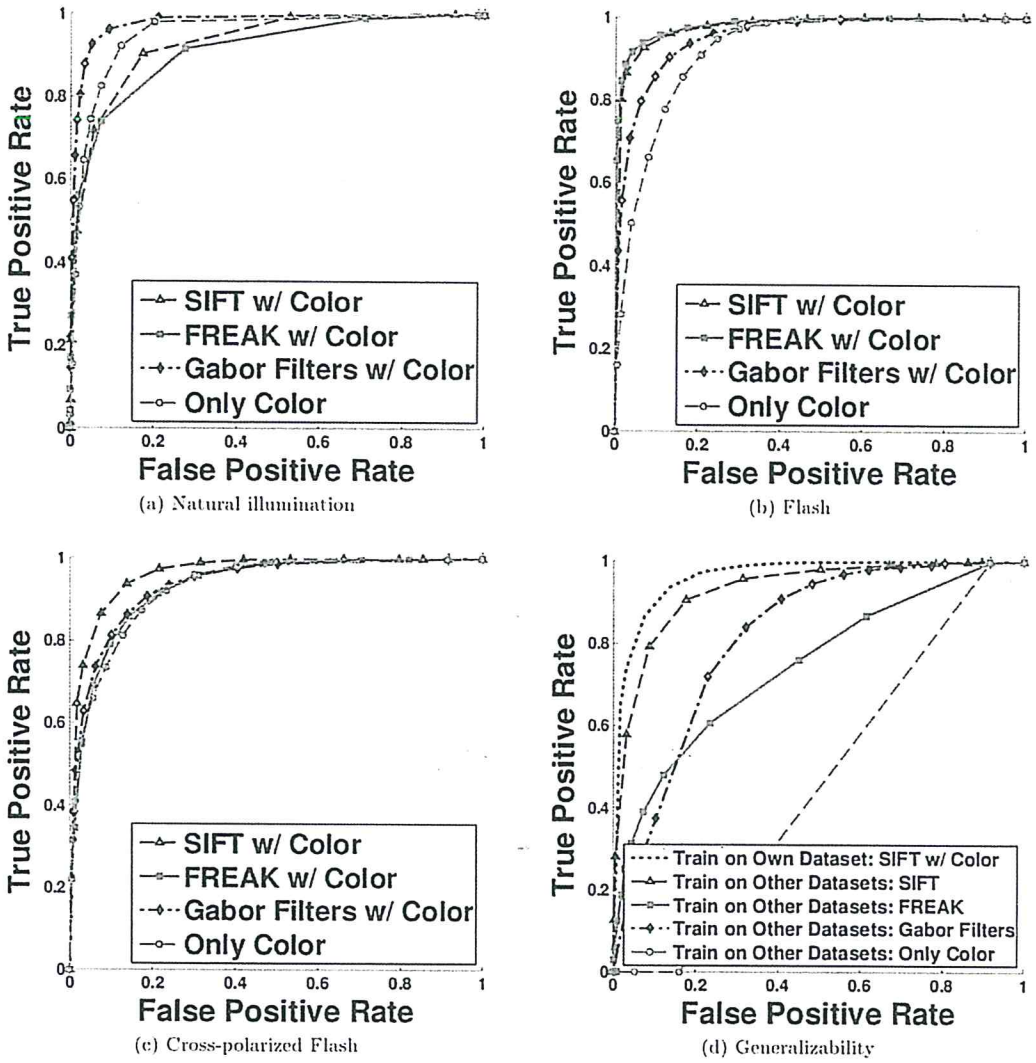


Figure 9. Classification performance curves computed from various feature descriptor types evaluated on a set of different illumination conditions. Natural illumination (a) causes grapes to be imaged with noise from uncontrolled lighting. Gabor filters perform best in this setting, as they are not as responsive to local image noise. Under conventional flash illumination (b) intense shading is seen upon the imaged grapes. The FREAK performs best with this shading on the grapes, which we hypothesize is because the relative-intensity comparisons of FREAK are successful at describing the steady decrease in intensity across the shaded grapes. Under cross-polarized illumination (c) grapes are free of noticeable shading. The exterior contour of the grape is the most noticeable feature on these grapes, and the SIFT descriptor seems to be better than FREAK at describing the value of this strong contour on the grape's exterior. When comparing across datasets, SIFT is the most generalizable descriptor across imaging conditions, as seen in (d).

berries. For this specific study, we use a controlled laboratory environment where we collected images individually of 56 grape clusters. We use ripe clusters of the Thompson Seedless variety. For each cluster, we collected several images from different orientations, at a fixed distance, and we collected a weight and a count of the number of berries. In the laboratory dataset, we do not use our automatic detection algorithm and instead hand-mark all berries visible

Table III. Berry-detection timings.

Algorithm step	Timing (s)
Image load and preprocess	0.097
Keypoint detection–radial symmetry	0.65
Keypoint detection–invariant maximal	0.55
Feature extract–Gabor	6.42
Feature extract–FREAK	0.045
Feature extract–SIFT	1.09
Grape/nongrape classification	0.017
Group berries into clusters	0.22
Total–slowest (radial/Gabor)	7.4
Total–fastest (maximal/FREAK)	0.92

Table IV. Cluster model correlation to fruit weight (laboratory dataset).

Measure-type	R^2 correlation	Mean-squared error
Total berry count (upper bound)	0.95	9.3%
2D Visible berry count [Eq. (4)]	0.88	15.4%
Ellipsoid 3D model [Eq. (6)]	0.85	17%
Convex hull 3D model [Eq. (5)]	0.92	13.7%

within the images to replicate a perfect detection algorithm and remove any bias from errors in the detection algorithm (k_d and k_r). Also, in the laboratory dataset there are no biases from the vine (k_v) or from other clusters (k_c) and hence we can isolate and study the bias from self-occlusions (k_b).

Initially, we compare the total berry count (gathered manually) of each cluster against its weight, Table IV. The correlation score for total berry count to weight is $R^2 = 0.95$ with a mean-squared error from a least-squares fit of 9.3%. We consider this an upper bound for the yield predictions, as the best yield prediction we could achieve depends on accurately knowing the berry count.

Next, we study different image measurements starting with the visible berry count, and we present the results in Table IV. The visible berry count correlates with $R^2 = 0.88$, which provides a mean-squared error of 15.4%; similar visible berry correlations have been found in Grossetete et al. (2012). The error is just 6% greater than the total berry count, and it indicates that a similar fraction of visible berries is present for small clusters as with large. The ellipsoidal model has a correlation score of $R^2 = 0.88$ and the lowest mean-squared error of 17%. Even though the ellipsoidal model attempts to predict the occluded berries behind the visible layer of berries, it correlates with a lower score than the visible berry measure. The ellipsoidal model could be less accurate because it violates one of our assumptions: the clusters do not have uniform density, or the clusters are not ellipsoidal, or the model could suffer from errors in the designation of the cluster contour.

Table V. Cluster model correlation to yield (vineyard datasets).

Measure-type	R^2 correlation
2D Visible berry count [Eq. (4)]	0.75
Ellipsoid 3D model [Eq. (6)]	0.61
Convex hull 3D model [Eq. (5)]	0.41

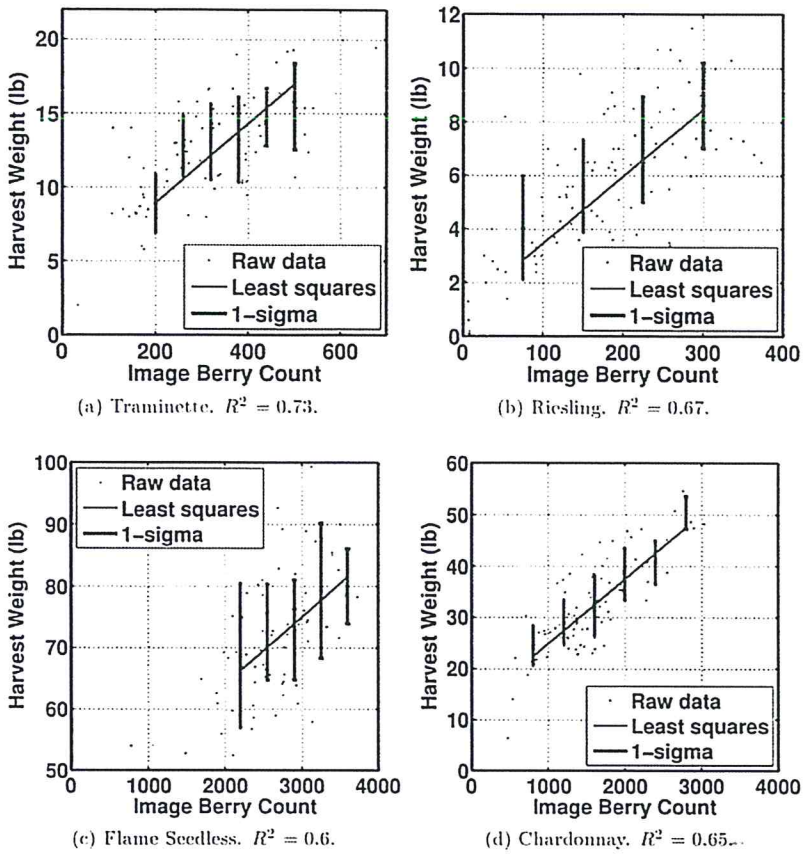
The final image measurement model we evaluate is the convex hull in Table IV. The correlation measures at $R^2 = 0.92$, which is the best of the three image measurements we study. One possible reason for the high correlation is because it encompasses the entire cluster contour. Therefore, it includes a measure of the partially visible berries as well as the completely visible berries, thus being more accurate than visible berry count alone. Despite finding that the contour area in the image is a more accurate measure other than the visible berry count, we do not yet deploy this measure outside the laboratory environment.

We evaluate our different cluster models in a vineyard. Table V presents results from the Traminette dataset. In the vineyard setting, several clusters are visible in each image and we have yet to develop a technique for successfully segmenting one cluster from another—a requirement of the volumetric ellipsoid and convex hull models. In practice, clusters grow to touch one another and it is difficult—without physically moving the clusters—to determine which berries belong to which cluster. Hence, at present we have only been able to demonstrate precise detection of individual berries, regardless of to which cluster they belong, and therefore in the following vineyard results we consider just the visible berry count.

5.5. Spatial Accuracy—Berry Count Correlation to Vine Yield

We compare our berry counts against actual harvest weights collected from the Traminette, Riesling, Flames Seedless, and Chardonnay datasets. First, we register images to vine spaces, and then we assign berry count measurements to each vine space. For the Traminette and Riesling datasets, we manually defined the vine spaces in images, but in the more recent Flame Seedless and Chardonnay datasets we deployed the stereo camera to perform this process automatically.

Once registered to specific vines, we compare our automated berry counts with the harvest crop weights (see



(a) Traminette. $R^2 = 0.73$. (b) Riesling. $R^2 = 0.67$.

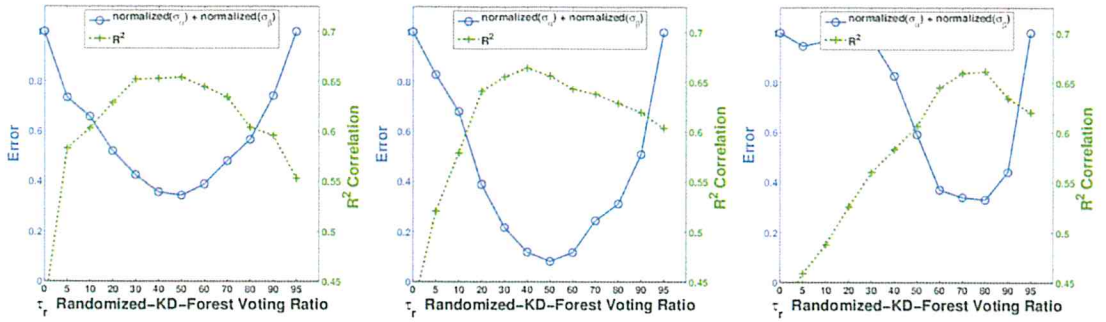
(c) Flame Seedless. $R^2 = 0.6$. (d) Chardonnay. $R^2 = 0.65$.

Figure 10. Correlation between our detected berry count and harvest crop weights. The black lines show the one-sigma standard deviation within the measurements, the red line represents a linear fit, and each of the blue data points represents the raw measurement of a single vine. The caption below shows the R^2 correlation score. These graphs illustrate our method generating a nondestructive measurement at every vine, whereas in conventional practice very sparse destructive samples are taken.

Figure 10 for details). The figure shows the raw data points and the distribution of measurements. Our automatically generated berry counts produced a linear relationship with actual harvest crop weights with correlation scores ranging from $R^2 = 0.6$ to 0.73 depending on the dataset (see Figure 10 for details). The R^2 correlation score quantifies how much of the actual variance in yield our method can estimate. Thus, we capture 60–75 % of the variance. Similar correlation scores to vine yield were achieved by Diago et al. (2012), although it is difficult to do a direct comparison because in that work the yield distribution was artificially induced by stages of defruiting, which reduces the true variance in vine occlusions and cluster-to-cluster occlusions. Furthermore, without considering individual berries, the correlation scores will not hold once the berry size changes since simple pixel count is used as an image measurement.

Our measurements achieve good spatial correlation, first through the high precision of our detection algorithm, which rarely counts false positives (Nuske et al., 2011), and also because there is some consistency in occlusion level across the vineyard. Further, increasing the correlation score could come from possible improvements to the detection algorithm and including a method to estimate the berries that are not visible to the camera. The variance unexplained, derived by simply subtracting $1 - R^2$, is around 25–40 %, which could either be due to variance in our detection performance (i.e., changes in the recall of the algorithm) or to variance in the occlusions, caused by either cluster, self-occlusions, or the vine.

Finally, we demonstrate how to optimize our system for spatial yield accuracy. Section 4.2 identified the metric and means to optimize the classification algorithm. Here we evaluate this approach by configuring with the set of



(a) Optimize Gabor against R^2 (b) Optimize FREAK against R^2 (c) Optimize SIFT against R^2

Figure 11. The randomized KD-forest voting threshold τ_r is optimized against a normalized deviation in true positive rate (σ_α) and normalized false positive (σ_β) [Eq. (16)] to maximize resultant R^2 correlation from berry counts, which is the quantitative measure of spatial yield accuracy.

different feature descriptors, and we optimize the key classification parameter, τ_r. The variance in both the true positive rate and the false positive level is dependent on τ_r, and we optimize over a small set of hand-labeled images (a set of 10 images for a given vineyard dataset) and then evaluate the resultant R^2 value against the harvest data; see Figure 11.

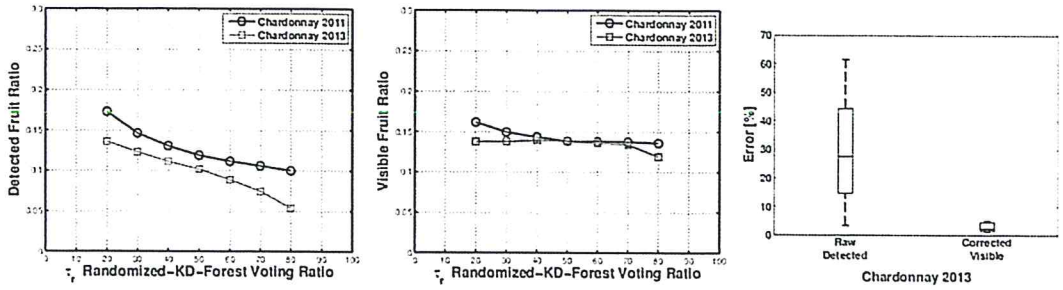
For the three types of feature descriptors, the graphs identify that minimizing the error in Eq. (15) when configuring the algorithm produces a similar peak in r^2 when correlating to harvest yield. For Gabor filters and FREAK features, the optimal τ_r value is between 40% and 50%, although it is noticeable that the error metric is slightly shifted from the r^2 peak, which equates to less than a 5% reduction in spatial yield accuracy. For SIFT features, the τ_r parameter is optimal at 80%, indicating the SIFT feature space has much different properties than FREAK and Gabor, and that in the SIFT feature space the positive berries are in a tighter cluster, but with negative features distributed sparsely but evenly. Nevertheless, the graphs reveal that the optimization strategy is adept and produces substantial increases in spatial yield accuracy in excess of 20%.

5.6. Yield Estimate Accuracy—Berry Count Calibrated to Harvest Yield

We have seen a linear correlation between our image berry counts and harvest yield illustrating that our estimates capture most of the true spatial variance in yield. The next metric to analyze is the accuracy of our system in estimating the overall yield. The calibration procedure to relate the image measurements to yield can be conducted in two different fashions. Either harvest and image data from prior seasons can be used, or destructive calibration samples could be collected at the time of imaging. Using prior seasons harvest data theoretically requires less human labor, since recording total yield at the time of harvest is a standard process in grape production, and further an entire sample of yield from all vines can be collected. The negative aspects to using prior harvest seasons are that you need one full season before you can establish a calibration for computing yield predictions, and the consistency of calibrations from season to season needs to be established so that the integrity and accuracy of the calibration are known. Using manual samples collected at the time of imaging has the advantage that a calibration can be established immediately and that the calibration can be assured to be representative of the current vineyard variety and training system. However, the human labor involved restricts the size of the sample set and, in turn, may adversely affect the accuracy of the calibration.

We present yield prediction results using calibration established from prior harvest data. We use different feature descriptors for different datasets, as per the earlier discovery that different features work better under different conditions. Gabor filters are used for the Riesling, Traminette, and Flame Seedless datasets, and FREAK feature descriptors are used for the Chardonnay and Petite Syrah datasets. The same classification threshold is used throughout, although we also demonstrate in Figure 12 that we correct for the particular classification performance using Eq. (10). Hence, a varying classification threshold can be compensated for, except at the extreme high and low classification thresholds which cause too much variance in classification performance. We also find that for some datasets, some keypoint-types and feature-types are not able to be compensated with Eq. (10), meaning that a changing threshold changes the fruit estimate. However, we repeat training for all features and keypoint combinations and pick the combination that has the most stable fruit estimate and consensus with fruit estimates of other types of features/keypoints.

The images for all datasets are processed otherwise in the same manner, with only one difference being an illumination normalization applied to the 2013 images at preprocessing to correct for the light attenuation at the peripheries



(a) Initial detected fruit (b) Estimated visible fruit (Eq. 10) (c) Calibration error

Figure 12. Illustration of how to correct for variations in berry detection performance due to variations in visual appearance or imaging conditions. The ratio of raw detected fruit to total fruit (a) is corrected by Eq. (10) to compute visible fruit (b) and plotted at different classification thresholds. The visible fruit is computed with measurements of berry detection accuracy computed from a small set of labeled images. The visible fruit in comparison to detected fruit is stable, with most of the different classification thresholds except the extremal thresholds. The extremal thresholds, which are shaded gray, correspond to deviation errors greater than 50% of the maximum from Figure 11(b). The stability in estimated visible fruit in comparison to the instability of the raw detected fruit highlights the necessity to compensate for detection performance when deriving yield estimates. The graph in (c) shows the yield prediction error when using the detected or visible fruit ratio computed from prior years with image measurements from the present year. The visible fruit estimate is stable even with substantial changes in the classification parameter [Eq. (10)].

Table VI. Yield estimation accuracy—Flame Seedless 2013.

Calibration source	Yield prediction error
Flame Seedless 2011	6.48%
Flame Seedless 2012	11.65%
Flame Seedless 2011 and 2012	9.07%

Table VII. Yield estimation accuracy—Chardonnay 2013.

Calibration source	Chardonnay 2013
Chardonnay 2011	−2.47%

of the image. We compute the visible fruit using Eq. (10) and then compute the visibility calibration function [Eq. (17)] from datasets collected in prior years, using the mean berry weight to compute an estimated total number of berries. We then apply the calibration to the image berry counts collected in 2013 and evaluate yield accuracy.

The Flame Seedless vineyard also has a consistent relationship between subsequent years. However, the Flame Seedless vineyard is grown in a substantially different manner than the wine grape vineyards. The vines are trained on a V-gable trellis and the shoots are thinned, the fruit is thinned, and leaves are pulled from in front of clusters such that there is far less occlusion from foliage and from cluster-to-cluster occlusion than the wine grape vineyards. This difference can be seen in the two sets of calibration lines in Figure 13, one for a table-grape split-cordon V-trellis and one for a wine grape single cordon. The yield prediction error for the table-grape data is between 6% and 11.5% accuracy from prior harvest data, and using all prior data to calibrate 2013 image data to an error of 9.07%; see Table VI.

Table VII shows the application of visibility calibration from the Chardonnay 2011 dataset and applied to the Chardonnay 2013 dataset, which achieves an accuracy of 2.5%. This result demonstrates that destructive calibration samples at the time of imaging may not be required, and prior season calibration may be sufficient. Further, calibration taken from the Traminette or Petite Syrah vineyards also achieves good accuracy on the Chardonnay 2013 dataset of 4% and 5%, respectively. However, calibration from Riesling or Pinot Noir vineyards produces 17% and 29% error, which we believe is due to lower levels of vegetation in the Riesling vineyard, and our hypothesis that the Pinot Noir vineyard had a combination of less vegetation in the fruiting zone and the system also detected a small percentage of grapes on the other side of the split cordon. The conclusion is that calibration must be site-specific to achieve high accuracy. For the Chardonnay vineyard the accuracy is 2.5%, and in the Flame Seedless vineyard the accuracy is 9% on average when calibrated from prior years in the same vineyard.

6. LESSONS LEARNED

We draw a number of lessons from this work. One of the most important is that consistent performance is required to ensure that image-measurements stay well-related to yield. To maintain consistent performance, both imaging and algorithmic robustness is important. The illumination and imaging configuration is vitally important, and operating

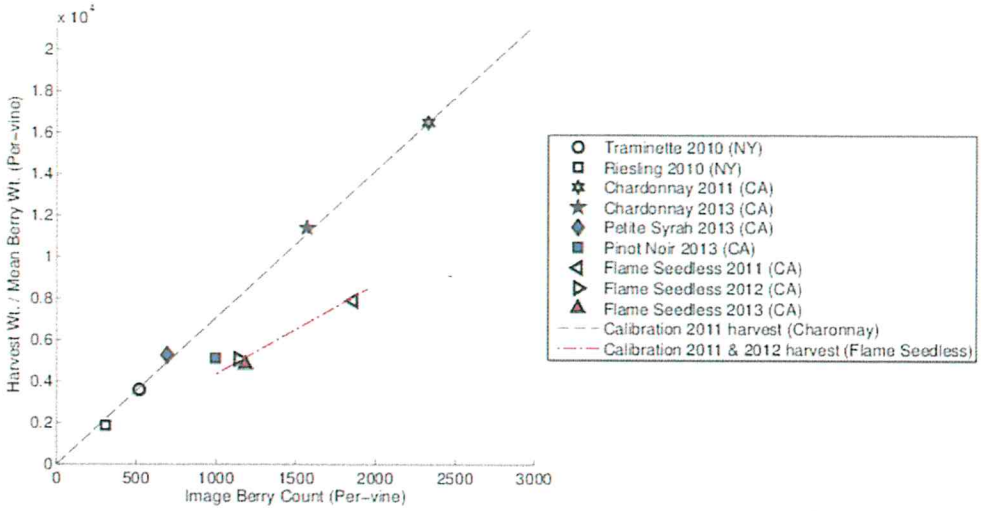


Figure 13. All yield and image berry count data. Calibration lines from prior seasons are drawn from Chardonnay 2011 and Flame Seedless 2011–2012 harvests. The site-specific calibration is consistent over years, within the Chardonnay and Flame Seedless vineyards, highlighting that destructive calibration samples at the time of imaging are potentially not needed. Similar trellising types of Petite Syrah and Traminette also show similar visibility to the Chardonnay. However, it must be noted that Riesling and Pinot Noir vineyards have noticeably higher visibility despite being from similar wine trellising, which indicates, contrary to past hypothesis, that site-specific calibration must be collected.

at night is the most reliable way to control the imaging. Through the development in this work, it became apparent that the imaging and illumination design was a major challenge for practical and robust sensing. Careful engineering is required to achieve the desired characteristics in the illumination power, the depth-of-focus, the frame-rate, flash recycle time, motion blur, and image quality. A naive approach to deploying a camera in a vineyard will simply be unsuccessful. Position estimation and image registration are also essential to maintain consistency, by accounting for the vehicle's motion along the vineyard rows and also the variable distance of the camera from the fruit-wall. If biases exist in the position estimation and image registration, the result will be an inaccurate relationship between image-measurements and yield. Finally, we have seen that with a few manually labeled images (we use sets of 10 labeled images), it is possible to characterize detection performance and configure the algorithm to increase the accuracy of the yield estimates. While it does take time to manually label images, this requires far less manual labor in comparison to the alternative of collecting destructive yield samples.

In terms of what has been learned from the multiple years, we reflect on the results presented in Nuske et al. (2012), and we see that the calibration relationships for the wine grapes aligned more accurately from 2010 to 2011, when errors were 4–5 %. The 2013 predictions report an accuracy of 3–11 % of total yield. The slight decrease in the worst-case accuracy for 2013 could be due to a few reasons. One may be the different intensity profile created by the new flash setup in 2013. In 2010 and 2011, a more uniform light source was used, whereas in 2013 the light was focused using reflectors to project the light forward for power efficiency reasons. This causes a gradual decline in the light toward the edge of the images, and a slight decrease in the amount of berries detected toward the peripheries of the image, which we do correct as best we can with intensity normalization across the image. Changes in the camera placement on the vehicle may introduce some unexpected variations in the fruit occlusion and appearance variations of both the fruit and foliage. Nevertheless, despite a change in camera and flash and vine appearance, the yield prediction accuracy is still within 6% on average. It is logical to conclude that keeping the imaging position and configuration consistent, there may be a further increase in the accuracy of yield predictions. We are also looking at methods to preprocess images to automatically normalize the lighting distributions between datasets and extend our approach to automatically configure the fruit detection algorithms (Section 4.2) to be more robust to dataset variations. Another learned lesson is that the vine-trellis type and management practices of the grower are also critical considerations. For instance, for the vineyards studied in this work, the trellis, the leaf-pulling, shoot/cluster thinning, and shoot positioning varied between the wine-grape vines and table-grape vines. These differences combined to affect the relationship between image-measurements and yield, although

it was shown that the relationship is stable in the same vineyard over several years. For example, the lower errors recorded in the wine-grape vineyards are expected to be due to more consistent visibility of the fruit zone and/or more consistent berry detection performance in these vineyards. In the table-grape vineyards the fruit zone is much larger, and due to the split vines, some of the fruit from one side would grow close to fruit from the other side, and there were challenges involved in positioning the camera and configuring the algorithm to not incorrectly count fruit from the other cordon.

7. CONCLUSION

The method presented in this paper demonstrates nondestructive yield estimation that can enable growers to perform management in a far more effective manner than was previously possible. Unlike previous work in this area, we demonstrate a complete system over several growing seasons and tens of acres that can be deployed efficiently. We address the key challenges relating to distinguishing the berry appearance, the challenges in imaging, estimating the position of the vehicle, and reducing variance in measurement, all together providing an automated solution to large-scale, high-resolution yield estimation.

The algorithm presented here exploits all visual cues of grapes, namely shape, texture, and color, enabling a detection system applicable across many different varieties and imaging conditions. While ideally there would be one particular visual texture-cue for all conditions, we discovered that different texture-descriptors work well in different conditions. While this poses a challenging question of which descriptor to use within the algorithm, we have shown that with a small amount of labeled data, it is possible to quantify the classification performance of individual feature descriptors at the training stage and identify the most optimal descriptor. We also show how to tune the algorithm and descriptor to maximize the accuracy of the resultant yield estimates. In particular, we distilled the model relating image measurements to yield down to the components that directly affect yield accuracy, and we were able to show that our theory results in substantial increases in accuracy.

We also attempted to improve the estimate of berries within a cluster with a volumetric prediction of cluster size, and we demonstrated a method that increases the accuracy of cluster-size estimates within a laboratory setting. When applied to real-world data, the volumetric method was found to decrease accuracy because grape clusters grow alongside one another, which makes it difficult to discover the boundary of neighboring clusters, leading to gross errors in extrapolating cluster size. Therefore, a first-order model that relates the visible berry count to yield is the most accurate in practice. An interesting observation can be drawn that humans are better at counting clusters per vine and weighing individual clusters, whereas, conversely, it seems robotic sensing struggles to accurately count mature grape clusters. Instead, it is easier to use robotic sensing to count the number of berries on a vine, a measure that would not be easy for a human to directly produce.

Finally, we identify stable calibration relationships for vineyards that hold true from year to year within 10% prediction of total yield. The results indicate that the system can be efficiently deployed without new labor-intensive manual calibration, given that a site-specific calibration has been established in prior years.

ACKNOWLEDGMENTS

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

REFERENCES

Alahi, A., Ortiz, R., & Vandergheynst, P. (2012). FREAK: Fast retina yeypoint. In IEEE Conference on Computer Vision and Pattern Recognition, New York.

Berenstein, R., Shahar, O. B., Shapiro, A., & Edan, Y. (2010). Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer. Intelligent Service Robotics, 3(4), 233–243.

Bergqvist, J., Dokoozlian, N., & Ebisuda, N. (2001). Sunlight exposure and temperature effects on berry growth and composition of cabernet sauvignon and grenache in the central San Joaquin Valley of California. American Journal of Enology and Viticulture, 52(1), 1–7.

Blom, P., & Tarara, J. (2009). Trellis tension monitoring improves yield estimation in vineyards. HortScience, 44, 678–685.

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: binary robust independent elementary features. In Proceedings of the 11th European Conference on Computer Vision: Part IV, ECCV'10 (pp. 778–792). Berlin, Heidelberg: Springer-Verlag.

Clingeleffer, P., Dunn, G., Krstic, M., & Martin, S. (2001). Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties: A national approach. Technical report, Grape and Wine Research and Development Corporation, Australia.

Crippen, D., Jr., & Morrison, J. (1986). The effects of sun exposure on the compositional development of cabernet sauvignon berries. American Journal of Enology and Viticulture, 37, 869–874.

Dey, D., Mummert, L., & Sukthankar, R. (2012). Classification of plant structures from uncalibrated image sequences. In IEEE Workshop on the Applications of Computer Vision (WACV).

Diago, M.-P., Correa, C., Millan, B., Barreiro, P., Valero, C., & Tardaguila, J. (2012). Grapevine yield and leaf area estimation using supervised classification methodology on rgb images taken under field conditions. Sensors, 12(12), 16988–17006.

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, 196–198.

Federici, J., Wample, R., Rodriguez, D., & Mukherjee, S. (2009). Application of terahertz gouy phase shift from curved surfaces for estimation of crop yield. Applied Optics, 48, 1382–1388.

Grocholsky, B., Nuske, S., Aasted, M., Achar, S., & Bates, T. (2011). A camera and laser system for automatic vine balance assessment. In American Society of Agricultural and Biological Engineers (ASABE) Annual International Meeting.

Grossetete, M., Berthoumieu, Y., Costa, J. P. D., 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—CIGR-AgEng.

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. 5314–5320).

Hung, C., Underwood, J., Nieto, J., & Sukkarieh, S. (2013b). A feature learning based approach for automated fruit yield estimation. In 9th International Conference on Field and Service Robotics (FSR).

Jimenez, A., Ceres, R., & Pons, J. (2000). A survey of computer vision methods for locating fruit on trees. In Transaction of the ASAE, 43, 1911–1920.

Kitt, B., Geiger, A., & Lategahn, H. (2010). Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In IEEE Intelligent Vehicles Symposium, San Diego.

Laws, K. (1980). Textured image segmentation. Ph.D. thesis, University of Southern California.

Lepetit, V., Lagger, P., & Fua, P. (2005). Randomized trees for real-time keypoint recognition. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference (vol. 2, pp. 775–781).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60, 91–110.

Loy, G., & Zelinsky, A. (2003). Fast radial symmetry for detecting points of interest. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, 959–973.

Marden, S., Liu, S., & Whitty, M. (2013). Towards automated yield estimation in viticulture. In Australasian Conference on Robotics and Automation.

Martinez-Casasnovas, J., & Bordes, X. (2005). Viticultura de precisión: Predicción de cosecha a partir de variables del cultivo e índices de vegetación. Revista de Teledetección, 24, 67–71.

Nuske, S., Achar, S., Bates, T., Narasimhan, S., & Singh, S. (2011). Yield estimation in vineyards by visual grape detection. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems.

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

Rabatel, G., & Guizard, C. (2007). Grape berry calibration by computer vision using elliptical model fitting. In ECPA 2007, 6th European Conference on Precision Agriculture.

Serrano, E., Roussel, S., Gontier, L., Dufourcq, T., et al. (2005). Early estimation of vineyard yield: Correlation between the volume of a vitis vinifera bunch during its growth and its weight at harvest. In FRUTIC 05, Information and technology for sustainable fruit and vegetable production: 7th Fruit Nut and Vegetable Production Engineering Symposium (vol. 2, pp. 311–318).

Singh, S., Bergerman, M., Cannons, J., Grocholsky, B., Hamner, B., Holguin, G., Hull, L., Jones, V., Kantor, G., Koselka, H., Li, G., Owen, J., Park, J., Shi, W., & Teza, J. (2010). Comprehensive automation for specialty crops: Year 1 results and lessons learned. Journal of Intelligent Service Robotics, Special Issue on Agricultural Robotics, 3(4), 245–262.

Swanson, M., Dima, C., & Stentz, A. (2010). A multi-modal system for yield prediction in citrus trees. In ASABE Annual International Meeting, Pittsburgh, PA.

Tabb, A., Peterson, D., & Park, J. (2006). Segmentation of apple fruit from video via background modeling. In American Society of Agricultural and Biological Engineers Annual International Meeting.

Taylor, J., Tisseyre, B., Bramley, R., Reid, A., Stafford, J., et al. (2005). A comparison of the spatial variability of vineyard yield in European and Australian production systems. In Precision agriculture'05. Papers presented at the 5th European Conference on Precision Agriculture, Uppsala, Sweden. (pp. 907–914). Wageningen Academic Publishers.

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

Wolpert, J. A., & Vilas, E. P. (1992). Estimating vineyard yields: Introduction to a simple, two-step method. American Journal of Enology and Viticulture, 43, 384–388.

Yang, L., Dickinson, J., Wu, Q. M. J., & Lang, S. (2007). A fruit recognition method for automatic harvesting. In Mechatronics and Machine Vision in Practice, 2007. M2VIP 2007. 14th International Conference (pp. 152–157).

Zhou, R., Damerow, L., Sun, Y., & Blanke, M. (2012). Using colour features of cv. âgalaâapple fruits in an orchard in image processing to predict yield. Precision Agriculture, 13(5), 568–580.