Classification of Plant Structures from Uncalibrated Image Sequences

Debadeepta Dey[1]   Lily Mummert[2*]   Rahul Sukthankar[2,1*]

debadeep@ri.cmu.edu   lily@cs.cmu.edu   rahuls@cs.cmu.edu

[1] The Robotics Institute, Carnegie Mellon University   [2] Intel Labs Pittsburgh

Abstract

This paper demonstrates the feasibility of recovering fine-scale plant structure in 3D point clouds by leveraging recent advances in structure from motion and 3D point cloud segmentation techniques. The proposed pipeline is designed to be applicable to a broad variety of agricultural crops. A particular agricultural application is described, motivated by the need to estimate crop yield during the growing season. The structure of grapevines is classified into leaves, branches, and fruit using a combination of shape and color features, smoothed using a conditional random field (CRF). Our experiments show a classification accuracy (AUC) of 0.98 for grapes prior to ripening (while still green) and 0.96 for grapes during ripening (changing color), significantly improving over the baseline performance achieved using established methods.

1. Introduction

Precision agriculture deals with characterizing spatial variation in the production of crops, which increases efficiency by reducing inputs such as fertilizer and more accurately predicting crop yield. Although extensive work has been based on remote sensing from satellite and airborne platforms, ground-based sensing is increasingly in demand because of its lower cost, ease of data acquisition, ability to observe smaller structures such as fruit that may not be visible from above, and potential for characterization at the level of individual plants.

Information about plant structures is useful in a variety of ways. Estimates of crop yield allow adjustments to be made during the growing season, and drive business decisions such as planning logistics for harvest and market preparation. Measurements of foliage provide an estimate of a plant's ability to intercept sunlight (and thereby mature a crop), and its water requirements [42]. The amount and condition of foliage can indicate water stress, nutrient excess or deficiency [34]. Other structures such as stem
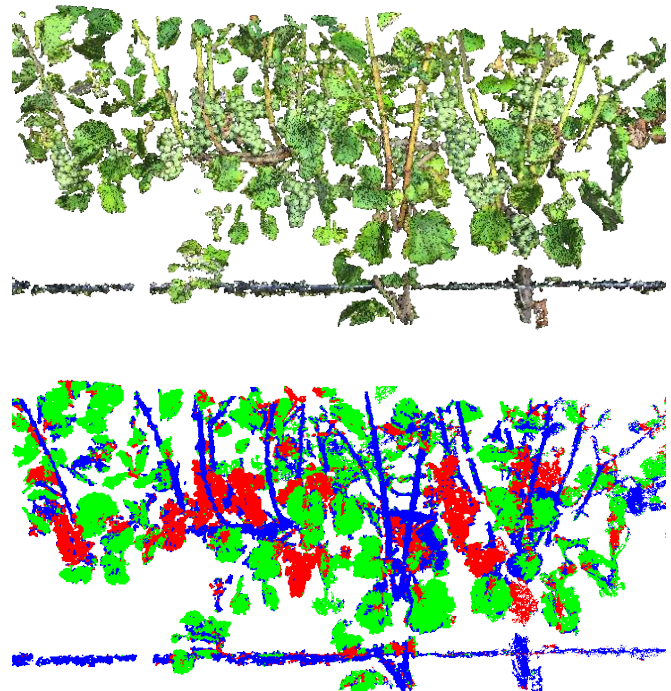

Figure 1. Reconstruction of grape vine from uncalibrated image data, and automatic classification using the proposed method into leaves (green), branches (blue), and fruit (red). Obtaining better non-destructive measurements of foliage and crop enables more accurate crop yield estimates.

length and branching can also indicate plant stress [26]. Current practice for obtaining these measurements often involves destructive sampling — for example, removing all of the fruit from a small percentage of plants to estimate crop weight. Work in agricultural sensing seeks to make such measurements non-destructively and automatically, with greater coverage of the crop, at multiple times during the growing season.

A variety of sensing modalities have been used to obtain measurements of plant structures, including LIDAR and ul-

---

trasonic sensing for 3D information [12, 27, 28, 32, 40, 41, 43]. Recent improvements in digital camera technology and structure from motion algorithms allow detailed 3D structures to be recovered from uncalibrated images at relatively low cost [18, 37]. Compared to analyses of 2D imagery, these reconstructions ameliorate the effect of occlusions and provide data registration. Compared to LIDAR, which is limited by spot size in the detail that can be sensed, image-based reconstructions use a lower cost sensor and provide both color and the potential to recover finer grained structures. Once a reconstruction is obtained, the next task is to automatically classify the content into semantically meaningful structures, such as leaves and fruit.

The main contributions of the paper are listed below:

- We propose a general vision-based processing pipeline for classification of plant structures that leverages recent advances in structure from motion and 3D point cloud segmentation.

- Our pipeline, which operates on uncalibrated images, replaces the need for expensive outdoor LIDAR and complicated data acquisition efforts with affordable commodity cameras and the ease of collecting data in a point-and-shoot manner.

- The plant structures are identified using both color and local 3D shape, which makes proposed the method robust to lighting and inevitable color changes as the crop matures.

- We describe an example application on real world datasets, motivated by the need to estimate crop yield during the growing season. Grapes are detected on vines at two different times during the growing season – before and after the start of ripening.

- Experiments show a classification accuracy as measured by area under the ROC curve of 0.98 for grapes prior to ripening (while still green) and 0.96 for grapes during ripening (changing color). These results are a significant improvement over standard techniques.

2. Related Work

Jiménez *et al*. [20] summarize the early progress on computer vision methods for locating fruit on trees. Moreda *et al*. [29] discuss methods for non-destructive estimation of fruit and vegetable volume using a broad range of sensing modalities, including but not limited to computer vision. McCarthy *et al*. [27] provide a review of current computer vision approaches (monocular, stereo, range and multi-spectral) for automated farming. Taylor *et al*. [39] present a survey of methods for incorporating new sensor information into viticultural practice. We organize the discussion of related work in two main categories: (1) fruit

detection and yield estimation, and (2) canopy and foliage reconstruction.

Work on fruit reconstruction primarily explores the imaging of individual pieces of fruit under controlled conditions, such as Forbes *et al*. [13], Chalidabhongse *et al*. on mango volume estimation [6] and Chaivivatrakul *et al*.'s on pineapple reconstruction [5]. While our proposed method is motivated by similar high-level goals, our applications demand robust fruit detection and yield estimation in natural settings, where the fruit must be segmented from nearby foliage. Swanson *et al*. [38] face similar challenges in yield estimation for citrus groves but their approach employs a combination of laser sensing to estimate canopy density and computer vision for fruit detection, while we require only a single commodity camera. The majority of recent computer vision approaches for yield prediction, such as Aggelopoulou *et al*. [2], employ straightforward image processing approaches (e.g., thresholding) by exploiting visual characteristics of their domain (e.g., white flowers against a dark background). Unfortunately, color is not a sufficient feature in our domain since green grapes and foliage are similar in color space. Nuske *et al*. [31] detect grapes using 2D techniques on artificially illuminated images taken at night using multiple flashes. In contrast, our approach can use but does not require specialized lighting.

Aguilar *et al*. [3] present early results on 3D modeling of tomato plant canopies to non-destructively estimate leaf area index, with applications in determining correct pesticide dosage. Bienert *et al*. [4] build a 3D representation of forest areas as input for a numerical wind flow simulation model. Although applied to different domains, their eigenvector analysis of 3D point clouds is structurally similar to the features we propose for classification (Section 3.2.1). Fleck *et al*. [12] recover the branch system and leaf-mass area of apple trees using 3D data obtained using a laser scanner, while Henning *et al*. [19] build 3D models by merging multiple calibrated laser scans. Several recent systems propose LIDAR for modeling canopy structure [32], assess fuelbed characteristics of trees [24], leaf area index estimation [28], branch modeling [16], fruit harvesting [21], and tree structure derivation [41]. There are only a few examples of vision-based systems related to 3D foliage reconstruction. Shlyakhter *et al*. [36] build 3D model of foliaged trees from instrumented (rather than uncalibrated) photographs with the goal of generating realistic trees in computer graphics applications. Phattaralerphong *et al*. [33] showed encouraging results (in simulation) that silhouette voxel carving could be used to reconstruct foliage. Zhang *et al*. [44] present preliminary work on recovering a 3D model of a corn stalk. Dandois and Ellis [9] acquire photos using an aerial kite platform and argue that reconstruction using imagery is competitive with LIDAR. Finally, Haala *et al*. [17] combine laser scanning with panoramic
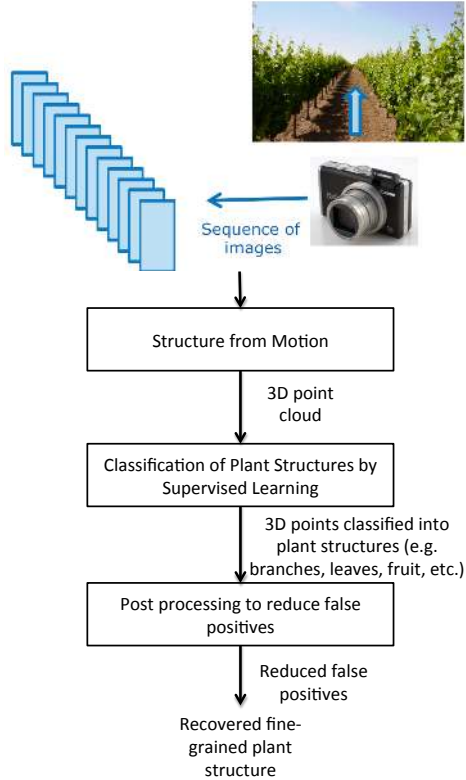
Figure 2. Overview of proposed pipeline for recovering fine-scale plant structure. We analyze local 3D geometry in the recovered point cloud to label plant structures, such as leaves and fruit.

imaging for classifying tree species; they acquire position and diameter of trees using the laser data and use visual texture parameters from images. In our proposed method, classification is used to segment the 3D scene into semantically meaningful structures (branches, leaves and fruit) rather than classifying whole trees by species.

Point cloud classification has recently been studied for applications such as object map construction in indoor settings [35], urban environment reconstruction [15, 30] and robot navigation in forested terrain [23]. Our approach is motivated by that of Lalonde *et al.* [23], where the tree trunks are identified by exploiting their locally linear structure. However, we focus on classifying finer-grained plant structures such as twigs, individual leaves and berries.

3. Overview of Processing Pipeline

Figure 2 shows an overview of the proposed vision-based processing pipeline. The input to the pipeline is a sequence of images acquired from a relatively inexpensive commodity digital camera. The data collection is simple and does not require time consuming calibration of the camera intrinsics. Furthermore, complicated setup procedures required for other sensors (e.g., LIDAR) are obviated. The sequence

of images is processed using a structure-from-motion module to recover a dense, colorized 3D reconstruction of the scene. In the next stage, local features are extracted from the point cloud and used to independently classify each point. The classified point cloud is smoothed in a post-processing step that uses generic spatial constraints and domain specific knowledge to reduce false positives. The fine-grained plant structure of interest is recovered at the end.

We briefly outline each module in the following sections with respect to the specific problem of estimating crop yield in grapevines.

3.1. 3D Reconstruction

Given an uncalibrated sequence of images, we first employ structure-from-motion to obtain a dense, colorized 3D point cloud. This is accomplished in a two stage procedure:

- obtaining a sparse 3D reconstruction of the scene using a scalable bundle adjustment algorithm [37];

- converting this representation into a dense, colorized 3D point cloud using recent progress in multi-view stereopsis [14].

These stages are detailed below.

Snavely *et al.* [37] propose a structure-from-motion algorithm suitable for internet-scale photo collections which has enabled *Photosynth* [1] and similar applications. We chose this particular algorithm for structure-from-motion because of its proven, robust performance on large scale image datasets with wide variations in viewpoint, illumination and scale. *Bundler* proceeds by extracting local features like SIFT [25] on each input image. Features are matched across pairs of images followed by finding *tracks* which are matches across multiple images. An initial pair of images which are well conditioned (large baseline, non-degenerate) is chosen. The fundamental matrix and camera parameters are estimated between the initial pair, and the locations of the sparse 3D points are estimated by bundle adjustment. To this initial 3D reconstruction other images are added and the expanding 3D reconstruction refined in an iterative fashion. This procedure is continued until no other suitable images are feasible to be added to the reconstruction. This results in a sparse 3D scene reconstruction.

The sparse 3D scene reconstruction from *Bundler* is input to the dense, robust, multi-view stereopsis software by Furukawa *et al.* [14]. This algorithm initializes rectangular 3D patches and iterates between an expansion step and a filtering step. In the expansion step, photometric consistency is used to spread the match to nearby pixels while in the filtering step visibility constraints are used to eliminate incorrect matches. The output at the end of this stage is a dense, colorized 3D reconstruction of the scene. An example reconstruction of a grape vine is shown in Figure 1.

Next, we extract shape and color features in the local neighborhood of each point and train a classifier to label each point into semantic categories, such as foliage, branch and fruit.

3.2. Classifying Grapevine Structures

The dense 3D point cloud generated from the uncalibrated image sequences captures the overall shape of the scene. Our goal is to recover the fine-grained plant structure, e.g., to determine which points correspond to berries, foliage or branches. We formulate this as a supervised classification problem and leverage recent work in robot obstacle detection [23] to segment the scene using the statistics of local spatial point distribution in conjunction with appearance information, such as color. Although color can be a useful cue (even for distinguishing unripe grapes from foliage), it is unreliable due to illumination conditions and variability within plants. Therefore, we focus primarily on classification based on shape features. Since classifications based solely on local information can be noisy, we spatially smooth the results using a conditional random field (CRF).

The key aspects of our approach are the choice of local features, the classifier and the spatial smoothing, all of which are detailed below.

3.2.1 Feature extraction

The features we employ to model plant structure can be domain specific. For the grapevine structures application, our features employ both color and shape information, since color alone cannot reliably discriminate foliage from unripe or green grape varieties. For color features, we use the (R, G, B) value of the 3D point (as obtained from appropriate images).

For shape, we employ a variant on the "saliency feature" proposed by Lalonde *et al.* [23] for traversability analysis of point cloud data in outdoor ground mobile robots. Intuitively, we characterize the local neighborhood around each 3D point in terms of its "point-ness", "curve-ness" and "surface-ness". In other words, we examine whether the local neighborhood is like a point (compact and spatially isotropic), a line (a distribution with a single dominant axis of spatial variation), or a thin flat sheet (two axes of variation). These mathematical concepts map directly to semantic categories of plant structures, such as berries (which tend to be round and compact), branches (roughly linear) and leaves (largely flat). The analysis of point cloud neighborhoods can be efficiently performed using singular value decomposition.

For computational efficiency, we store the point cloud in a kd-tree, which enables fast lookup of the local neighborhood around each point. For each point $\mathbf{x}$ in the data set, we perform the following operations. First, we retrieve the

set of points in its neighborhood $\mathcal{N} = \{\mathbf{x}_i : ||\mathbf{x}_i - \mathbf{x}|| < d\}$, where d specifies the support region. Next, we compute the covariance matrix for this neighborhood as:

$$\sum_{\mathbf{x}_i \in \mathcal{N}} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top, \qquad (1)$$

where $\bar{\mathbf{x}}$ denotes the mean of the 3D points in the neighborhood $\bar{\mathbf{x}} = \sum_{\mathbf{x}_i \in \mathcal{N}} \mathbf{x_i}$.

We perform Singular-Value Decomposition on this covariance matrix to identify the principal components of $\mathcal{N}$'s spatial distribution. Specifically, we analyze the relative magnitudes of the top three eigenvalues $\lambda_0 \geq \lambda_1 \geq \lambda_2$. For isotropic spatial distributions (corresponding to berries), we expect $\lambda_0 \approx \lambda_1 \approx \lambda_2$; for predominantly linear distributions (branches), $\lambda_0 \gg \lambda_1 \approx \lambda_2$; and for roughly planar distributions (leaves), $\lambda_0 \approx \lambda_1 \gg \lambda_2$.

We represent the shape of the local neighborhood $\mathcal{N}$ (specified by d) around a given point $\mathbf{x}$ using the following feature vector:

$$\text{saliency}_{\mathbf{x}} = \begin{bmatrix} \text{point-ness} \\ \text{curve-ness} \\ \text{surface-ness} \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ \lambda_0 - \lambda_1 \\ \lambda_1 - \lambda_2 \end{bmatrix}. \qquad (2)$$

For each point in the point cloud we compute this saliency at 3 spatial scales (d), concatenate these and append $\mathbf{x}$'s color, $(R(\mathbf{x}), G(\mathbf{x}), B(\mathbf{x}))^\top$ to obtain a 12-dimensional feature vector for each point in the cloud.

3.2.2 Classification

Our grapevine data set was manually labeled into three semantic classes (berry, branch and leaf). Using a portion of this data, we trained a multi-class support vector machine (SVM) classifier [8] to classify points. Section 4 presents classification results on a separate test set using a variety of parameters.

For completeness we investigated the Gaussian Mixture Model (GMM) classifier as used by Lalonde *et al.* [23] but found that SVM consistently outperforms GMM.

The output of the SVM generates an independent classification for each 3D point based solely upon its local spatial neighborhood. As confirmed in our experiments, such a classification can be quite noisy since it ignores the labels of surrounding 3D points.

3.3. Spatial smoothing

We address the classification noise induced by independent 3D point classification by spatially smoothing the labels generated by the SVM. Specifically, we formulate this task as an energy minimization problem using a conditional random field (CRF) framework with discrete labels [22]. Each 3D point is treated as a node in the CRF, initialized

(a) Before ripening

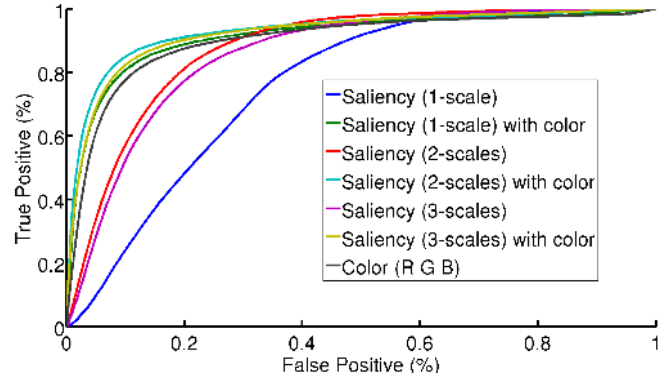

(b) During ripening

Figure 3. Grape image examples

Figure 4. ROC curves for green Gewürztraminer grapes using SVM classification prior to spatial smoothing.

4. Evaluation

We present a selection of experiments on recovering fine structure from uncalibrated image sequences of grapevines bearing unripe and ripe berries.

4.1. Dataset

Since there are no standard, publicly available datasets for this domain, we collected two image sequences of Gewürztraminer grapes in a research vineyard. The sequences were collected three weeks apart, before and after the start of ripening. In the first sequence, the grapes were green, while in the second, they had begun changing color. The vines were trained with vertical shoot positioning, a common cultural practice that places the fruit in a predictable zone on the vine. In addition, the leaves had been pulled from the fruit zone, making the fruit visible. Leaf pulling is another common practice that increases sunlight exposure and air circulation around the fruit, making the vine less susceptible to certain diseases. Both image sequences concentrated on the fruit zone.

Images were taken with a Canon PowerShot SX200IS camera with a 5mm focal length lens and fill-in flash. A total of 133 8-megapixel images were acquired of a 15 foot section containing four vines. We ensured that consecutive images in a given sequence overlap by at least 50% to provide sufficient correspondences for the structure from motion pipeline. Figure 3 shows sample images from both sequences. The data was manually labeled and partitioned into training and test sets.

4.2. Results

Figure 1 shows a snapshot of the reconstruction of the green Gewürztraminer grapes. The reconstruction of the green and purple grape data set had 839,000 and 933,000 total 3D points respectively. Of these, 87,900 and 73,308 points from the green and purple grape data set respectively were used to train a support vector machine with the ra-
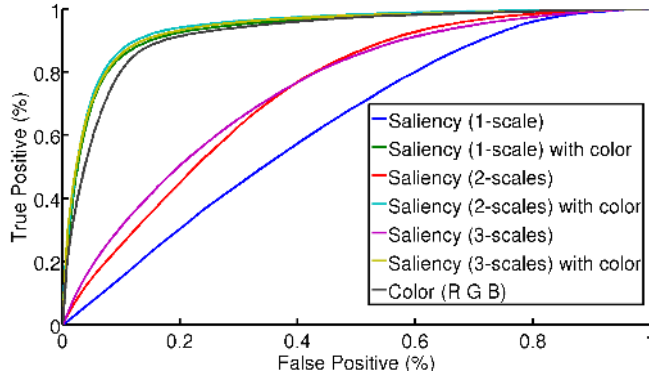
with the output generated by the SVM. Each node is connected to those 3D points in its spatial neighborhood and we employ a simple pairwise term that penalizes adjacent nodes with differing labels. The size of the spatial support region should be selected using domain knowledge; using too small a volume does not significantly change the labels, while employing too large a volume can result in the loss of desirable small objects (such as individual berries) because they are treated as outliers. In practice, we smooth over regions that are similar in size to the support region of our spatial features. We use the algorithm by Delong *et al*. [10] to solve the minimization problem. Spatial smoothing significantly improves the accuracy of the point classification system as shown in Section 4.

Figure 5. ROC curves for purple Gewürztraminer grapes using SVM classification prior to spatial smoothing.

Figure 6. ROC curves for green Gewürztraminer grapes before and after CRF-based spatial smoothing. Spatial smoothing improves the AUC from 0.92 to 0.98.

Figure 7. ROC curves for purple Gewürztraminer grapes before and after CRF-based spatial smoothing. Spatial smoothing improves the AUC from 0.94 to 0.96.

dial basis kernel ($\sigma = 1.0$) using the LIBSVM implementation [7]. The training data contained equal numbers of points from each of the three semantic classes to ensure good performance on the grape class as grapes are only approximately 10% of the points with the majority of the points belonging to foliage and branches. The rest of the points in the data set were held for testing.

Figures 4 and 5 show ROC curves for classification of green and purple grapes, respectively, using the SVM prior to spatial smoothing. The figures show the performance of saliency features with and without color at three different combinations of scales, as well as color alone.

The nominal scale was chosen by inspection of the point cloud such that points within approximately a two inch radius of the point under consideration contributed to the computation of the saliency feature for that point. For features using two scales, the second was half the nominal scale. For features using three scales, the third was twice the nominal scale.

The figures show that the multi-scale saliency features with two scales performed better than either single scale or three scales, with an area under the curve (AUC) of 0.92 and 0.94 for green and purple grapes, respectively. For three scales, the highest scale incorporates contributions from points which are far away from the point under consideration, and provide distracting information about the local shape around the point. Similarly the single scale alone does not encapsulate all the necessary information about the characteristic local shape around the point.

Saliency features alone did not perform as well for purple grapes compared to green grapes. Many of the purple grape images were taken from a greater distance from the vines compared to the green grape images. This increased distance resulted in a noisier reconstruction, which made the local shape of leaf points similar to the spatially isotropic local shape of grapes, instead of being predominantly flat. As a result, a large number of foliage points were misclassified as grapes and vice versa. This problem
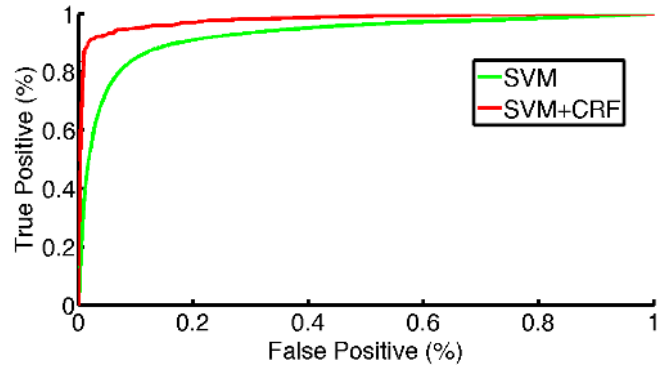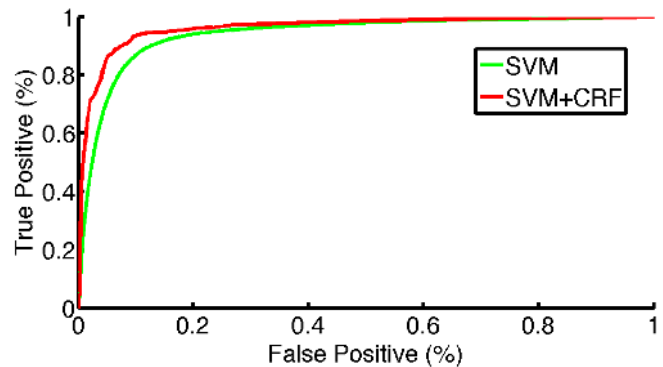
can be fixed by ensuring that images are collected at a closer distance to the vines.

As expected we observe that adding color features can improve results, provided that the images were acquired under similar illumination conditions. For purple grapes, we see that color alone performs well, but combining color with shape features further improves accuracy. For green grapes, the improvement due to color can be explained by the fact that under our illumination conditions, the camera perceives the fruit and foliage as slightly but sufficiently different shades of green. As discussed earlier, the use of color alone is to be avoided, particularly since the color of the grapes changes as they ripen. Since the local 3D shape of fruit and foliage is both invariant to lighting changes and does not change (in terms of statistical properties) through the season. Employing color in conjunction with shape enables us to be more robust to color changes.

Figures 6 and 7 show the ROC curves obtained after spatially smoothing the output of the independent classifier using the CRF, for the green and purple grape datasets, respectively, as compared to the ROC curve obtained from

SVM classification alone. For clarity, we only show results using the best set of features: saliency computed at two scales along with color. These results show that our CRF-based spatial smoothing produces a significant improvement in the true positive rate and a drop in false positive rate. For instance, on green grapes a nominal operating point of 84.8% true positive and 11.6% false positive rate improves to 91.9% true positive and 2.7% false positive rate. Overall, for green grapes the area under the curve (AUC) improves from 0.92 to 0.98 due to the CRF-based smoothing. Similarly for purple grapes, the AUC is boosted from 0.94 to 0.96.

As discussed in Section 3.3, the size of the support region can impact the effectiveness of spatial smoothing. For these experiments, on the green grapes dataset we defined the support region to be the same size as the larger of the two spatial saliency features (corresponding to approximately 2 inches). Since the 3D point cloud reconstructions from the purple grape dataset are less reliable, we find that a smaller support region (corresponding to approximately 1 inch) results in a greater improvement on this dataset. Automatically selecting the support region using cross-validation is practical since the spatial smoothing procedure is not computationally time-consuming.

5. Conclusion

This paper demonstrates the feasibility of recovering fine-scale plant structure in 3D point clouds obtained from uncalibrated image sequences using structure from motion. The proposed method employs a combination of shape and color features to model the local neighborhood around a given 3D point in terms of its spatial distribution. This enables us to label each point as either a fruit (isotropic distribution), leaf (planar) or branch (linear). Our experiments on both unripe and ripe grapes show that an SVM classifier using our features can detect fruit (for yield estimation) with high accuracy, particularly when the labels are spatially smoothed using a conditional random field.

We plan to build upon these results by extending our work in several directions. First, preliminary experiments show that we can obtain more accurate 3D reconstructions by exploiting the sequential nature of the image data (e.g., using [11]). We will also continue to work towards a yield estimate in grapes by counting either bunches or berries. Finally, we will evaluate the generality of the proposed approach by applying it to a variety of other crops.

Acknowledgments

References

[1] http://photosynth.net/default.aspx. 331

[2] A. Aggelopoulou, D. Bochtis, S. Fountas, K. Swain, T. Gemtos, and G. Nanos. Yield prediction in apple orchards based on image processing. *Precision Agriculture*, pages 1–9, 2010. 330

[3] M. A. Aguilar, J. L. Pozo, F. J. Aguilar, J. Sanchez-Hermosilla, F. C. Páez, and J. Negreiros. 3D Surface Modelling of Tomato Plants Using Close-Range Photogrammetry. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008. 330

[4] A. Bienert, R. Queck, A. Schmidt, C. Bernhofer, and H.-G. Maas. Voxel Space Analysis Of Terrestrial Laser Scans In Forests For Wind Field Modeling. In *ISPRS Commission V Mid-Term Symposium*, 2010. 330

[5] S. Chaivivatrakul, J. Moonrinta, and M. N. Dailey. Towards Automated Crop Yield Estimation: Detection and 3D Reconstruction of Pineapples in Video Sequences. In *International Conference on Computer Vision Theory and Applications*, 2010. 330

[6] T. Chalidabhongse, P. Yimyam, and P. Sirisomboon. 2D/3D Vision-Based Mango's Feature Extraction and Sorting. In *International Conference on Control, Automation, Robotics and Vision*, pages 1–6, 2006. 330

[7] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm. 334

[8] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20, 1995. 332

[9] J. P. Dandois and E. C. Ellis. Remote Sensing of Vegetation Structure Using Computer Vision. *Remote Sensing*, 2(4):1157–1176, 2010. 330

[10] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *CVPR*, 2010. 333

[11] A. W. Fitzgibbon and A. Zisserman. Automatic Camera Recovery for Closed or Open Image Sequences. In *ECCV*, 1998. 335

[12] S. Fleck, D. van der Zande, M. Schmidt, and P. Coppin. Reconstructions of Tree Structure from Laser-Scans and Their Use To Predict Physiological Properties and Processes in Canopies. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2004. 330

[13] K. Forbes. Volume estimation of fruit from digital profile images. Master's thesis, University of Cape Town, 2000. 330

[14] Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multiview Stereopsis. *PAMI*, 32:1362–1376, August 2010. 331

[15] A. Golovinskiy, V. Kim, and T. Funkhouser. Shape-based Recognition of 3D Point Clouds in Urban Environments. In *ICCV*, 2009. 331

[16] B. Gorte and D. Winterhalder. Reconstructions of Laser-Scanned Trees Using Filter Operations in the 3D Raster Do-

main. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2004. 330

[17] N. Haala, R. Reulke, M. Thies, and T. Aschoff. Combination of Terrestrial Laser Scanning with High Resolution Panoramic Images for Investigations in Forest Applications and Tree Species Recognition. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2004. 330

[18] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 330

[19] J. G. Henning and P. J. Radtke. Ground-based Laser Imaging for Assessing Three-dimensional Forest Canopy Structure. *Photogrammetric Engineering & Remote Sensing*, 72(12):1349–1358, 2006. 330

[20] A. R. Jiménez, R. Ceres, and J. L. Pons. A Survey of Computer Vision Methods for Locating Fruit on Trees. *Transaction of the American Society of Agricultural and Biological Engineers*, 43(6):1911–1920, 2000. 330

[21] A. R. Jiménez, R. Ceres, and J. L. Pons. A vision system based on a laser range-finder applied to robotic fruit harvesting. *Machine Vision and Applications*, 11:321–329, 2000. 330

[22] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*, 2001. 332

[23] J.-F. Lalonde, N. Vandapel, D. F. Huber, and M. Hebert. Natural terrain classification using three-dimensional ladar data for ground robot mobility. *Journal of Field Robotics*, 23(10):839–861, 2006. 331, 332

[24] E. L. Loudermilk, J. K. Hiers, J. J. O'Brien, R. J. Mitchell, A. Singhania, J. C. Fernandez, J. Wendell P. Cropper, and K. C. Slatton. Ground-based LIDAR: a novel approach to quantify fine-scale fuelbed characteristics. *International Journal of Wildland Fire*, 18:676–685, 2009. 330

[25] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 331

[26] C. McCarthy, N. Hancock, and S. Raine. On-the-go Machine Vision Sensing of Cotton Plant Geometric Parameters: First Results. In J. Billingsley and R. Bradbeer, editors, *Mechatronics and Machine Vision in Practice*. 2008. 329

[27] C. McCarthy, N. Hancock, and S. Raine. Applied machine vision of plants: a review with implications for field deployment in automated farming operations. *Intelligent Service Robotics*, 3:209–217, 2010. 330

[28] I. Moorthy, J. R. Miller, B. Hu, J. A. J. Berni, P. J. Zareo-Tejada, and Q. Li. Extracting tree crown properties from ground-based scanning laser data. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 2830–2832, 2007. 330

[29] G. P. Moreda, J. Ortiz-Cañavate, F. J. García-Ramos, and M. Ruiz-Altisent. Non-destructive technologies for fruit and vegetable size determination — a review. *Journal of Food Engineering*, 92(2):119–136, 2009. 330

[30] D. Munoz, N. Vandapel, and M. Hebert. Directional Associative Markov Network for 3-D Point Cloud Classification. In *International Symposium on 3D Data Processing, Visualization and Transmission*, 2008. 331

[31] S. Nuske, S. Achar, T. Bates, S. Narasimhan, and S. Singh. Yield estimation in vineyards by visual grape detection. In *IROS*, 2011. 330

[32] K. Omasa, F. Hosoi, and A. Konishi. 3D lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of Experimental Botany*, 58(4):881–898, 2007. 330

[33] J. Phattaralerphong and H. Sinoquet. A method for 3D reconstruction of tree crown volume from photographs: assessment with 3D-digitized plants. *Tree Physiology*, 25:1229–1242, 2005. 330

[34] A. G. Reynolds and T. K. Wolf. Grapevine Canopy Management. In T. K. Wolf, editor, *Wine Grape Production Guide for Eastern North America*. 2008. 329

[35] R. Rusu, Z. Marton, N. Blodow, M. Dolha, and M. Beetz. Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems*, 56(11), 2008. 331

[36] I. Shlyakhter, M. Rozenoer, J. Dorsey, and S. Teller. Reconstructing 3D Tree Models from Instrumented Photographs. *IEEE Computer Graphics and Applications*, 21(3):53–61, May/Jun 2001. 330

[37] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. *IJCV*, 80:189–210, 2008. 330, 331

[38] M. Swanson, C. Dima, and A. Stentz. A Multi-Modal System For Yield Prediction in Citrus Trees. In *American Society of Agricultural and Biological Engineers Annual International Meeting*, 2010. 330

[39] J. Taylor, B. Tisseyre, and J.-P. Praat. Bottling Good Information: Mixing Tradition and Technology in Vineyards. In *Information and Technology for Sustainable Fruit and Vegetable Production*, Fruit, Nut and Vegetable Production Engineering Symposium, 2005. 330

[40] S. D. Tumbo, M. Salyani, J. D. Whitney, T. A. Wheaton, and W. M. Miller. Investigation Of Laser and Ultrasonic Ranging Sensors for Measurements of Citrus Canopy Volume. *Applied Engineering in Agriculture*, 18(3):367–372, 2002. 330

[41] D. Van der Zande, W. Hoet, I. Jonckheere, J. van Aardt, and P. Coppin. Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. *Agricultural and Forest Meteorology*, 141:147–160, 2006. 330

[42] L. Williams and J. Ayars. Grapevine water use and the crop coefficient are linear functions of the shaded area measured beneath the canopy. *Agricultural and Forest Meteorology*, 132:201–211, 2005. 329

[43] Q.-U. Zaman and A. W. Schumann. Performance of an Ultrasonic Tree Volume Measurement System in Commercial Citrus Groves. *Precision Agriculture*, 6:467–480, 2005. 330

[44] W. Zhang, H. Wang, G. Zhou, and G. Yan. Corn 3D Reconstruction with Photogrammetry. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008. 330