

A Novel Benchmark RGBD Dataset for Dormant Apple Trees and its Application to Automatic Pruning

Shayan A. Akbar Somrita Chattopadhyay Noha M. Elfiky Avinash Kak
School of Electrical and Computer Engineering, Purdue University

sakbar@purdue.edu, chattops@purdue.edu, nelfiky@purdue.edu, kak@purdue.edu

Abstract

Dormant pruning is a necessary procedure in the field of specialty crop production. In order to mitigate the need of huge labor, automation of this pruning process has become a topic of utmost importance in the field of horticulture. 3D modeling and reconstruction is a major step in such robotics precision agriculture. In this paper, we introduce a new public dataset which can be used for reconstructing dormant apple trees. Our dataset comprises of 9 different apple trees in both indoor and outdoor environment. The images are collected using a portable Kinect2 sensor. To the best of our knowledge, this is the first publicly available dataset for the application like 3D modeling of dormant trees. We hope that the dataset will provide the entire research community working towards mechanizing dormant pruning a baseline benchmark for evaluating different 3D reconstruction and modeling algorithms.

1. Introduction

Dormant pruning is a mandatory procedure in specialty crop production. This process is necessary for healthy growth of the trees and quality fruit production. In winter, when the trees have no foliage, trained pruners are employed to carefully remove certain primary branches (i.e., branches that are connected directly to the trunk) of the fruit trees based on some predefined rules by the horticulture community. This entire process requires huge manpower, and thus, turns out to be an extremely expensive as well as time consuming process.

In the past few decades, researchers have conducted extensive research on mechanization of fruit tree pruning. Their main focus has been on accurate representation of objects and scenes in 3D to support the development of automatic pruning machines. Thus, 3D reconstruction and modeling of the trees with accurately measured trunk and branch diameters has emerged to be an essential stage of this pruning process. Accurate information about the trunk

and the primary branches of the concerned tree is required to precisely identify the potential pruning points, the final goal being supplying all these information to an intelligent robotic agent for automatic pruning.

Automation in such robotics precision agricultural activities require detailed and accurate information about different kinds of tree structures. An increase in the available datasets would allow researchers to conduct their experiments in more detailed, holistic and accurate fashion. One main bottleneck in the research community performing 3D reconstruction of trees is the lack of available dataset. In this paper, we introduce a new dataset¹ consisting of images of several dormant apple trees from different orchards of varied tree structure and varied complexity level. We expect that this dataset will also serve as a benchmark dataset providing a platform for an objective performance comparison of different reconstruction algorithms on real-world data.

The remainder of this paper is organized as follows. An overview of the current state-of-the-art is presented in Section 2. A detailed description of our new dataset and its acquisition process is given in Section 3. We provide qualitative results for some of the trees in the dataset in Section 4. In Section 5, we explain how we have processed the data for reconstructing its 3D structure. The metrics that can be used for performance evaluation are described in Section 6. Experimental results are discussed in Section 7. Finally, in Section 8 the content of the paper is summarized and potential future works are highlighted.

2. Related Works

Existing literature on geometric reconstruction and 3D scene modeling is vast. This section first provides a brief overview of the state-of-the-art of 3D reconstruction and modeling techniques in the field of automatic pruning of trees. We also mention few approaches that use Kinect sensor for this purpose. The section ends with a list of publicly available datasets in the computer vision community.

¹https://engineering.purdue.edu/RVL/CVPRW_Dataset

Researchers have investigated 3D technologies such as stereo vision, laser scanners and structured light for 3D sensing and object reconstruction. However, only limited studies have been reported which aim at detecting 3D appearance of trees. Moreover, the literature mostly focused on reconstructing trees with foliage [29, 19]. In [36], a shape-from-silhouette is employed for reconstructing the dormant tree structure. However, the experiments were conducted in a indoor setup; thus, still has open questions about its performance in actual high density modern orchards. Wang et al. propose a RGB-D sensor based scheme for 3D reconstruction of dormant trees for robotic pruning purpose in [37]. Their reconstruction results look promising in a laboratory setup; however, the results deteriorate in outdoor orchard environment in presence of direct sunlight. Some other methods [4, 17] employ generative branching rules without taking into account the fact that the trees are predominantly cylindrical in structure, and using this information may lead to simplification of the tree structure. Thus, these methods may result in inaccurate reconstruction for complex tree structures in the absence of large number of hypotheses. Another method [20] focus on reconstructing complex unfoliated trees from a sparse set of images by integrating 2D/3D tree topology as shape priors into the modeling process. The authors of [11] and [5, 8] have aimed at modeling dormant trees using stacked circles and semi-circle fitted 3D reconstruction schemes, and have achieved promising results.

All the images listed in our proposed dataset are acquired with a Kinect2 sensor. Although other laser scanners like LIDAR provide images of better quality, they come along with side-effects of being expensive and heavyweight. We chose Kinect2 sensor due its low cost and easy portability. It is observed that Kinect provided us with comparable quality depth images (at a very high frame rate of 30 fps) as of other expensive sensors. Several other traditional approaches [24, 31, 21, 22, 15] have used Kinect to capture data because of these benefits.

In scientific domains like computer vision, robotics, software engineering, agriculture and many other, publicly available datasets play an important role. Open databases pose challenging questions to a wider community and allow direct comparisons of the state-of-the-art. In computer vision and machine learning communities, there exists a wide variety of public datasets addressed towards different kinds of applications. Few such datasets include — the RGB-D datasets for evaluation of visual SLAM and odometry systems [35, 34], the KITTI dataset for the purpose of autonomous driving based on stereo imaging [13], the VAIS dataset for recognizing maritime images in the visible and infrared spectra [40], etc. There is a huge collection of databases and benchmarks for applications related to stereo vision and optical flow — such as the Malaga urban

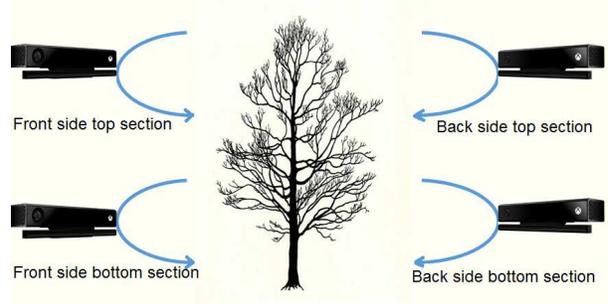


Figure 1: Data Collection Procedure: Images are captured from front side top section, front side bottom section, back side top section, and back side bottom section while moving Kinect sensor as illustrated using arrows.

dataset [7], Middlebury datasets [27, 14, 28], the Cityscapes dataset for semantic understanding of urban scenes through visual perception [9], benchmarks for evaluating optical flow algorithms [6], pedestrian detection benchmarks [10] and many more. There are several other publicly available databases for purposes like recognition [18, 25], classification [39, 26, 12] and image retrieval [23]. Several other datasets exists in these fields. However, we have noticed that no open benchmarks are available for applications related to automatic pruning of dormant trees. The goal of this paper is to provide a real world field image dataset to the computer vision, robotics and horticulture communities who are dealing with 3D reconstruction and modeling of dormant apple trees for pruning purposes.

3. Dataset and Acquisition

In this section, we describe the information contained in our apple tree dataset, the format of the dataset, and the data acquisition procedure along with the hardware and software details of the sensor.

3.1. Dataset Details

The apple tree dataset contains information about nine trees. Out of these nine trees, three are present indoor while six are located in the outdoor environment. The trees belong to two states of the USA: Indiana and Pennsylvania.

For data acquisition, each tree is divided into two sections: top and bottom. And for each section, data is collected from both front and back sides of the tree. Consequently, data collection procedure is performed four times for each side and section of a tree. Figure 1 illustrates the data acquisition procedure for each section and side of tree.

There are five types of information that are collected for building the dataset: the depth images of the tree, the color images corresponding to the depth images, ground truth images, diameter values of primary branches, and distances

between consecutive branches. In Table 1, we summarize all the information about each individual tree.

The depth images of the tree are stored in a binary file format. The reason for choosing this format is because storing binary files is very time-efficient. The binary file contains an array of floating point numbers which correspond to the depth values. This array of numbers can be reshaped to form a matrix of size 424 x 512, which is the resolution of the depth image.

Apart from depth images, their corresponding color images are also captured and stored in PNG file format. The resolution of these images is 1920 x 1080.

The branches of the tree are labeled from bottom to top and ground truth images of the tree are captured from different angles using a regular camera.

Apart from images, ground truth diameter measurements and distance measurements between consecutive branches are also noted using caliper. The diameter measurements are taken at approximately 2.5 inches from the origin of the branch from the trunk as recommended by horticulturists. And distance between two consecutive primary branches is the distance between origin of the first primary branch from the trunk to the origin of second primary branch from the trunk. The ground truth images and measurements assist in the evaluation procedure in which the 3D model is compared with the actual tree.

3.2. Dataset Acquisition

The depth images and their corresponding color images are captured using Kinect2 sensor in the form of a continuous stream while moving the sensor around the tree. To capture depth images, the Kinect2 sensor uses the Time-of-Flight principle in which the phase delay between emitted and reflected IR signals are measured to calculate the distance from sensor to the target object. The lightweight Kinect2 sensors are inexpensive and easily available.

Kinect SDK 2.0 [1] is used to develop the data acquisition software. There are several methods available in the .NET Framework to store data in a file. Since writing images to a file at such high frame rate is time consuming, *BufferedStream* class of .NET framework is used for storage. As mentioned in [3], this method of storage is more time efficient than the other methods. All five types of information mentioned in the previous Section are stored in a single directory for each individual tree.

4. Qualitative Result

In this section, we provide qualitative results for five trees in the dataset. Specifically, we show a labeled ground truth image, a sample depth image, its corresponding color image, and 3D point cloud obtained using Meshlab [2] for each tree in Fig. 2.

Specifically, columns 1, 2, 3, 4, and 5 provide results for trees having tree IDs *Tree1*, *Tree3*, *Tree6*, *Tree7*, and *Tree9*, respectively.

Note that the red color labels for branch numbers are shown in the ground truth images (first row of the Fig. 2). The labels are marked from bottom to top near the origination of the primary branch at trunk. Also, note that the depth images and their corresponding color images belong to the same camera viewpoint.

The 2D depth images are converted to 3D point cloud format using the intrinsic camera matrix \mathbf{K} as follows:

$$\mathbf{V}(u) = \mathbf{D}(u)\mathbf{K}^{-1}u \quad (1)$$

Where u represents the pixel coordinate of depth image, $\mathbf{D}(u)$ represents depth value at locate u , and $\mathbf{V}(u)$ represent the 3D coordinates corresponding to the pixel in depth image. The point clouds for five depth images of the trees are visualized using MeshLab software and a snapshot of the visualization is shown in the last row of the Fig. 2 for each tree.

5. Data Processing

This section provides a detailed discussion about how images of our dataset can be processed for final 3D reconstruction. In Section 5.1, we discuss how to remove lens distortion from the captured images. Then, in Section 5.2 we present the preprocessing operations used to denoise the images, and technique applied for filling in the internal gaps present in our data is described in Section 5.3. Finally, Section 5.4 demonstrates how we perform reconstruction of the dormant tree structures.

5.1. Sensor Calibration

The first step in our processing framework involves calibration of the Kinect2 depth sensor to remove the lens distortion from our raw depth images. Our application thus requires a perfect knowledge of the intrinsic calibration parameters (focal length, principal point, radial, and tangential distortion coefficients) of the Kinect depth sensor for an accurate 3D reconstruction [16]. For this purpose, we have used a sphere based calibration technique for RGB-D sensors proposed by Staranowicz et al. in [32, 33]. The process involves moving a spherical object in front of the sensor for a few seconds and a robust feature extraction pipeline is used to automatically detect and track the object in the images. The calibration algorithm is based on a novel least squares method that provides initial and robust estimate of the sensor parameters. Popular ellipse and sphere fitting techniques like Hough transform and RANSAC are employed to minimize the measurement noise and presence of outliers.

Table 1: Information about dataset: The first column specifies the ID of the tree. There are nine trees in the dataset with IDs *Tree1* to *Tree9*. The state of the USA to which the tree belongs is given in the second column. IN denotes Indiana, while PA indicates Pennsylvania. In the third and fourth columns, checkmarks are made based on whether the tree is present indoor or outdoor. The fifth, sixth, seventh, and eighth columns specify the number of depth images and their corresponding color images in front side top section, back side top section, front side bottom section, and back side bottom sections of the tree, respectively. Ninth and tenth columns mention the number of branches present in the top and bottom sections of the tree. Note that there are no entries for Top portion of *Tree7* because no branches were found on the top portion of the tree.

Tree ID	Tree Location			Number of images				Number of branches	
	State	Indoor	Outdoor	Front Top	Back Top	Front Bottom	Back Bottom	Top	Bottom
<i>Tree1</i>	IN	✓		80	60	30	75	5	12
<i>Tree2</i>	IN	✓		100	70	70	80	9	11
<i>Tree3</i>	PA	✓		80	70	60	65	11	12
<i>Tree4</i>	PA		✓	77	220	148	250	5	10
<i>Tree5</i>	PA		✓	175	195	240	195	6	13
<i>Tree6</i>	PA		✓	130	120	115	110	15	18
<i>Tree7</i>	PA		✓	-	-	160	200	-	16
<i>Tree8</i>	PA		✓	180	175	175	180	7	9
<i>Tree9</i>	IN		✓	170	150	230	240	3	3

The intrinsic parameter vector \mathbf{k} is obtained by calibrating the sensor. Subsequently, the undistorted depth map can be obtained as follows:

$$\mathbf{X}_t = \begin{bmatrix} 2k_3xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4xy \end{bmatrix} \quad (2)$$

$$\mathbf{X}_u = (1 + k_1r^2 + k_2r^4 + k_5r^6)\mathbf{X} + \mathbf{X}_t \quad (3)$$

Where $\mathbf{X} = (x, y)$ is the input depth map, \mathbf{X}_t is the tangentially undistorted map, \mathbf{X}_u is the radially and tangentially undistorted map, $r^2 = x^2 + y^2$, and $\mathbf{k} = (k_1, k_2, k_3, k_4, k_5)$ is the parameter vector, where k_1, k_2 and k_5 are the radial distortion parameters; while k_3 and k_4 are the tangential distortion parameters.

As of now, we have only used depth images for reconstruction. Our next plan is to incorporate the color information in our framework. To obtain the intrinsic calibration matrix and distortion parameters of the color sensor, and the extrinsic parameters i.e. the relative position and orientation between the color and depth sensors, we would be calibrating the RGB sensor as well using the method stated above [32, 33].

5.2. Filtering Noise from Raw Depth Images

A substantial portion of any raw RGB-D frame is occupied by the background. Once sensor calibration is done, the next step in reconstruction process is to remove most of the background by taking only the points within a 2D bounding box where we expect to find the object. Moreover, the background contains a lot of clutter, specially around the boundary pixels of the trees, which needs to be removed before further processing. In our approach, we segment the

background from the foreground object by setting a value of 0 to the pixel values of the depth image which are greater than a certain threshold. We call this threshold the **Depth Threshold**, and represent it by the symbol τ . We use τ to define the maximum depth beyond which all the depth values in the depth image are set to zero. The value of τ should be altered depending upon the position of the object from the sensor. In our experiments, we set the value of τ between 1m and 2m for different trees.

5.3. Filling up the Gaps in the Depth Images

Another challenge in accurate 3D reconstruction of the dormant apple trees is the missing data i.e. the presence of gaps and holes in the depth images. As a solution to this problem, we apply grayscale morphological operations to fill up the internal gaps present in the images. We perform a closing operation, where the depth image A is first dilated, and then eroded with an all-one 3×3 structuring element B as follows:

$$A \bullet B = (A \oplus B) \ominus B \quad (4)$$

Where \oplus and \ominus denote the morphological operations, dilation and erosion, respectively. These operations also help in elimination of the bays along the boundaries and getting smoother edges.

5.4. Application of Dataset for 3D Reconstruction

The purpose of this section is to demonstrate the usability of the proposed dataset and provide baseline results by applying it to one of the most challenging computer vision problems i.e. 3D reconstruction. In particular, we explain one of the previously proposed 3D reconstruction ap-

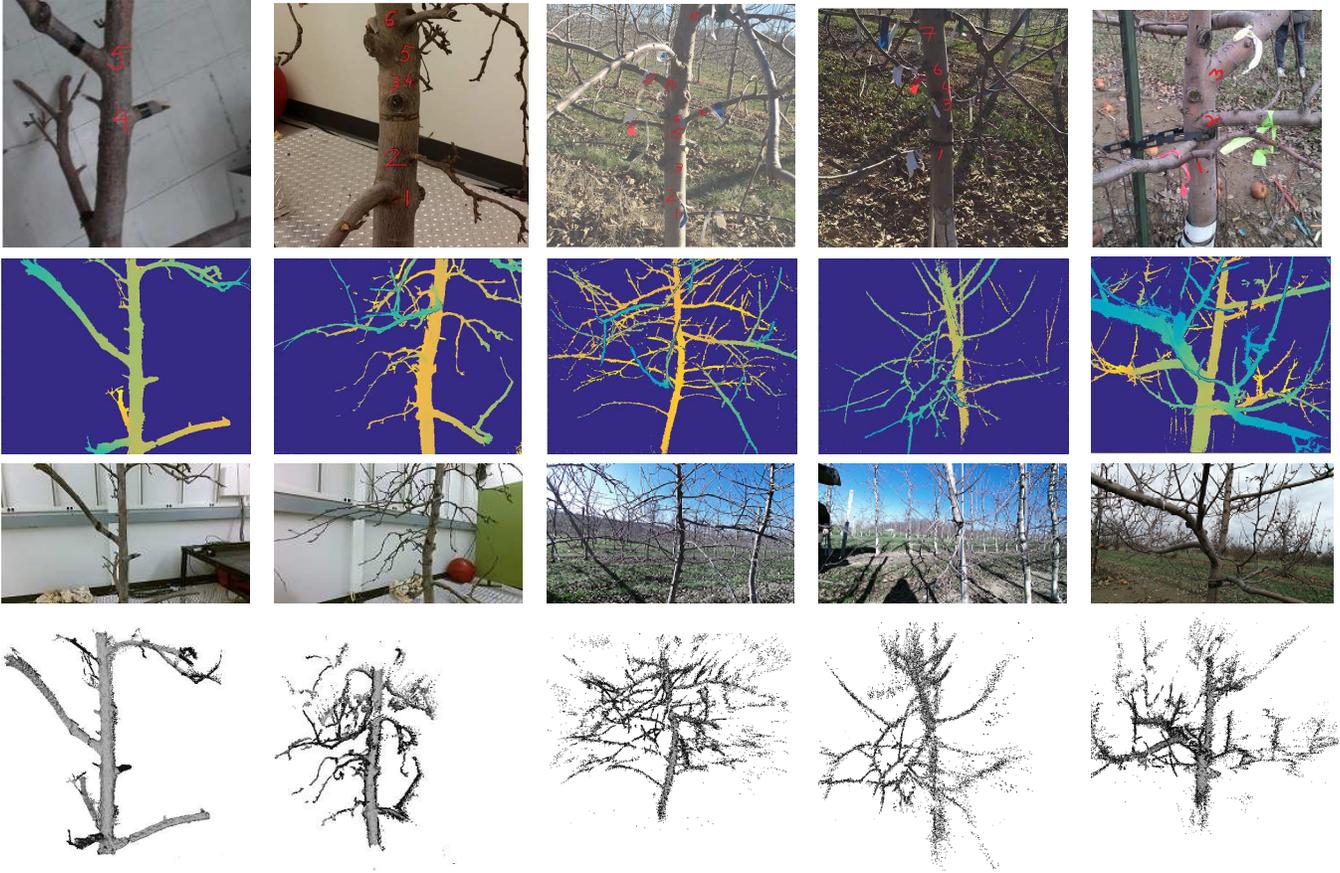


Figure 2: Qualitative results for five trees. The first row shows the labeled ground truth images of the five trees. The second and third rows show the depth images and their corresponding color images of the trees. The final row shows the point clouds obtained from the depth images of the trees.

proaches [8] which can be applied on our dataset. However, the users can also employ their own pipeline on the images of our dataset for modeling.

The method begins by filtering the depth images as explained in Sections 5.1, 5.2, and 5.3. Once we have a filtered clean depth image, we extract the tree skeleton from the depth image, as proposed by Yohai et al. in [30]. The centers and radii of the semicircles are estimated using the skeleton nodes and their interconnections. Then, depth images are converted to point clouds which are aligned to provide transformation matrices. ICP algorithm [38] is used for the purpose of aligning pairs of 3D point clouds. The transformation is applied on the 3D semicircle fit models to form the final 3D reconstruction. Please refer to [8] for details. We will use *Incremental Approach* algorithm mentioned in their paper.

6. Benchmark Evaluation Methodology

A critical point with any dataset is how to measure the performance of the algorithms. To provide a common baseline for comparing different dormant tree reconstruction algorithms, in this section we have provided few methodologies that can be used for the purpose of performance evaluation. To quantitatively measure the performance of the proposed approach in [8, 5], three different metrics, namely, *Branch Identification Accuracy*, *Estimation Error* and *Confidence Value* are used.

- **Branch Identification Accuracy (BIA)** : *BIA* basically gives the accuracy of modeling primary branches. It is computed as the percentage of the detected branches verified to be true over the actual number of ground truth branches of the tree. For example, a 100% *BIA* means if there exists 10 branches in the actual tree, our proposed method could reconstruct all 10 of those branches; whereas, 50% *BIA* indicates that our algorithms could reconstruct only 5 out of 10 branches.

- **Estimation Error** : This is calculated based on the absolute difference between the ground truth diameter (measured using caliper) and the estimated diameter (obtained using the reconstruction approach). Lesser estimation error indicates a better reconstruction result.
- **Confidence Value** : This indicates the percentage of the primary branches whose *Estimation Error* are within a certain threshold value ϵ and is given as follows:

$$\text{Confidence Value} = \frac{P}{Q} * 100 \quad (5)$$

Where P is the number of branches with *Estimation Error* less than ϵ and Q is the total number of reconstructed branches.

These three metrics together can provide us with a thorough quantitative understanding of different reconstruction algorithms. However, we believe that some metric indicating the presence of false positives and false negatives in our system would make the evaluation process even more robust. In the future, we plan to design a metric which would take care of the false alarms.

7. Experimental Results and Analysis

As stated previously, goal of this paper is to provide a set of baseline results to define the state of the art on the database and allow researchers to get a sense of what is good performance on the data.

3D reconstruction and modeling is an important step towards robotic pruning. Once we have the accurate information about the trunk and primary branches of a tree, further developments can be made towards automation of dormant pruning.

In this section, we provide initial results of 3D reconstruction of apple trees on our new dataset using the novel method explained in Section 5.4. We have tested and analyzed the method qualitatively and quantitatively on all the trees present in our dataset. Few of those results are illustrated in this paper. From Fig. 3, it can be observed that the reconstructed trees look very similar to the ground truth data.

To quantitatively analyze the performance of these algorithms on our dataset, we have used the evaluation methodologies described in Section 6. Some performance evaluation statistics are displayed in Table 2 and Fig. 4 to depict the usage of our dataset. Table 2 summarizes the performance of the stated algorithm for 3 test trees. Figure 4 provides a comparison between the ground truth diameters and estimated diameters of the primary branches of these 3 trees in the dataset. It can be seen that the diameters estimated using our methods are very close to the actual diameter values; thus, the corresponding estimation errors are

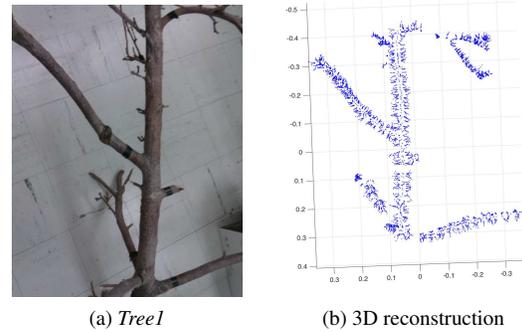


Figure 3: The groundtruth image of the tree ‘*Tree1*’ is shown in (a) and its 3D reconstruction using *Incremental Mean* algorithm mentioned in [8] is displayed in (b).

very small. Specifically, the mean estimation errors are 3.6 mm, 2.7 mm, and 3.3 mm for *Tree1*, *Tree3*, and *Tree6*, respectively.

8. Conclusion and Future Work

This paper introduces a new database comprising of images of dormant apple trees. The images were collected using a lightweight, portable Kinect2 sensor. Our dataset contains five different information for 9 different dormant apple trees which can be used for the reconstruction purpose — labeled groundtruth images, color as well as depth images, groundtruth diameter measurements of the primary branches, and the relative distance between a consecutive pair of primary branches. The database can be useful for different research communities that deal with 3D reconstruction and modeling of dormant trees for automatic pruning purpose, and can also serve as a common baseline for evaluation of different algorithms designed for such applications. To the best of our knowledge, this is the first open dataset available for automatic pruning purpose. In the future, we plan to include more trees in this database.

Acknowledgment

This research is funded by the United State Department of Agriculture. We would like to thank Prof. Peter Hirst’s group at Purdue University, Prof. James Schupp’s group at Pennsylvania State University, and the Bear Mountain Orchards at Aspers, PA, for their help in the tree scanning process.

References

- [1] Developing with Kinect for Windows. <https://dev.windows.com/en-us/kinect/develop>. [Online; accessed 17-Feb-2016].

(Tree ID, NI)	Confidence Values (%)			Mean Estimation Error (mm)	BIA (%)
	$\epsilon = 3$ mm	$\epsilon = 5$ mm	$\epsilon = 7$ mm		
(Tree1, 30)	55.5	77.8	77.8	3.6	100.0
(Tree3, 60)	54.5	72.3	81.82	2.7	81.82
(Tree6, 115)	38.9	66.6	66.6	3.3	72.20

Table 2: The first column represents the tree ID with number of depth images (NI) used to form reconstruction using *Incremental Mean* approach. The next three columns indicate the confidence values in estimating the diameters of each tree with a tolerance less than $\epsilon = 3, 5,$ and 7 mm, respectively. Note that the reconstructions are obtained using depth images of front side bottom section of the tree. Mean Estimation Error and BIA for each tree are given in the last two columns of the table, respectively.

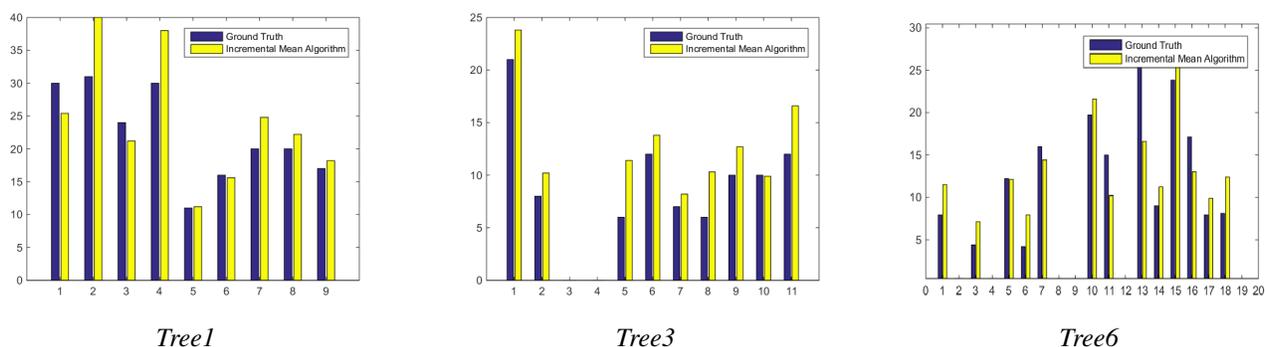


Figure 4: The groundtruth diameters (shown in blue color) and the estimated diameters (shown in yellow color) of the primary branches of the *Tree1*, *Tree3*, and *Tree6*, using *Incremental Mean* approach are shown. Note that we have not shown the bars for branches that are not reconstructed by the algorithm. For example, branches 3 and 4 are not reconstructed in *Tree3*.

- [2] MeshLab. <http://meshlab.sourceforge.net/>. [Online; accessed 17-Feb-2016].
- [3] C# .Net: Fastest Way to Read Text Files. <http://cc.davelozinski.com/c-sharp/fastest-way-to-read-text-files>, 2013. [Online; accessed 17-Feb-2016].
- [4] B. Adhikari and M. Karee. 3d reconstruction of appletrees for mechanical pruning. 2012.
- [5] S. Akbar, N. Elfiky, and A. Kak. A novel framework for modeling dormant apple trees using single depth image for robotic pruning application. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2016.
- [6] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.
- [7] J.-L. Blanco-Claraco, F.-Á. Moreno-Dueñas, and J. González-Jiménez. The Málaga urban dataset: High-rate stereo and lidar in a realistic urban scenario. *The International Journal of Robotics Research*, 33(2):207–214, 2014.
- [8] S. Chattopadhyay, S. Akbar, N. Elfiky, H. Medeiros, and A. Kak. Measuring and modeling apple trees using time-of-flight data for automation of dormant pruning applications. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
- [9] M. Cordts, M. Omran, S. Ramos, T. Scharwächter, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, volume 3, 2015.
- [10] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 304–311. IEEE, 2009.
- [11] N. Elfiky, S. Akbar, J. Sun, J. Park, and A. Kak. Automation of dormant pruning in specialty crop production: An adaptive framework for automatic reconstruction and modeling of apple trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 65–73, 2015.
- [12] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–

- 338, 2010.
- [13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, page 0278364913491297, 2013.
- [14] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [15] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.
- [16] S. Jin, L. Fan, Q. Liu, and R. Lu. Novel calibration and lens distortion correction of 3d reconstruction systems. In *Journal of Physics: Conference Series*, volume 48, page 359. IOP Publishing, 2006.
- [17] M. Karkee, B. Adhikari, S. Amatya, and Q. Zhang. Identification of pruning branches in tall spindle apple trees for automated pruning. *Computers and Electronics in Agriculture*, 103:127–135, 2014.
- [18] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1817–1824. IEEE, 2011.
- [19] Y. Livny, F. Yan, M. Olson, B. Chen, H. Zhang, and J. El-Sana. Automatic reconstruction of tree skeletal structures from point clouds. In *ACM Transactions on Graphics (TOG)*, volume 29, page 151. ACM, 2010.
- [20] L. D. Lopez, Y. Ding, and J. Yu. Modeling complex unfoliated trees from a sparse set of images. In *Computer Graphics Forum*, volume 29, pages 2075–2082. Wiley Online Library, 2010.
- [21] T. Mallick, P. P. Das, and A. K. Majumdar. Characterizations of noise in kinect depth images: a review. *Sensors Journal, IEEE*, 14(6):1731–1740, 2014.
- [22] S. Meister, S. Izadi, P. Kohli, M. Hämmerle, C. Rother, and D. Kondermann. When can we use kinectfusion for ground truth acquisition? In *Proc. Workshop on Color-Depth Camera Fusion in Robotics*, volume 2, 2012.
- [23] H. Müller and T. Tsirikas. Global pattern recognition: The imageclef benchmark. *IAPR Newsletter*, 32(1):3–6, 2010.
- [24] C. V. Nguyen, S. Izadi, and D. Lovell. Modeling kinect sensor noise for improved 3d reconstruction and tracking. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 524–530. IEEE, 2012.
- [25] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. T. Lee, S. Mukherjee, J. Aggarwal, H. Lee, L. Davis, et al. A large-scale benchmark dataset for event recognition in surveillance video. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3153–3160. IEEE, 2011.
- [26] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3):157–173, 2008.
- [27] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *Pattern Recognition*, pages 31–42. Springer, 2014.
- [28] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [29] I. Shlyakhter, M. Rozenoer, J. Dorsey, and S. Teller. Reconstructing 3d tree models from instrumented photographs. *IEEE Computer Graphics and Applications*, (3):53–61, 2001.
- [30] Y. B. Sinai. Skeletonization using voronoi. <http://www.mathworks.com/matlabcentral/fileexchange/27543-skeletonization-using-voronoi>, 2010 (accessed October, 2015).
- [31] J. Smisek, M. Jancosek, and T. Pajdla. 3d with kinect. In *Consumer Depth Cameras for Computer Vision*, pages 3–25. Springer, 2013.
- [32] A. Staranowicz, G. R. Brown, F. Morbidi, and G. L. Mariottini. Easy-to-use and accurate calibration of rgb-d cameras from spheres. In *Image and Video Technology*, pages 265–278. Springer, 2014.
- [33] A. N. Staranowicz, G. R. Brown, F. Morbidi, and G.-L. Mariottini. Practical and accurate calibration of rgb-d cameras using spheres. *Computer Vision and Image Understanding*, 137:102–114, 2015.
- [34] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 573–580. IEEE, 2012.
- [35] J. Sturm, S. Magnenat, N. Engelhard, F. Pomerleau, F. Colas, D. Cremers, R. Siegwart, and W. Burgard. Towards a benchmark for rgb-d slam evaluation. In *RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf.(RSS)*, 2011.
- [36] A. Tabb. Three-dimensional reconstruction of fruit trees by a shape from silhouette method. In *2009 Reno, Nevada, June 21-June 24, 2009*, page 1. American Society of Agricultural and Biological Engineers, 2009.
- [37] Q. Wang and Q. Zhang. Three-dimensional reconstruction of a dormant tree using rgb-d cameras. In *2013 Kansas City, Missouri, July 21-July 24, 2013*, page 1. American Society of Agricultural and Biological Engineers, 2013.
- [38] J. Wilm. Iterative closest point. <http://www.mathworks.com/matlabcentral/fileexchange/27804-iterative-closest-point>, 2013 (accessed October, 2015).
- [39] L. Yang, P. Luo, C. Change Loy, and X. Tang. A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3973–3981, 2015.
- [40] M. Zhang, J. Choi, K. Daniilidis, M. Wolf, and C. Kanan. Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 10–16, 2015.