Crop Row Detection and Waypoint Definition Using Aerial Images of Fields

Alireza Amiri
Dept. Mechatronics Engineering
Khaje Nasir Toosi University of Technology (KNTU)
Tehran, Iran
ali.amiri@email.kntu.ac.ir

Dr. Saeed Khankalantary
Dept. Mechatronics Engineering
Khaje Nasir Toosi University of Technology (KNTU)
Tehran, Iran
s.kalantary@kntu.ac.ir

Abstract—As autonomous agriculture evolves, using wheeled mobile robots for various tasks necessitates precise waypoint generation to define the robots' paths accurately. This paper introduces a method that leverages aerial imagery to detect crop positions and determine waypoints. A specialized hardware setup, consisting of a high-resolution camera, wireless transmitter, and receiver, is developed to capture and transmit live images of the agricultural field. In the image processing stage, crops are identified through two parallel techniques—Unet and K-means clustering. Subsequently, the Hough Transform is applied to detect crop row lines, refined through filtering to ensure that a single, accurate line represents each row. Finally, by selecting specific points on the paths between these rows and converting them into global coordinates, the system facilitates real-time crop detection and precise waypoint generation, supporting autonomous navigation for agricultural robots.

Index Terms—Crop Row Detection, Hough Transform, Image Processing

I. INTRODUCTION

The rising need for autonomous agriculture systems is driven by their potential to enhance farm productivity and sustainability significantly. These systems increase efficiency and crop yield by automating critical tasks such as crop monitoring, harvesting, and treatment application, making them essential in modern farming practices. Additionally, they contribute to sustainable food production by optimizing resource usage, reducing waste, and minimizing environmental impact. The capability of these systems to make real-time decisions based on changing field conditions further underscores their importance, primarily as they address the challenges posed by complex and varied agricultural environments.

Mobile robots offer significant advantages in autonomous agriculture by enhancing flexibility, precision, and operational efficiency. Their ability to navigate diverse terrains and crop layouts provides the flexibility needed for efficient field operations. Equipped with advanced vision systems, these robots ensure precise navigation and targeted treatments, improving productivity and resource utilization. The autonomous operation of mobile robots reduces manual labor, lowering operational costs and increasing efficiency. Their real-time monitoring capabilities also enable timely interventions for crop management tasks such as weeding and spraying. The scalability and adaptability of mobile robots further make them a versatile solution, suitable for various tasks and farm sizes in modern agriculture.

Designing a trustworthy and accurate navigation system is crucial for mobile robots to perform tasks in agricultural fields effectively. These robots must navigate complex field environments, including varying terrains and non-uniform crop layouts, to precisely execute operations such as planting, weeding, and harvesting. Without a reliable navigation system, the efficiency and effectiveness of these robots are compromised, leading to potential errors in task execution and resource wastage. Therefore, developing robust navigation systems to ensure accurate path planning and obstacle avoidance is essential for successfully deploying autonomous mobile robots in agriculture.

Various methods and algorithms are employed for effective crop detection in agricultural fields. Traditional image processing techniques include linear regression, the Hough Transform, and object-by-object image analysis (OBIA), which detect crop rows and individual plants. Machine learning approaches, such as support vector machines (SVM) and random forests (RF), classify crops and weeds by analyzing pixel intensities, textures, shapes, and geometric features in field images. Deep learning methods, especially convolutional neural networks (CNNs), offer advanced capabilities in crop row detection by automatically learning complex features from large datasets, thereby improving accuracy. These methods collectively contribute to the precise identification of crop rows and plants, enhancing the efficiency of agricultural automation.

II. METHODOLOGY

A. Data Acquisition

The initial step in this research involves acquiring essential environmental data, including aerial images and the global coordinates of the camera at the time of image capture. This data is crucial for converting the local pixel positions in the images into global coordinates. Various cameras, such as Sequoia, RedEdge, Micasense, and DJI Zenmuse X7, were considered based on their technical properties and available resources. The GoPro Hero 4 camera was selected for this project because it captures high-resolution RGB images at a high frame rate. Although it lacks NIR and thermal capabilities, RGB images are sufficient for the image processing algorithms used in

this study. In addition to aerial images, global coordinates are obtained using a Ublox-Neo-6m GPS module, which transmits the data from the camera's location to the processing unit. The camera's height is predefined rather than measured and transmitted in real-time to simplify the setup.

B. Image Processing

This section focuses on the crucial task of crop detection within the captured images, which is essential for identifying crop rows and determining the path for subsequent operations. The process begins with a preprocessing step that converts the raw images into a format compatible with the models used for detection. Two crop detection methods, K-means and Unet, are then implemented, with a detailed explanation of each approach. The performance of these models is compared to select the most effective one for this application. Additionally, insights from previous studies using these algorithms inform the analysis.

1) Preprocessing: High-resolution aerial images covering large areas are typically too large to use in machine learning algorithms directly. To address this, the images are initially split into equal-sized sections, making them manageable for algorithmic processing. The size of these sections is determined by the specific requirements of the algorithms used, ensuring compatibility and optimized performance. Maintaining uniform image sizes across all sections enhances processing speed and efficiency.

2) Crop Detection: This project's most computationally intensive aspect is crop detection, which is crucial in ensuring accurate outcomes for subsequent procedures. A well-defined layer mask for field images is essential, as the crop detection system must effectively distinguish plants and crops from other objects, including background soil and unwanted vegetation such as grass and weeds.

a) Color Filtering and Based Image Segmentation: By analyzing aerial images of crop fields, the idea emerged of applying color filtering to isolate areas within a specific range of green, corresponding to vegetation. Given that the input images are in RGB format, the initial approach involved using only the green channel to filter out pixels based on their green intensity. However, this method proved unsatisfactory for several reasons. First, white pixels containing high green intensity could not be effectively filtered using just the green channel. Additionally, areas with a more yellowish hue were incorrectly excluded by this filter. To address these issues, the RGB images were converted to the HSV (Hue, Saturation, and Value) format, which allows for more precise color filtering based on hue. The vegetation color range was then defined and applied to the images, resulting in improved segmentation of crops from the soil. The output of this process was a binary black-and-white mask, where white pixels represented areas containing plants.

A binary K-means clustering algorithm was developed and applied to group the white pixels to refine the layer masks further, effectively clustering vegetation areas. Through an

iterative process, the algorithm parameters were tuned to achieve optimal results.

b) Unet-Based Image Segmentation: Although the color filtering method demonstrated satisfactory performance, it struggled to distinguish crops from other vegetation, such as grass or weeds, present in the field. This limitation arises because the algorithm passes all vegetation through the color filter, disrupting crop detection. These challenges highlight the inadequacy of color filtering for this task and underscore the need to utilize machine-learning models that intelligently differentiate crops from other vegetation.

Access to a suitable dataset is essential to train a machine learning model. For this purpose, the vineyard aerial images dataset [REF] was employed, which includes annotated layer masks indicating the positions of plants within the images. The images were initially split into equal-sized segments of 128 by 128 pixels to ensure consistency in the input to the model, eliminating the need for resizing during processing. Subsequently, a Unet model was defined and compiled based on the architecture outlined in [REF]. The model achieved a final accuracy of 96

C. Crop Row Detection

Crop row detection involves assigning a line with a defined slope and intercept to each crop row. Based on the binary mask of segmented crops obtained in the previous step, lines must be fitted to the pixels to minimize the least square error from the line. This study analyzed and implemented two approaches—linear regression and the Hough Transform—for this purpose.

1) Linear Regression: This experiment used the binary layer mask of segmented crops to assign each pixel to its corresponding crop row. The first step involved determining the optimal angle of the crop rows through an iterative algorithm, assuming that all rows are parallel and share the same slope. Once the slope was defined, the x and y coordinates of the white pixels were rotated to align the crop rows vertically, simplifying the clustering process.

At this stage, the K-means algorithm was applied with a customized loss function designed to minimize the least square error of the distances between the white pixels and a vertical line. Unlike standard loss functions that measure the distance from a centroid point, this approach considered each cluster centered around a line, not a point.

Given the variability in the number of clusters across different images, the clustering algorithm is initiated by grouping nearby pixels. When a pixel was too distant from the existing group, it was treated as the center of a new cluster. However, this approach led to the formation of numerous unwanted clusters. To address this, adjacent clusters were merged in a subsequent step. Finally, the initial rotation of the pixel coordinates was reversed to visualize the resulting clusters. As illustrated in the results, even after merging close clusters, the method failed to accurately identify crop rows, leading to complications in the later stages of the project. To conclude the experiment, linear regression was applied to fit a line to the

pixels within each cluster. However, the results indicated that this clustering and linear regression approach is ineffective for defining crop rows, necessitating exploring alternative methods.

2) Hough Transform: Another widely used solution for crop row detection is the Hough Transform. This computer vision technique is primarily designed to identify geometric shapes in images, making it particularly effective for detecting lines and curves. In this project, the Hough Transform was employed as an alternative to the linear regression model for predicting crop rows. Implementing this method requires significantly less preprocessing and image manipulation, mainly due to the availability of related packages in OpenCV. When the Hough Transform was applied to the images, multiple lines were detected for each crop row, as illustrated in the results. These lines successfully covered the crop rows, indicating the method's effectiveness in line detection. To refine the results and assign only one line per crop row, the detected lines were merged, and a single line was calculated using the average slope and intercept of the detected lines. The final results are presented in FIGURE.

D. Path Waypoint Detection

Defining waypoints becomes straightforward once the crop rows have been established. Initially, a path is defined as a line parallel to the crop rows and equidistant from two neighboring crop rows. A predefined number of equally spaced points are selected along each path line. The coordinates of these points are recorded for use in subsequent steps. FIGURE

E. Reconstruction of Aerial Image

Following the image splitting described in section B.1, the image processing tasks outlined earlier will be applied to each segment of the leading aerial image. The goal is to determine and record the waypoint coordinates within each segment. Once the waypoint coordinates are computed, the initial image must be reconstructed and assembled. Subsequently, the waypoint coordinates need to be converted to align with the coordinates of the newly assembled image.

F. Image Coordination to Global Coordination Conversion

An analytical approach and mathematical algorithm are required to obtain the corresponding global coordinates of the points identified in the images. This process involves using the global coordinates of the camera at the time of image capture, along with the camera's height. The conversion process is broken down into two phases: first, converting pixel coordinates into meter distances, and second, converting these meter distances into the global coordinate system.

1) Pixel to Meter Conversion: Assuming the camera is positioned at the center of the image at a height of h, and given the angle θ (defined as the angle between the standard line from the camera to the ground and the line connecting the camera to the edge of the image), the conversion from pixel position to ground distance is described by the following equations:

$$\tan\theta = \frac{x}{h} \implies x = \frac{L}{2} \cdot \frac{h}{\tan\theta}$$

where x is the distance from the camera to the edge of the image in meters.

If the image length in pixels is denoted as L, then:

$$1 \text{ pixel} = 2h \cdot \frac{\tan\theta}{L} \text{ (m)}$$

Using these equations, the position of a pixel with coordinates (x, y) in the image can be determined relative to the camera's position on the ground.

2) Meter to Global Coordination Conversion: The final step involves converting the meter distances obtained from the previous calculations into global coordinates. The relationship between meters and degrees of latitude or longitude is given by:

$$1 \text{ (m)} = 0.00001°$$

This conversion factor enables the translation of local pixel positions, measured in meters, into global coordinates.

III. CONCLUSION

In response to the increasing demand for autonomous agricultural systems, there is a critical need for accurate and reliable waypoints for navigation. This paper presents a solution involving a comprehensive live image capturing, processing, and waypoint generation system. The system is divided into three main sections: data acquisition, which captures and transmits aerial images; image processing, which identifies crops, determines crop rows, and assigns waypoints; and global coordinate conversion, which translates local waypoint coordinates into global coordinates. This approach ensures that the waypoints are precise and suitable for use by mobile robots or other devices requiring accurate navigation within agricultural fields.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.