In Situ Width Estimation of Biofuel Plant Stems

Arda Sahiner, Franklin Heng, Adith Balamurugan, Avideh Zakhor; University of California, Berkeley; Berkeley, CA

Abstract

Efficient plant phenotyping methods are necessary in order to accelerate the development of high yield biofuel crops. Manual measurement of plant phenotypes, such as width, is slow and error-prone. We propose a novel approach to estimating the width of corn and sorghum stems from color and depth images obtained by mounting a camera on a robot which traverses through plots of plants. We use deep learning to detect individual stems and employ an image processing pipeline to model the boundary of each stem and estimate the pixel and metric width of each stem. This approach results in 13.5% absolute error in the pixel domain on corn averaged over 153 estimates and 13.2% metric absolute error on phantom sorghum averaged over 149 estimates.

1. Introduction

Understanding the relationship between genotypes and phenotypes of plants is essential for clean energy and optimizing biofuel production. By collecting physiological traits of plants, it is possible to find links between plant gene sequences and biomass yield. Sorghum and corn plants have been demonstrated to be suitable sources for fuel in practice [13, 2]. To determine the genotype-phenotype map for sorghum and corn, rapid phenotyping methods are necessary for efficient data collection. Currently, measurement of plant phenotypes, such as width, is often done manually. This process is slow and error-prone. Therefore, it is essential to develop algorithms that can accurately and efficiently phenotype in situ plants [11].

We propose a practical, robust method that phenotypes sorghum and corn grown in outdoor conditions from RGB and depth images. Specifically, we develop an algorithm to estimate the width of plant stems using an Intel RealSense R200 camera, which is a stereo camera collecting RGB, infrared, and stereoscopic depth images at a relatively low cost [9]. This camera is mounted onto a mobile robotic platform, which traverses through rows of densely positioned plants.

Our proposed algorithm detects individual plant stems, estimates the width of each stem in pixels, and converts these pixel widths to metric widths using depth data. For detection, we train a Region-based Convolutional Neural Network (R-CNN) to propose bounding boxes for individual stems. Given a region within which a stem lies, we employ edge detection, multi-resolution morphological operations, and Random Sample Consensus (RANSAC) in order to model the boundary of each stem and estimate stem widths.

In our experimental setup, a robot with a mounted stereo camera moves in the space between two rows of a given plot, collecting images at a fixed rate, as shown in Figure 1. Each plot contains a unique genetic plant strain, and since geneticists are primarily interested in comparing statistical parameters of phenotypes across plots, our goal is to estimate the histogram of stem widths for each plot. As such, we have designed our algorithm to identify and discard width estimates with low confidence values so as not to adversely affect the statistical characterization of a plot. Since each row of a plot has 50 plants, discarding low confidence estimates is unlikely to significantly alter the estimated histogram.

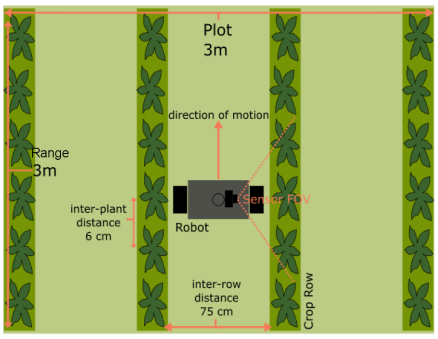


Figure 1. *Sensor setup for measuring in situ plants.*

2. Related Work

Previous work to estimate widths of biofuel plant stems in outdoor settings models the shape of stems in order to produce width estimates. Notably, [1] develops a width estimation algorithm from 2.5D infrared images by employing Frangi filters to locate the tubular, narrow shapes of stems within an image and by applying a Hough transform to compute the lines that outline the boundary of each stem.

Similar to [1], we propose a method which first identifies individual stems in each image, and then models the boundary of each stem using two lines. However, rather than hand-tuning an image transformation to accentuate plant stems, we implicitly learn this transformation through deep learning on known data, similar to [4]. We also deploy RANSAC, which has been shown to outperform Hough transforms in cases involving line-fitting from noisy observations, in both accuracy and computational speed [15, 10].

Rather than using images, some plant phenotyping methods model stems by constructing 3D point clouds from data obtained from a time-of-flight (ToF) sensor [8, 7]. Despite promising results, this method is limited by the cost of 3D ToF sensors, as well as the computational cost of generating and operating on point clouds.

Previous works phenotype one type of plant, with the most emphasis on phenotyping sorghum [1, 8, 12, 3]. The disadvantage of this approach is that it may not generalize to multiple kinds of plants, which is important for ease and robustness of use. Thus, we provide results on both corn and sorghum.

The rest of the paper is organized as follows: in Section 3, we outline our proposed approach; in Section 4, we describe the

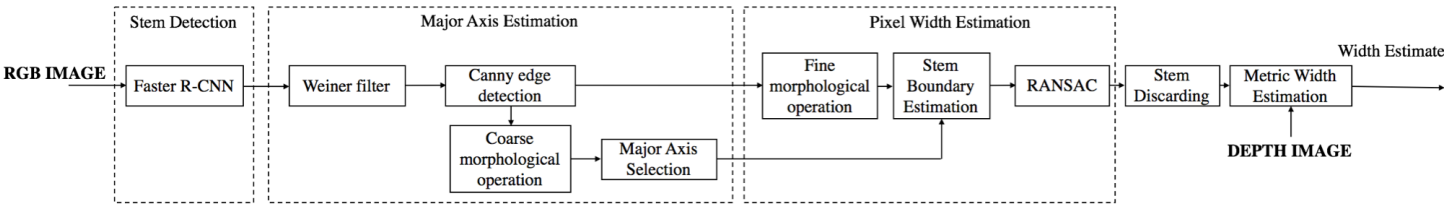


Figure 2. *Diagram of the proposed approach. Inputs from the stereo camera are indicated in bold.*

experimental setup and results of our algorithm on two datasets; and in Section 5, we discuss implications of our findings and future work.

3. Proposed Approach

As shown in Figure 2, our proposed algorithm consists of four main stages: stem detection, major axis detection, pixel width estimation, and metric width estimation. The stem detection stage uses Faster R-CNN, a Convolutional Neural Network that can predict bounding boxes around objects of interest, in order to identify general locations of individual stems [14]. Second, the major axis estimation stage applies a low-resolution morphological closing operation to the edge profile of an individual stem in order to estimate the major axis of the stem. Third, the pixel width estimation stage uses points on either side of the stem and fits them to lines on both sides of the stem using RANSAC. These lines are evaluated to determine whether they are nearly parallel, and if so they are used to generate an estimate for the pixel width of the stem; otherwise, no estimate is generated for the stem. Lastly, depth data and knowledge of the camera parameters are used to generate a metric width estimate for the given stem.

3.1 Stem Detection with Faster R-CNN

Since we aim to estimate the width of each individual plant, our algorithm first isolates individual stems. This is achieved by feeding RGB images captured by the stereo camera, each containing one or more plants, into a pre-trained Faster R-CNN model [14]. The novelty of Faster R-CNN is best identified by the design of its deep convolution layers, which are used to deduce region proposals.

We leverage an existing pre-trained model, Faster R-CNN with ResNet101 [18], and fine-tune it with 2000 images of corn and sorghum crops. The Faster R-CNN outputs a list of bounding box coordinates and corresponding confidence levels corresponding to regions of interest which indicate plant stems, as shown in Figure 3. We use these predictions to crop the original RGB image and its corresponding depth image into areas solely focusing on individual stems. These cropped images of detected stems are fed into the subsequent steps of the algorithm.

Note that in the above process, it is possible to detect stems of the same plant multiple times in multiple images due to the overlap between successive images. It is possible to mitigate this by tracking detected stems across frames and fuse the width estimates of the same stem in multiple pictures in order to arrive at a more accurate estimate. However, since we are mostly interested in statistical characterization of stem widths in a plot containing the same genetic strain with 50 plants, we speculate that ignoring the duplications across images by not tracking them is not likely to result in significant change in the mean stem width estimate for the plot.

Figure 3. *An example of corn stems as detected by Faster R-CNN.*

3.2 Stem Major Axis Estimation

After detecting stems with Faster R-CNN, we estimate the major axis of each stem to represent its location and orientation. First, we de-noise the image with an adaptive low-pass Wiener filter and increase the contrast using histogram equalization to enhance edges, as shown in Figure 4(b) [17]. Next, a Canny edge detector is used to generate an edge profile for each stem, such as the one shown in Figures 4(c) and 5(a). Canny edge detection is well-suited for this since it discourages disconnected edges, which are not of use when representing a stem outline [5].

Given a binary image representation of an edge profile, we would like to extract a larger structure from this image, under the assumption that a stem is the most prominent structure present, surrounded by smaller obfuscating structures such as leaves. Thus, we desire a low-resolution representation of the stem which preserves larger structures. We use a morphological closing operation applied to the complement of the binary image [16]. By using a large, rectangular structuring element of size 30×15, we can extract large structures from the edge profile of the stem, as shown in Figure 5(b).

Assuming the stem is the primary structure in the image, it should be represented by one of the remaining connected components after applying morphological closing. If there are multiple connected components, such as in Figure 5(b), a component that is large, near the center of the image, and near-vertical most likely represents the stem. In practice, this can be chosen by calculating a weighted average of these features for each component and selecting the component with the highest average. Given a chosen component, its location and orientation can be used to obtain a major axis line which resembles the stem's major axis. Figure 5(c) illustrates a connected component closer to the center being chosen to represent the stem, along with a superimposed major

axis line.

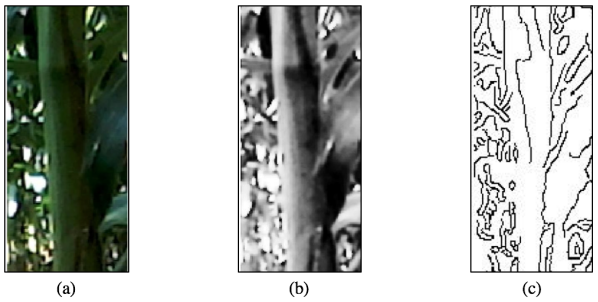


Figure 4. *(a) Bounding box around a stem; (b) result of Wiener filtering and contrast enhancement; (c) Canny edge profile of resulting image.*

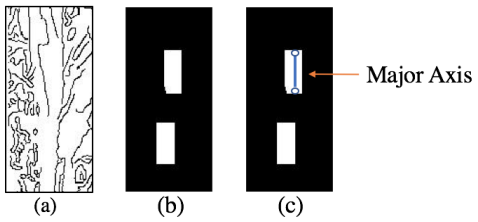


Figure 5. *(a) Canny edge profile of a stem; (b) result of coarse-resolution morphological closing; (c) proposed stem major axis.*

3.3 Stem Boundary Estimation

The boundary of the stem can be represented by edge points on either side of the stem. Since the Canny edge profile demonstrated in Figures 4(a) and 5(a) is too noisy to accurately represent the boundary of the stem, we apply a smaller morphological closing operation of size 12×4 in order to obtain a finer resolution image representation as shown in Figure 6(b). We next superimpose the major axis on this new edge profile and draw lines perpendicular to the axis with equidistant spacing along the length of the axis, such as in Figure 6(c). Points on either side of the axis at which these lines first intersect an edge are chosen as candidate boundary points, such as in Figure 7(a).

3.4 RANSAC and Pixel Width Estimation

Having generated candidate points that delineate the boundary of a stem, we seek to estimate the final width of the stem. Given that these proposed points are derived from a possibly noisy edge profile of the detected stem, it is often the case that some points do not truly represent the boundary of the stem, such as the candidate points on the left side of the stem in Figure 7(a). Thus, we wish remove outliers from the points on both sides of the stem and use the remaining points to represent the stem boundary. Random Sample Consensus (RANSAC) is a robust method for removing outliers and modeling data from remaining inlier points [6]. Given that stems of corn and sorghum have almost no curvature, we represent the boundary of the stem with a line on each side of the stem. Thus, we use RANSAC independently on each side of the stem, finding a linear fit to the data which ignores outlier candidate points, as demonstrated in Figures 7(b) and 7(c). The pixel width of the stem is computed by finding the average distance between the two segments found by RANSAC.

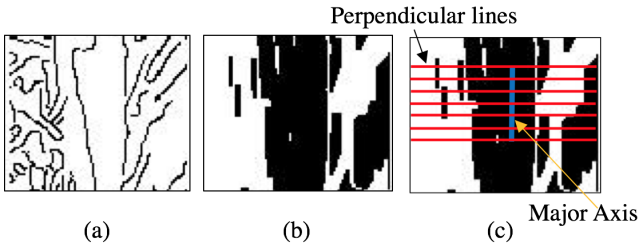


Figure 6. *(a) Canny edge profile of the stem in the region of the major axis; (b) result of applying fine-resolution morphological closing; (c) Perpendicular lines to the major axis drawn along the axis.*

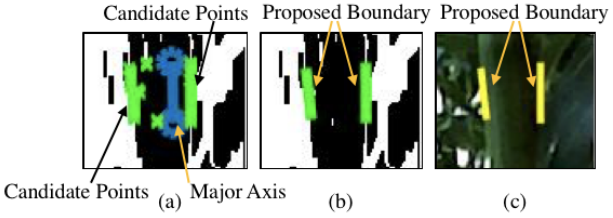


Figure 7. *(a) Candidate points on boundary of the stem; (b) lines proposed by RANSAC superimposed on a fine-resolution morphological representation; (c) lines proposed by RANSAC superimposed on the original RGB image.*

3.5 Discarding Low-Confidence Estimates

The proposed methods rely on the assumption that the edge profile of a stem accurately represents the stem's physical outline. In practice, this may not always hold, due to multiple factors, such as obfuscating leaves or blurry images due to shaky robot movement. A case of a blurry image and its corresponding noisy edge profile is shown in Figures 8(a) and 8(b). In order to detect the presence of these factors, we compute the angle between the two line segments proposed by RANSAC. Since actual sorghum and corn plants should have nearly parallel edges, we can discard our estimated width if the angle between the two lines are sufficiently large, i.e. greater than 5 degrees, such as in Figure 8(c). Additionally, since our final goal is to statistically characterize the stem width of an entire plot of plants of a single genetic strain, even if we do not generate width estimates for all stems in a plot, we can still characterize the phenotype of a given strain.

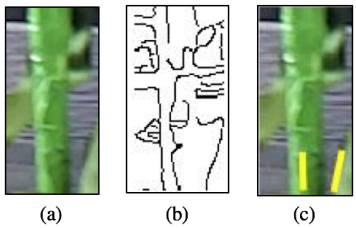


Figure 8. *(a) A blurry RGB image of a stem; (b) Canny edge profile of the stem; (c) proposed stem boundary lines.*

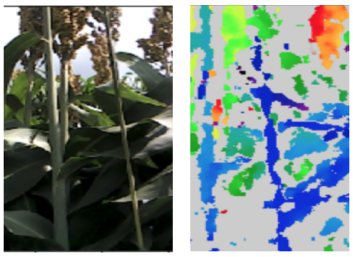
3.6 Metric Width Estimation

Lastly, we determine the metric width of a stem given its boundary and corresponding depth data from the depth camera. RGB images and depth images obtained by the camera may not

be aligned, so we manually align by shifting all depth pixels by some constant value. This shift may vary according to the speed of the robot, so it must be hand-tuned to each individual dataset. An example of the misalignment between RGB images and depth images is shown in Figure 9. Next, to account for noise in depth measurements, we average the depth values of all pixels enclosed by the boundary of the stem to estimate the distance between the stem and the camera. Once an average depth is obtained, this depth, the camera's focal length, the estimated pixel width, and the distance from the stem to the center of the image are used to compute the metric width of the stem with the following formula:

$$w_m = \frac{w_p * d}{f_x}$$

where w_m is the metric width, w_p is the pixel width, d is the distance to the stem, and f_x is the sensor focal length in units of pixels.



(a) (b)

Figure 9. *(a) An RGB image of two stems; (b) the corresponding depth map visualized through a heat map. Note that the two vertical lines in the depth map correspond to the two stems in the RGB image. The stems are seen further to the left of the image in the depth map than the RGB image.*

4. Results

We present two experiments to test different components of the proposed algorithm. First, we use data from in situ corn plants in outdoor conditions to test pixel width estimation methods. Second, we use phantom sorghum plants placed in an outdoor setting to evaluate our metric width estimation.

4.1 Corn Pixel Width Estimation

The data collected for testing corn width estimates was obtained from a RealSense R200 camera mounted on a robot, which traversed through 6 rows of distinct corn plots of 50 plants each, with the setup described in Section 1. Due to difficulties with hand-labeling metric stem widths on a large number of plants, we opt to use ground truth data in the pixel domain by labeling the pixel width of each stem after being detected by Faster R-CNN at three locations along the stem and averaging these measurements, as shown in in Figure 10.

Our pre-trained Faster R-CNN model generated 531 individual bounding box images, from which our algorithm generated pixel width estimates for 153 images, achieving a discard rate of 71%. The results are shown in Table 1. "Average % Error" in column 2 is obtained by calculating the signed percent error of each individual pixel width estimate and averaging this error across images, while "Average % Absolute Error" in column 3 is obtained by determining the absolute percent error of each estimate and averaging this error across images. Despite discarding a large proportion of images, we are able to characterize the width for each stem to a high accuracy.

Figure 10. *An RGB image of a corn stem, with three red lines indicating the locations along the stem in which ground truth measurements in the pixel domain were taken.*

Table 1: Pixel width estimation of detected in situ corn plants.

Plot	Average % Error	Average % Absolute Error	% Discarded
1	8.4	10.6	55 (40/73)
2	8.9	15.8	64 (87/135)
3	3.4	10.9	73 (77/106)
4	3.3	17.8	70 (52/74)
5	-2.7	10.2	85 (64/75)
6	-3.3	13.3	85 (55/65)
All	**5.3**	**13.5**	**71 (378/531)**

4.2 Phantom Sorghum Metric Width Estimation

Since no ground truth metric width measurements were collected in the field, we opt to also evaluate the performance of our algorithm on an additional dataset, obtained by moving a stereo depth camera moved across five closely-positioned phantom sorghum plants in 117 frames, such as in Figure 11. Ground truth metric data was acquired using caliper measurements of each phantom at three locations along the stem and averaging these measurements, while ground truth pixel data was acquired in the same manner as the previous experiment.

Figure 11. *An RGB image of phantom sorghum plants captured by the stereo camera.*

From the 117 captured frames, Faster R-CNN detected 390 portions of stems. From these 390 detected stem images, our algorithm generated pixel and metric width estimates for 149 images, achieving a discard rate of 62%. Individual estimates were then manually matched against individual stems to evaluate the quality of estimates for each stem. Pixel width estimation results are

shown in Table 2, while metric width estimation results are shown in Table 3. As seen, the average % error and average % absolute error for both tables are on par with those in Table 1.

Table 2: Pixel width estimation of detected phantom sorghum plants. "GT" in column 4 stands for ground truth.

Plant	Average % Error	Average % Absolute Error	% Variation in GT	% Discarded
1	-0.1	10.1	4.3	41 (23/56)
2	9.8	9.8	9.8	97 (32/33)
3	-14.3	15.1	7.0	63 (74/118)
4	-6.5	13.1	4.6	48 (39/81)
5	0.2	22.0	6.1	72 (73/102)
All	**-6.0**	**14.7**	**6.4**	**62 (241/390)**

Tabl 3: Metric width estimation of detected phantom sorghum plants. "GT" in column 4 stands for ground truth.

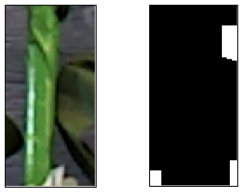
Plant	Average % Error	Average % Absolute Error	% Variation in GT
1	1.4	9.4	1.7
2	0.9	0.9	5.8
3	-12.8	13.7	1.3
4	-7.8	13.6	4.7
5	3.5	16.8	8.0
All	**-5.0**	**13.2**	**4.3**

Percent variation in ground truth in Tables 2 and 3 is defined as the average percent absolute difference of each individual measurement from the mean of all three measurements of a stem. From Table 2, we see that even in the portion of a stem that is detected by Fast R-CNN, there is high variability in the stem's pixel width in the ground truthing process. Additionally, from Table 3, it is clear that along an entire individual stem, there are large variations in ground truth metric width. Thus, it appears likely that a significant proportion of our estimation error can be explained by variability of the ground truth width measurement of an individual stem, both for pixel and metric width estimates.

As shown in Table 2, for plant 2, only 1 estimate was generated out of the 33 different detected instances. Frequently discarding was due to an obfuscating leaf, which negatively impacts major axis detection, an essential component for accurate width estimation, as shown in Figure 12. However, the single estimate generated for plant 2 is accurate, indicating that we can accurately phenotype plants even when a large proportion of images is discarded.

6. Discussion

The results in Tables 1, 2, and 3 confirm that our algorithm performs as well on phantom sorghum as it does on actual corn plants. In comparing our results in the pixel domain to those in the metric domain, we also find that the vast majority of our error is in the pixel width estimation step, specifically in detecting the boundary of the stem. As a result, future work could develop more modern approaches to detecting stem boundaries, such as a deep learning approach to segmentation. In addition, manual depth-RGB alignment must also be replaced by automatic methods.



(a) (b)

Figure 12. *(a) Image of plant 2 with leaf on the right side; (b) coarse-resolution morphological representation.*

7. Acknowledgments

The information, data, and work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000598. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] Tavor Baharav, Mohini Bariya, and Avideh Zakhor. In situ height and width estimation of sorghum plants from 2.5d infrared images. *Electronic Imaging*, 2017:122–135, 2017.

[2] Mustafa Balat and Havva Balat. Recent trends in global production and utilization of bio-ethanol fuel. *Applied energy*, 86(11):2273–2282, 2009.

[3] Jose Batz, Mario A Méndez-Dorado, and J Alex Thomasson. Imaging for high-throughput phenotyping in energy sorghum. *Journal of Imaging*, 2(1):4, 2016.

[4] Harjatin Singh Baweja, Tanvir Parhar, Omeed Mirbod, and Stephen Nuske. Stalknet: A deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. In *Field and Service Robotics*, pages 271–284. Springer, 2018.

[5] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[6] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[7] William Gélard, Philippe Burger, Pierre Casadebaig, Nicolas Langlade, Philippe Debaeke, Michel Devy, and Ariane Herbulot. 3d plant phenotyping in sunflower using architecture-based organ segmentation from 3d point clouds. In *5th International Workshop on Image Analysis Methods for the Plant Sciences*, 2016.

[8] Jihui Jin and Avideh Zakhor. Point cloud based approach to stem width extraction of sorghum. *Electronic Imaging*, 2017:148–155, January 2017.

[9] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. Intel realsense stereoscopic depth cameras. *CoRR*, abs/1705.05548, June 2017.

[10] M Kröger, W Sauer-Greff, R Urbansky, M Lorang, and M Siegrist. Performance evaluation on contour extraction using hough transform and ransac for multi-sensor data fusion applications in indus-

trial food inspection. In *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2016*, pages 234–237. IEEE, 2016.

[11] M. Minervini, H. Scharr, and S. A. Tsaftaris. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, 32(4):126–131, July 2015.

[12] Seth C Murray, Leighton Knox, Brandon Hartley, Mario A Méndez-Dorado, Grant Richardson, J Alex Thomasson, Yeyin Shi, Nithya Rajan, Haly Neely, Muthukumar Bagavathiannan, et al. High clearance phenotyping systems for season-long measurement of corn, sorghum and other row crops to complement unmanned aerial vehicle systems. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*, volume 9866, page 986607. International Society for Optics and Photonics, 2016.

[13] S Prasad, Anoop Singh, N Jain, and HC Joshi. Ethanol production from sweet sorghum syrup for utilization as automotive fuel in india. *Energy & Fuels*, 21(4):2415–2420, 2007.

[14] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[15] Fayez Tarsha-Kurdi, Tania Landes, and Pierre Grussenmeyer. Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. In *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, volume 36, pages 407–412, 2007.

[16] Alexander Toet. A morphological pyramidal image decomposition. *Pattern recognition letters*, 9(4):255–261, 1989.

[17] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA, 1949.

[18] Eniac Xie. Faster-rcnn-resnet. https://github.com/Eniac-Xie/faster-rcnn-resnet, 2017.

Author Biography

Avideh Zakhor is Qualcomm Chair and Professor in EECS at U.C. Berkeley. She has won a number of best paper awards, including the IEEE Signal Processing Society, Circuits and Systems Society and IEEE Solid Circuits Society in 2008. She was a Hertz fellow, General Motors scholar, IEEE fellow, received the Presidential Young Investigators (PYI) award in 1990, and was selected as Electronic Imaging scientist of the year by SPIE in 2018.

Arda Sahiner is currently pursuing a B.S. in Electrical Engineering and Computer Science at the University of California, Berkeley. His current research focuses on signal processing, optimization, and machine learning for a variety of applications.

Franklin Heng received his bachelor's degree in Computer Science from University of California, Berkeley in 2018. He is interested in various applications of computer vision and image processing. His current research projects pertain to patient care and national defense.

Adith Balamurugan is currently pursuing a MS in Electrical Engineering and Computer Science at the University of California, Berkeley. He received his bachelor's degree in both Computer Science and Statistics from UC Berkeley in 2018. His current research focuses on few-shot and one-shot learning algorithms for object detection, targeting indoor spaces.