

Article

Harvesting Route Detection and Crop Height Estimation Methods for Lodged Farmland Based on AdaBoost

Yanming Li ^{1,2,*}, Yibo Guo ¹, Liang Gong ^{1,2}  and Chengliang Liu ^{1,2}

¹ School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; dudugyb@sjtu.edu.cn (Y.G.); gongliang_mi@sjtu.edu.cn (L.G.); chlliu@sjtu.edu.cn (C.L.)

² Key Laboratory of Intelligent Agricultural Technology (Yangtze River Delta), Ministry of Agriculture and Rural Affairs, Shanghai 200240, China

* Correspondence: yml@sjtu.edu.cn

Abstract: Addressing the challenge of the current harvester route detection method's reduced robustness within lodging-affected farmland environments and its limited perception of crop lodging, this paper proposes a harvesting operation image segmentation method based on **SLIC superpixel segmentation** and the **AdaBoost ensemble learning algorithm**. This segmentation enables two essential tasks. Firstly, the RANSAC algorithm is employed to extract the harvester's operational route through straight-line fitting from the segmented image. Secondly, the method utilizes a 3D point cloud generated by binocular vision, combined with IMU information for attitude correction, to estimate the height of the harvested crop in front of the harvester. Experimental results demonstrate the effectiveness of this method in successfully segmenting the harvested and unharvested areas of the farmland. The average angle error for the detected harvesting route is approximately 1.97°, and the average error for crop height detection in the unharvested area is around 0.054 m. Moreover, the algorithm exhibits a total running time of approximately 437 ms. The innovation of this paper lies in its simultaneous implementation of two distinct perception tasks, leveraging the same image segmentation results. This approach offers a robust and effective solution for addressing both route detection and crop height estimation challenges within lodging-affected farmland during harvesting operations.



Citation: Li, Y.; Guo, Y.; Gong, L.; Liu, C. Harvesting Route Detection and Crop Height Estimation Methods for Lodged Farmland Based on AdaBoost. *Agriculture* **2023**, *13*, 1700. <https://doi.org/10.3390/agriculture13091700>

Academic Editor: Roberto Alves Braga Júnior

Received: 28 July 2023

Revised: 22 August 2023

Accepted: 23 August 2023

Published: 28 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Unmanned operation is the current trend in the development of intelligent agricultural machinery [1]. Whether high-quality and high-efficiency operation directly determines whether unmanned agricultural machinery can be popularized and applied, and the real-time sensing detection of crop growth state is the premise of high-quality operation of unmanned agricultural machinery [2]. Harvesting is an important part of rice production. In the rice harvesting process, the harvested area and the area for harvesting are similar in color, different areas have high repeatability and rice lodging often exists, which brings challenges to the real-time detection and recognition of the state of unmanned rice harvesting. In automatic rice harvesting operations, the distinction between harvesting area and unharvested area, line recognition, and height detection of fallen crops have attracted the interest of many researchers.

During the rice harvest period, occurrences of crop lodging are frequent. This phenomenon is typically triggered by agronomic variables such as cultivation practices and external natural forces like wind and rain [3]. The extent of lodging can vary significantly. An illustrative image of lodging-affected farmland is provided in Figure 1. On the one hand, lodging-affected crops exhibit distinct image characteristics and varying heights compared to their non-lodging counterparts. Most of the existing harvester route extraction methods,

designed for conventional harvesting scenarios, are not suitable for lodging conditions. Moreover, studies and experimental research dedicated to lodging scenarios are relatively limited. On the other hand, parameters such as breakage rate, impurity rate, and stubble height are critical performance indicators in rice harvesting activities. Generally, in the context of harvesting non-lodging crops, the harvester's header is positioned at a relatively elevated level. Conversely, when encountering lodging-affected crops, it becomes essential to lower the header's height appropriately to minimize harvesting losses [4,5]. Presently, the adjustment of header height is predominantly reliant on manual observation of crop conditions and empirical height estimation by the operator. Unfortunately, there lack a mature real-time crop height estimation method during operational phases [6].



(a)



(b)



(c)



(d)

Figure 1. Image of farmland with different levels of lodging (a) Lodging Farmland Scene 1 in Jiading District; (b) Lodging Farmland Scene 2 in Jiading District; (c) Lodging Farmland Scene 1 in Chongming District; (d) Lodging Farmland Scene 2 in Chongming District.

This paper will delve into the precision segmentation of both the harvested and unharvested regions within farmland images influenced by lodging. The primary objective is to extract the harvester's route while concurrently achieving real-time height detection of unharvested crops positioned in front of the harvester. This paper proposes a harvesting operation image segmentation method based on SLIC superpixel segmentation and the AdaBoost ensemble learning algorithm. This segmentation enables two essential tasks. Firstly, the RANSAC algorithm is employed to extract the harvester's operational route through straight-line fitting from the segmented image. Secondly, the method utilizes a 3D point cloud generated by binocular vision, combined with IMU information for attitude correction, to estimate the height of the harvested crop in front of the harvester.

Initially, we employed the Simple Linear Iterative Clustering (SLIC) [7] image segmentation algorithm to preprocess the collected images, generating a substantial number of superpixel samples for machine learning. After performing multi-dimensional feature extraction and manual labeling of these superpixels, we utilized the AdaBoost ensemble learning algorithm [8] for model training. During the harvester's operation, we execute superpixel segmentation, feature extraction, and predictive classification on real-time collected images. Subsequently, we extract the harvester's operating route through straight-line fitting using the random sampling consensus (RANSAC) algorithm [9]. Simultaneously, we transform the disparity map into a 3D point cloud and apply attitude correction based on the IMU information. By leveraging the range of the unharvested area determined through machine learning, we extract the crop height within a specific range of this area. The algorithm's flow is visually represented in Figure 2.

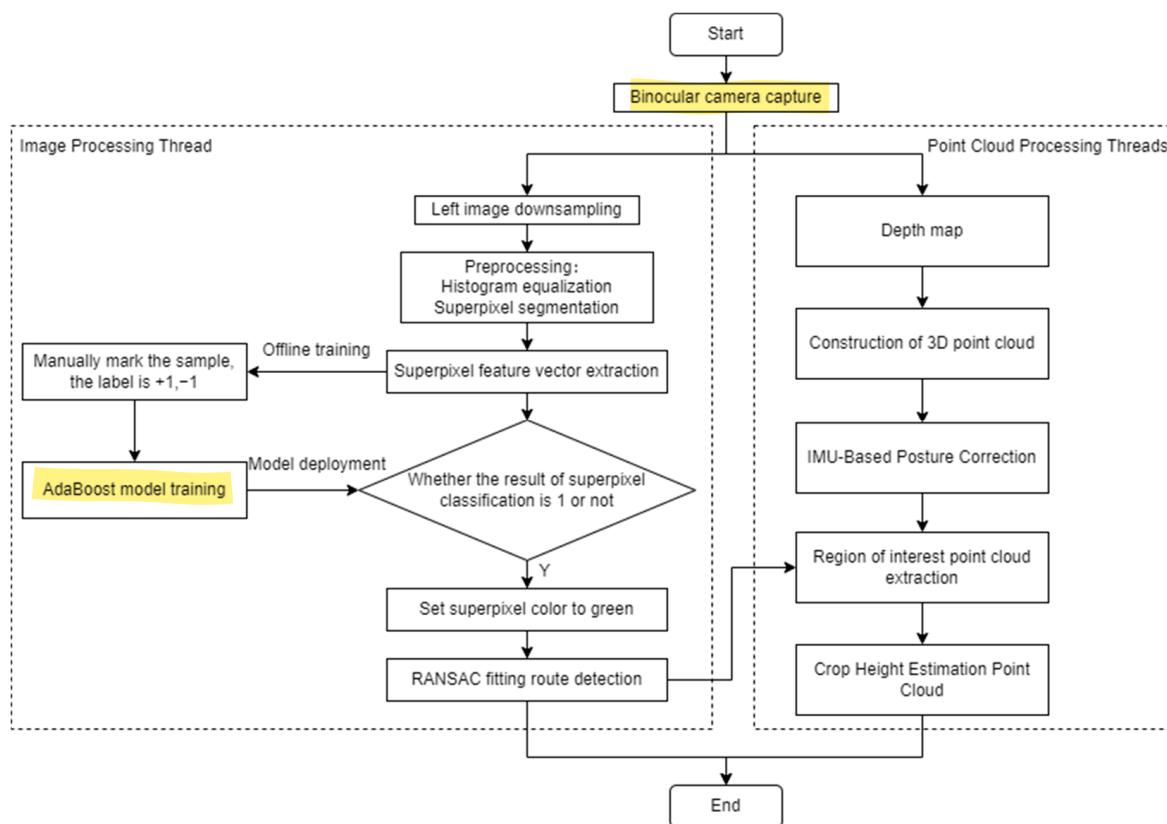


Figure 2. Algorithm flowchart.

This paper is structured as follows: Section 2 presents related work. Section 3 introduces the materials and methods. Section 3.1 introduces the experimental platform and equipment installation. Sections 3.2–3.4 introduce sample acquisition based on superpixel segmentation, feature selection, and identification of crops to be harvested based on the AdaBoost algorithm. Section 3.5 presents a method for route detection of harvesters based on the RANSAC algorithm. Section 3.6 presents a method for crop height estimation based on 3D point clouds. Section 4 presents the results. Sections 4.1 and 4.2 present validation of the quality of the AdaBoost model and experiment results for route extraction and crop height estimation. Section 5 presents a discussion. Section 6 presents the conclusions.

2. Related Work

Scholars have used various methods to study the regional perception of crops to be harvested. According to the sensors used, it can be roughly divided into two types: image processing methods based on monocular cameras [10] and 3D information processing methods based on laser sensors and binocular cameras [11]. Detection based on a monocular

camera is the extraction of specific edge features based on the color and texture of two-dimensional images, including the extraction method based on wavelet filter and fuzzy pattern recognition, the method based on inverse perspective mapping, the region growth algorithm based on adaptive threshold, and the image segmentation algorithm based on maximum interclass variance [12]. Methods based on binocular cameras are mostly based on depth maps after binocular stereo matching. Luo et al. [13], for example, proposed a fast and robust detection method for harvesting edges of multiple crops based on stereo vision, realizing simultaneous detection of unharvested crop edges and crop end edges. However, the detection based on the laser sensor is based on the extraction of specific boundary points based on the three-dimensional location information of the scene. For example, Chateau et al. [14] used filters to limit the interference caused by dust on the laser sensor to achieve crop edge detection.

However, the aforementioned detections of crops intended for harvesting have all been conducted in ideal farmland conditions, neglecting a significant phenomenon—lodging. During the rice harvest period, crop lodging is a prevalent occurrence, primarily caused by agricultural factors such as cultivation methods and external forces like wind and rain. The degree of lodging can vary considerably. In the previously mentioned route detection algorithm for the harvester, the traditional image processing method typically operates at a speed between 100 ms and 500 ms due to its low computational requirements, allowing for real-time detection during the harvesting process. However, since this algorithm is primarily based on the color contrast between harvested and unharvested plots, variations in light intensities and directions during farmland operations can significantly influence the image's color. Moreover, when there are lodging crops in the field, the color difference between the lodging area and the harvested area is not substantial. Consequently, the robustness of such algorithms is limited, and no algorithmic improvements or experiments have been conducted to address large-scale crop lodging. The three-dimensional information processing algorithms based on binocular cameras or lidars solely rely on the height difference in the farmland scene for detection. Nevertheless, when a significant area of crops is severely lodged, the height of the crops in this area becomes similar to that of the harvested stubble. Consequently, the lodged unharvested crops will inevitably be mistaken for the harvested area. As a result, such algorithms cannot be applied in lodging-affected farmland scenarios at all.

Additionally, the occurrence of crop lodging has given rise to another essential perception requirement for the autonomous operation of the harvester, which involves the height detection of crops to be harvested in front of the machine. Crop lodging detection plays a crucial role in providing reference information for adjusting header height and harvesting direction [15]. During the operation of lodging farmland, ensuring that the harvester's header changes reasonably with crop height is vital to achieving a low breakage rate, minimizing impurity content, and maintaining an appropriate stubble height during rice harvesting. When harvesting normal crops, the position of the harvester's header is relatively high, but when faced with lodging crops, the header's height should be appropriately reduced to minimize harvesting losses [16]. Currently, the adaptive control of combined harvester header height has emerged as the main trend in its development [17,18]. According to relevant data, the average loss rate of mechanical harvesting caused by crop lodging exceeds 5% [19]. Therefore, it is necessary to detect crop height in lodging areas.

At present, the detection of crop height is generally based on typical three-dimensional sensors such as lidar to construct farmland maps, and according to the detection platform, it is divided into two categories: unmanned aerial vehicle systems and agricultural vehicle systems. UAVs fly smoothly and do not damage crops, so many scholars use UAVs to carry sensors for ground data collection [20–23]. For example, Chu et al. [24] used a small UAS platform equipped with consumer-grade RGB and near-infrared cameras to collect low-altitude view images of cornfields and reconstruct the 3D structural information of plants to calculate the height and growth of crops. Chang et al. [25] proposed a framework for monitoring crop height using UAS data and estimated the crop growth curve of each

sorghum variety. Liu et al. [26] used UAVs to collect visible light and thermal infrared images and established a comprehensive rice lodging identification model based on particle swarm optimization and the SVM algorithm. However, UAV systems are all used to monitor crop growth information. The method is neither real-time nor connected with ground agricultural machinery, so it cannot provide a real-time reference for ground agricultural machinery operations. In their research on in-vehicle systems, Weiss and Biber [27] used a MEMS-based 3D lidar sensor to discriminate between ground and plants. Moreno et al. [28] installed a laser scanner on the front of the cab of the sprayer, creating a rolling 3D terrain map as a representation of the environment. Masuda et al. [29] proposed a method to detect the lodging degree of rice plants using a laser rangefinder and a camera. It can be concluded that the method of the vehicle-mounted system is basically based on a lidar or binocular camera to realize the height detection of crops, which can provide crop object information for real-time spraying, harvesting, and other operations, and then realize the dynamic adjustment of equipment such as headers.

Based on the above analysis, we choose to seek a better engineering application method for the perception problem of harvesting operations based on binocular vision sensors. On the one hand, with the improvement in the computing power of edge computing devices, machine learning algorithms in the field of vision have spawned a large number of real-time application scenarios [30]. According to the machine learning algorithm, the lodging area and the harvested area of the farmland can be accurately distinguished based on the comprehensiveness of multi-dimensional color and texture features. Moreover, due to the existence of some features that are not sensitive to light brightness, compared with traditional image processing methods, the performance of such algorithms in the actual farmland operation environment is more stable. On the other hand, stereo matching can obtain a 3D point cloud of the surrounding environment from binocular images, which also provides rich information for crop height detection. In this way, with a single binocular camera sensor, we successfully achieve the perception of both problems.

3. Materials and Methods

3.1. Experimental Platform and Equipment Installation

Due to the large amount of real-time environment perception data based on vision and the high computing speed of the algorithm, NVIDIA's Jetson TX2 series modules were selected as the main controller of the system. The NVIDIA Jetson TX2 is a powerful edge computing processor that is very suitable for artificial intelligence. It is equipped with a dual-core 64-bit CPU and an NVIDIA Pascal GPU with 256 NVIDIA CUDA cores. This module fully supports all functions that a discrete NVIDIA GPU can achieve. The CUDA framework and many application programming interfaces based on it, such as the ZED 2 camera used in this paper, is applied to the function interface in the Jetpack 4.5 software development kit, which is suitable for fast computing, image processing, and multi-threaded application scenarios in the field of robotics.

The experimental platform is a Kubota 4LZ-4J (PRO988Q-Q) full-feed crawler harvester, and the experimental sites are harvested farmland in Jiading District and Chongming District, Shanghai. The ZED 2 binocular camera produced by Stereolabs is used as the visual sensor; the side-by-side output resolution is 3080×1080 , and the frame rate is 30 frames per second. The vehicle-mounted Jetson TX2 product of NVIDIA is also used as the algorithm processor, and the camera is connected to the processor through USB for image transmission. In addition to using some functions in the computer vision open source library OpenCV for image processing, the ZED 2 camera software development kit applied to Jetpack 4.5 is used for image acquisition, depth map output, and camera pose detection.

The depth module of the ZED 2 camera has a measurement accuracy of 1% to 3 m and 5% to 15 m. In order to ensure the reliability of stereo camera ranging, we discussed the range and resolution of stereo cameras more accurately according to the binocular ranging formula and realized the derivation of the installation position of the vision sensor.

For the perception task based on stereo vision, we can generally estimate the depth range of the target object in this task and deduce the corresponding parallax value of the stereo camera at this depth from Equation (1).

$$\begin{cases} d_1(Z_{\text{measured}}) = \left[\frac{b \times f_x}{Z_{\text{measured}}} \right] \\ d_2(Z_{\text{measured}}) = d_1(Z_{\text{measured}}) + 1 \end{cases} \quad (1)$$

where Z_{measured} represents the depth distance expected to be measured, in mm, $d_1(Z_{\text{measured}})$ and $d_2(Z_{\text{measured}})$ represent the parallax that may correspond to the depth measurement. Since it is an integral pixel resolution, it is necessary to round the calculated decimal, so there are two parallax differences of 1 pixel, so the parallax resolution at this time is 0.5 pixels.

When the measurement distance is Z_{measured} , the ranging resolution is approximately calculated by Equation (2).

$$\begin{aligned} \varepsilon(Z_{\text{measured}}) &= \left(\frac{b \times f_x}{d_1(Z_{\text{measured}})} - \frac{b \times f_x}{d_2(Z_{\text{measured}})} \right) / 2 \\ &= \frac{b \times f_x}{d_1(Z_{\text{measured}}) \times d_2(Z_{\text{measured}}) \times 2} \end{aligned} \quad (2)$$

where $\varepsilon(Z_{\text{measured}})$ represents the ranging resolution of the stereo camera when the measurement depth is Z_{measured} , in mm.

For the identification and height detection experiments of unharvested crops, in order to ensure the accuracy of 5 cm crop height measurement, it is calculated that $Z_{\text{measured}} \leq 3.50$ m. At the same time, the viewing angle of the camera installation position needs to avoid the occlusion of the harvester header. The experimental platform and equipment installation are shown in Figure 3. The installation height of the camera relative to the ground is 2.4 m, protruding 1 m in front of the cab of the harvester; the vertical viewing angle parameter of the camera is a maximum of 70°; and the installation angle of the camera is calculated to be 18° tilted relative to the ground and viewed from above.



Figure 3. The experiment platform and equipment installation.

3.2. Sample Acquisition Based on Superpixel Segmentation

The size and quality of datasets significantly impact the performance of machine learning algorithms. However, in the agricultural field, research and applications are relatively limited, leading to a scarcity of large-scale datasets, particularly in the context of the harvesting farmland scenarios explored in this chapter. To address this data limitation, a large number of samples are initially acquired through superpixel segmentation and

manual annotation, effectively mitigating the dataset scarcity. Moreover, employing such image processing granularity reduces dimensionality and enhances subsequent image processing efficiency without compromising accuracy.

For the superpixel segmentation of farmland images, the SLIC algorithm is chosen due to its superior overall performance in terms of algorithm speed and superpixel shape. Before segmentation, histogram equalization is applied to the original farmland images to minimize the adverse effects of lighting. The results of the superpixel segmentation of farmland images are depicted in Figure 4. The image collected by the left camera is 1920×1080 pixels. Considering the running time of the machine learning algorithm, the scale after downsampling (960×540 pixels) is selected as the appropriate scale for the classification of farmland harvesting areas and cropped.

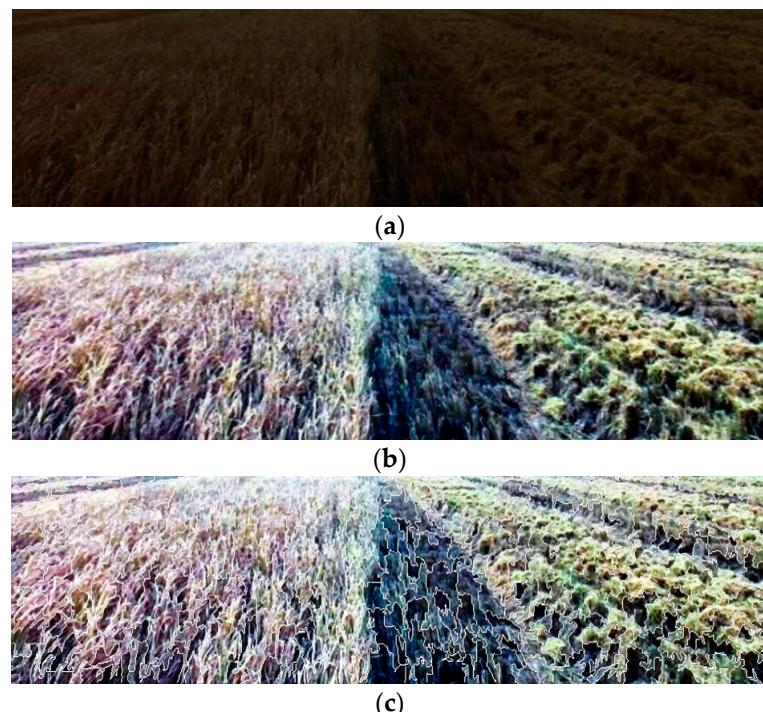


Figure 4. (a) Image of harvested farmland in Jiading District (backlight); (b) Image after preprocessing by histogram equalization; (c) The results of the superpixel segmentation.

The number of superpixels obtained in Figure 4c amounts to 259. It can be seen that the size of each superpixel is relatively moderate, the pixel characteristics inside the superpixels are similar, the obvious edges are all separated, and the boundary between the harvested area and the unharvested area is also more accurate.

3.3. Feature Selection

Commonly used image features mainly include color features and texture features. In this paper, the 6-dimensional color feature vector and the 10-dimensional texture feature vector of superpixels are extracted, which constitute a total of 16-dimensional feature vectors to characterize the characteristics of superpixels. The specific feature selection method is as follows.

3.3.1. Color Feature Extraction

Choose to calculate the first-order color moment (mean) and second-order color moment (variance) of the HSV three-channel to represent the color characteristics of the image. The HSV color space is more visually intuitive. Compared with the highly correlated RGB

color space, the HSV color space is more convenient for color processing and identification. For each channel, the color moment is calculated by Equation (3):

$$\begin{cases} \mu_i = \frac{1}{N} \sum_{j=1}^N p_{i,j} \\ \sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{i,j} - \mu_i)^2 \right)^{\frac{1}{2}} \end{cases} \quad (3)$$

where μ_i represents the first-order color moment of superpixel i , σ_i represents the second-order color moment of the superpixel i , N represents the number of pixels in the superpixel i and $p_{i,j}$ represents the channel value of the j th pixel in the superpixel i .

The visualization effect of the color moment feature is shown in Figure 5.

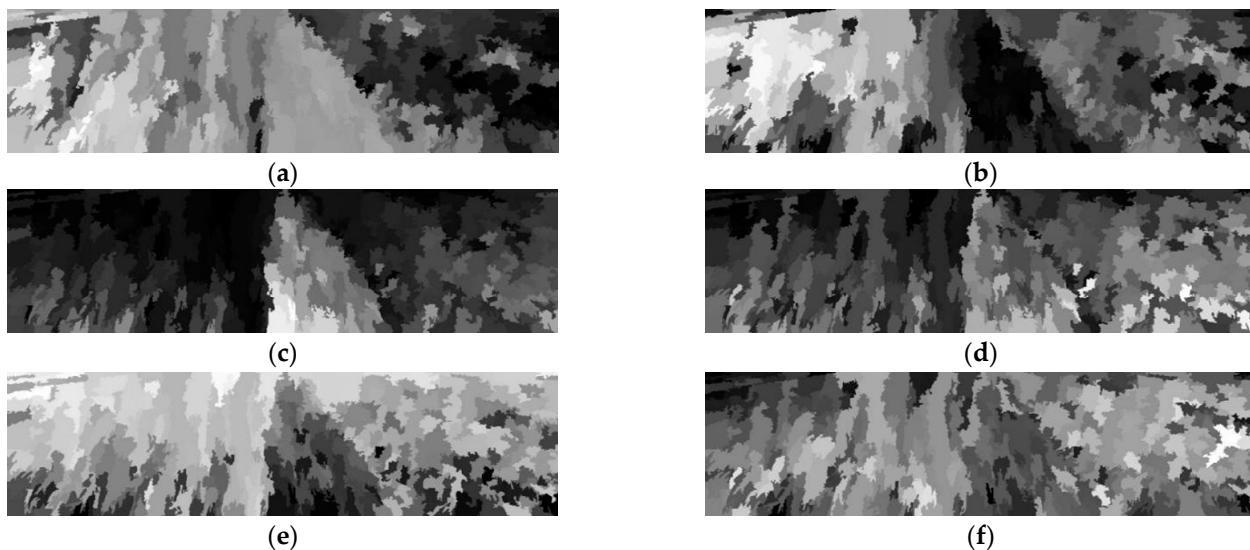


Figure 5. (a) H-channel first-order color moments; (b) H-channel second-order color moments; (c) S-channel first-order color moments; (d) S-channel second-order color moments; (e) V-channel first-order color moments; (f) V-channel second-order color moments.

3.3.2. Texture Feature Extraction

Considering the texture characteristics of the harvested farmland images under investigation, it is hypothesized that intensity features and orientation features of textures would be effective for the image detection process. As such, the mean gradient magnitude and Histogram of Oriented Gradient (HOG) features of pixels within each superpixel are computed as texture features [31].

The Sobel operator is employed for plane convolution with the image, serving the dual purpose of noise reduction and gradient calculation. The gradient calculation formula of the Sobel operator in both the horizontal and vertical directions is presented in Equation (4):

$$\begin{cases} G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \times I \\ G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \times I \end{cases} \quad (4)$$

where G_x represents the horizontal gradient of the image, G_y represents the vertical gradient of the image, and I represents the grayscale image of the image.

Then, for each pixel $p(x, y)$ in the image, the calculation formula of its gradient magnitude and gradient direction is shown in Equation (5):

$$\begin{cases} G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ \alpha(x, y) = \tan^{-1}\left(\frac{G_x(x, y)}{G_y(x, y)}\right) \end{cases} \quad (5)$$

where $G(x, y)$ represents the gradient magnitude of $p(x, y)$ and $\alpha(x, y)$ represents the gradient direction of $p(x, y)$.

For each superpixel, its gradient mean value can be directly calculated, as depicted in Figure 6. For its HOG feature, the directional gradient histogram of 9 bins is used for statistics; that is, 360° is divided into fan-shaped areas in 9 directions, and the gradient magnitude of each pixel in the superpixel is used as the weight coefficient, and each pixel is projected to the fan-shaped area to which its gradient direction belongs. Finally, a 1-dimensional gradient intensity feature vector and a 9-dimensional gradient direction feature vector corresponding to the pixel are obtained.

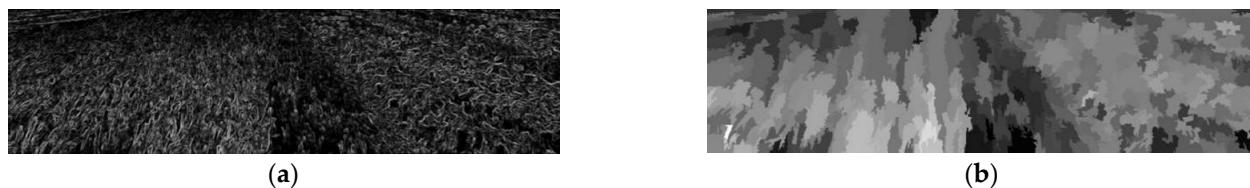


Figure 6. (a) Sobel gradient map; (b) Gradient mean feature map.

3.3.3. Consolidation of Eigenvectors

According to the above calculation process, the 6-dimensional color feature vector and the 10-dimensional texture feature vector of the superpixel are finally extracted; that is, a total of 16-dimensional feature vectors are formed to represent the characteristics of the superpixel, and subsequent machine learning training and prediction are performed.

3.4. Identification of Unharvested Crops Based on the AdaBoost Model

3.4.1. AdaBoost Model Training Sample Processing

AdaBoost is a representative boosting algorithm among integrated learning algorithms. Its principle is to form a strong classifier by properly integrating multiple weak classifiers. The reason for choosing this algorithm is that under the framework of Adaboost, various regression classification models can be used to build a weak learner that is very flexible and has high precision, and the training error decreases at an exponential rate.

The farmland rice harvest images collected in the Chongming and Jiading districts of Shanghai were screened and preprocessed by superpixel segmentation, and superpixel training samples were obtained from 24 images. Manually label the training samples: set the harvested superpixels as negative samples, marked as -1 , and the unharvested superpixels as positive samples, marked as $+1$. Finally, a training sample set $(x_i, y_i), i = 1, 2 \dots 6383$ containing 2862 positive samples and 3521 negative samples is obtained, where x_i is the 16-dimensional feature vector extracted from the i th superpixel sample, and y_i is the pre-labeled label of this sample, $y_i = \{-1, +1\}$.

3.4.2. AdaBoost Model Training for Unharvested Crop Recognition

AdaBoost forms a strong classifier by appropriately integrating multiple weak classifiers, which are widely used in the image field.

First, a weak classifier, $G_1(x)$, is trained with the initial weights, and the weights of the training samples are updated according to the performance of the learning error rate of $G_1(x)$. By reducing the weight of the correctly classified samples and increasing the weight of the wrongly classified samples, the next weak classifier, $G_2(x)$, is guided to make a correct judgment on the wrongly classified samples. A weak classifier $G_2(x)$ is then

trained based on the weighted training set... This is repeated until the number of weak learners reaches a pre-specified number K , and finally, these weak classifiers are integrated according to Equation (6) to obtain the final strong classifier $G(x)$.

$$G(X) = \text{sign} \left(\sum_{k=1}^K \alpha_k G_k(x) \right) \quad (6)$$

where α_k represents the weight coefficient of the k th weak classifier, which is related to the classification error rate of the weak classifier $G_k(x)$ on the training samples.

Use the labeled training sample set to train the AdaBoost model, set the maximum number of iterations of the weak classifier of the AdaBoost model to 15, the weak classifier adopts the decision tree model and set the maximum possible depth of the decision tree to 15.

In addition, for model selection, cross-validation is a means to avoid overfitting and a method to solve overfitting. We can choose the model with the best generalization performance among multiple candidate models [32–34]. Due to the small sample size obtained, the 10-fold cross-validation method was used to increase the generalization ability of the model. The training sample set is randomly divided into 10 parts: 9 samples are used as the training sample set, and 1 sample is used as the test sample set. The AdaBoost model training based on different training and test sets was repeated 10 times, and the model with the highest accuracy was selected as the final model to participate in the subsequent prediction process.

3.4.3. Verification of AdaBoost Model

The prediction effect of the AdaBoost model is shown in Figure 7, where the green part is the superpixel predicted by the model as the unharvested area. It can be seen that the classification results of this model for farmland harvest images are basically accurate. Although there are a few superpixels that have not been correctly classified, they will not have an excessive impact on the route extraction process.



Figure 7. AdaBoost classification results diagram.

3.5. Route Detection of Harvester Based on RANSAC Algorithm

The RANSAC algorithm is used to perform straight-line fitting on the classified images to detect the route of the harvester. The RANSAC algorithm performs well for straight-line fitting of samples containing noisy points.

For the harvested farmland images classified in the previous section, the route detection process based on the RANSAC method is as follows:

1. Perform Otsu binarization, small connected domain removal, and Canny edge detection operations on the farmland classification results;
2. Perform RANSAC straight line fitting on the obtained edge curve to obtain the route of the harvester.

The route detection process based on the RANSAC method is shown in Figure 8.

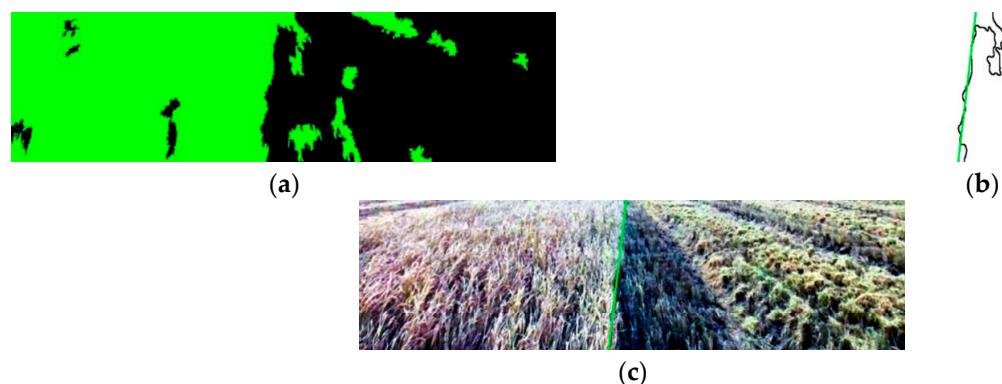


Figure 8. (a) Superpixel binary classification results diagram; (b) RANSAC line fitting graph; (c) Route extraction graph.

3.6. Crop Height Estimation Based on 3D Point Cloud

An IMU is generally composed of three single-axis accelerometers, gyroscopes, and magnetometers, which can respectively obtain the three-axis acceleration, three-axis attitude angle (or angular rate), and magnetic field information of the object. The magnetic field information can be corrected at the three-axis angle. The essence of attitude correction based on the IMU is to transform the space coordinate system.

Due to the muddy land in the rice harvesting farmland, the harvester will inevitably bump during the harvesting process. The ground directly under the head of the harvester is regarded as the origin of the ground coordinate system. At this time, there will be a dynamic angular deviation between the camera coordinate system and the ground coordinate system. We believe that crops are stationary relative to the ground coordinate system, and the 3D point cloud constructed using the camera depth map is based on the camera coordinate system. Therefore, to accurately describe the actual height of the crop with 3D point cloud data, use the inertial sensor (Inertial Measurement Unit, IMU) module that comes with the camera, which obtains the attitude of the camera in real time and converts the camera coordinate system to the ground coordinate system.

Using the stereo camera as a visual sensor, while detecting the unharvested area of the farmland, the point cloud of the area of interest is constructed according to the depth map obtained in real time, and the height of the crops ahead can be estimated.

3.6.1. Construction of Farmland 3D Point Cloud Based on IMU Correction

According to the depth map obtained by stereo matching, the three-dimensional point cloud of the image is constructed, and the calculation formula is shown in Equation (7), as follows:

$$\begin{cases} X_C = \frac{(u - c_x)}{f_x} \times Z_C \\ Y_C = \frac{(v - c_y)}{f_y} \times Z_C \\ Z_C = \frac{f_x \times b_l}{d} \end{cases} \quad (7)$$

where c_x, c_y, f_x, f_y and b_l are both obtained by camera calibration, this paper adopts the calibration method in the literature [35].

The ground directly below the front of the harvester is regarded as the origin of the ground coordinate system, and there will be a dynamic angular deviation between the camera coordinate system and the ground coordinate system during harvesting. To accurately describe the actual height of crops with 3D point cloud data, the Inertial Measurement Unit (IMU) that comes with the ZED 2 camera is used to obtain the attitude of the camera in real time to convert the camera coordinate system to the ground coordinate system. The schematic diagrams of the two coordinate systems are shown in Figure 9. The $O_C - X_C Y_C Z_C$ coordinate system represents the camera coordinate system, and the $O_G - X_G Y_G Z_G$ coordinate system represents the ground coordinate system. The relation-

ship between the camera coordinate system and the ground coordinate system is shown in Equation (8).

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_G \\ Y_G \\ Z_G \\ 1 \end{bmatrix} \quad (8)$$

where R represents the rotation matrix, which is calculated by the built-in IMU module of the camera, and t represents the translation matrix, which is obtained by manual measurement.

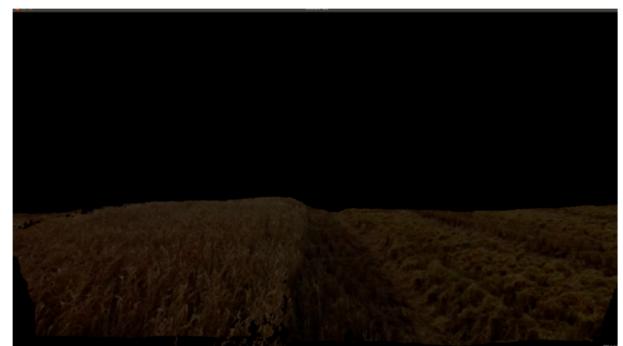


Figure 9. Coordinate system diagram.

Figure 10 shows the 3D point cloud of the farmland image obtained after IMU attitude adjustment.



(a)



(b)

Figure 10. (a) The original image of the harvesting farmland; (b) 3D point cloud of the harvesting farmland.

3.6.2. Point Cloud Extraction of Interest Region

According to the actual application scenario, the following constraints are put forward for the region of interest:

1. The detection of crop height only needs to pay attention to the crop area that is about to be harvested in the farmland, so the point cloud extraction is restricted to the

- unharvested area identified above, and the right area of the route in the image is directly used for the convenience of processing;
2. The experimental platform is a Kubota 4LZ-4J (PRO988Q-Q) full-feed crawler harvester. According to the working width of the harvester, the width of the point cloud extraction is limited to the area to be harvested is about 2 m;
 3. Since the ranging accuracy of the stereo camera is directly related to the measurement distance, it is meaningless for the stereo ranging of too-far targets. In order to ensure the accuracy of crop height measurement, and according to the operating speed of the harvester of about 1.2 m/s, it is determined to measure the 1.5 m-long area in front of the harvester in each frame of the image;
 4. In the ground coordinate system, the crop must be higher than the ground. Generally, the height of rice grown in the mature stage is about 100–110 cm, and there may be lodging. Therefore, point clouds with a height of 0–150 cm are considered as crops, so as to quickly filter some noise interference.

The constraints for specific regions of interest are summarized in Equation (9).

$$\begin{cases} p \in I_{pos} \\ x_G > \max(x_{G_pos}) - 2000 \\ 2000 < z_G < 3500 \\ 0 < y_G < 1500 \end{cases} \quad (9)$$

where I_{pos} represents the set of all the pixels in the image that the model predicts to be 1, that is, the image of the unharvested area, (x_G, y_G, z_G) represents the three-dimensional coordinates of the pixel in the ground coordinate system after coordinate transformation, in mm, x_{G_pos} represents the coordinate value in the x-direction of the pixel predicted by the model to be 1 after coordinate transformation, in mm.

The point cloud extraction effect of the region of interest is shown in Figure 11a. After the point cloud extraction of the area of interest is completed, a square grid with a side length of 10 cm is established for it on the ground coordinate system, and the average height of the point cloud in each grid is calculated to realize the estimation of the crop height in front of the harvester operation. The height estimation point cloud is shown in Figure 11b.

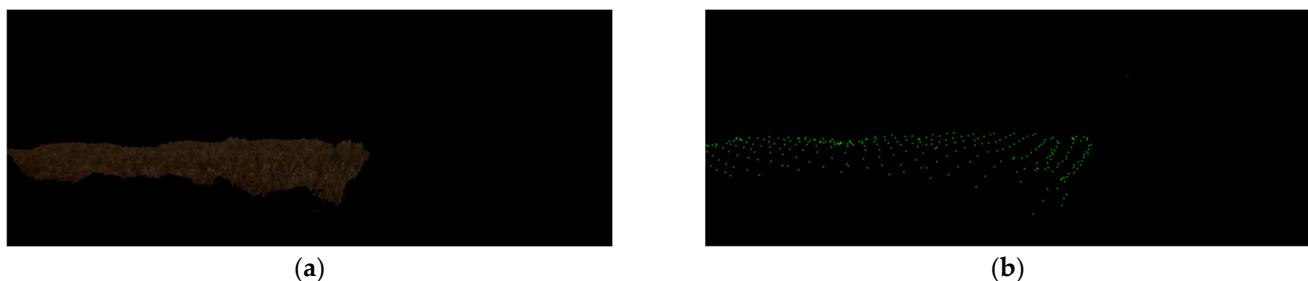


Figure 11. (a) The original point cloud image of the region of interest; (b) The height estimation point cloud.

4. Results

4.1. Validation of the Quality of the AdaBoost Model

To verify the quality of the AdaBoost model, the obtained model is evaluated by the F1 score (F1-Score), which is an indicator used in statistics to measure the accuracy of the binary classification model. It can be seen as a weighted average of model precision and recall. The higher the F1 score, the better the prediction performance of the model. The formula for calculating the F1 score is shown in Equation (10), as follows:

$$F1 = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

where precision represents model accuracy, the ratio of positive samples predicted to be positive to all samples predicted to be positive. Recall represents model recall, the proportion of positive samples predicted to be positive to all samples that are actually positive. Their calculation formulas are shown in Equation (11), as follows:

$$\begin{cases} \text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \end{cases} \quad (11)$$

where TP represents true positives, positive samples whose predictions are correct. TN represents true negatives, negative samples whose predictions are correct. FP represents false positives—positive samples whose predictions are wrong. FN represents false negatives—negative samples whose predictions are wrong.

Generally speaking, precision and recall are negatively correlated. The F1-Score is introduced here as a comprehensive indicator to balance the impact of the accuracy rate and the recall rate and evaluate a classifier more comprehensively. F1 is the harmonic mean of precision and recall. The larger the F1 score, the higher the quality of the model.

In addition to the photos collected from farmland in Chongming and Jiading districts, photos taken on farmland in Yanzhou, Shandong, were added to test the robustness of the model. A total of 50 farmland images during the harvest period were selected, and a total of 14,410 superpixels were obtained. After manual annotation, there are 6224 positive samples and 8186 negative samples, and the AdaBoost model classification results of all superpixels are counted. The data are shown in Table 1.

Table 1. Statistics and evaluation of model classification results.

TP	TN	FP	FN	Precision (%)	Recall (%)	F1 Score (%)
5701	7387	799	523	87.7	91.6	89.6

It can be seen that the effect of this model is good. Although some scattered superpixels are misclassified, experiments show that this model can already meet the needs of subsequent steps and can achieve the research goals of unharvested crop identification and height estimation.

4.2. Experiment Results of Route Extraction and Crop Height Estimation

A Kubota harvester equipped with a stereo system was used to carry out field experiments. The harvester was harvesting at a speed of about 1.2 m/s. The ZED 2 camera captured images in real time and processed them by the Jetson TX2. The farmland images in the two districts were tested, and the original farmland images, superpixel classification results, route extraction effects, and crop height estimation point clouds are shown in Figure 12. The average angle error of the algorithm-detected route and the manually marked route is used as the evaluation index for the identification of unharvested crops, and the average error of the height of the crop at at least four sampling points measured randomly in the region of interest and the height of the corresponding position in the point cloud is used as the evaluation index of the crop height measurement. The experimental results show that the average angle error of the route is about 1.97° and the average error of the height is about 0.054 m. It is believed that this method can provide a reliable reference for the harvesting operation of the harvester.

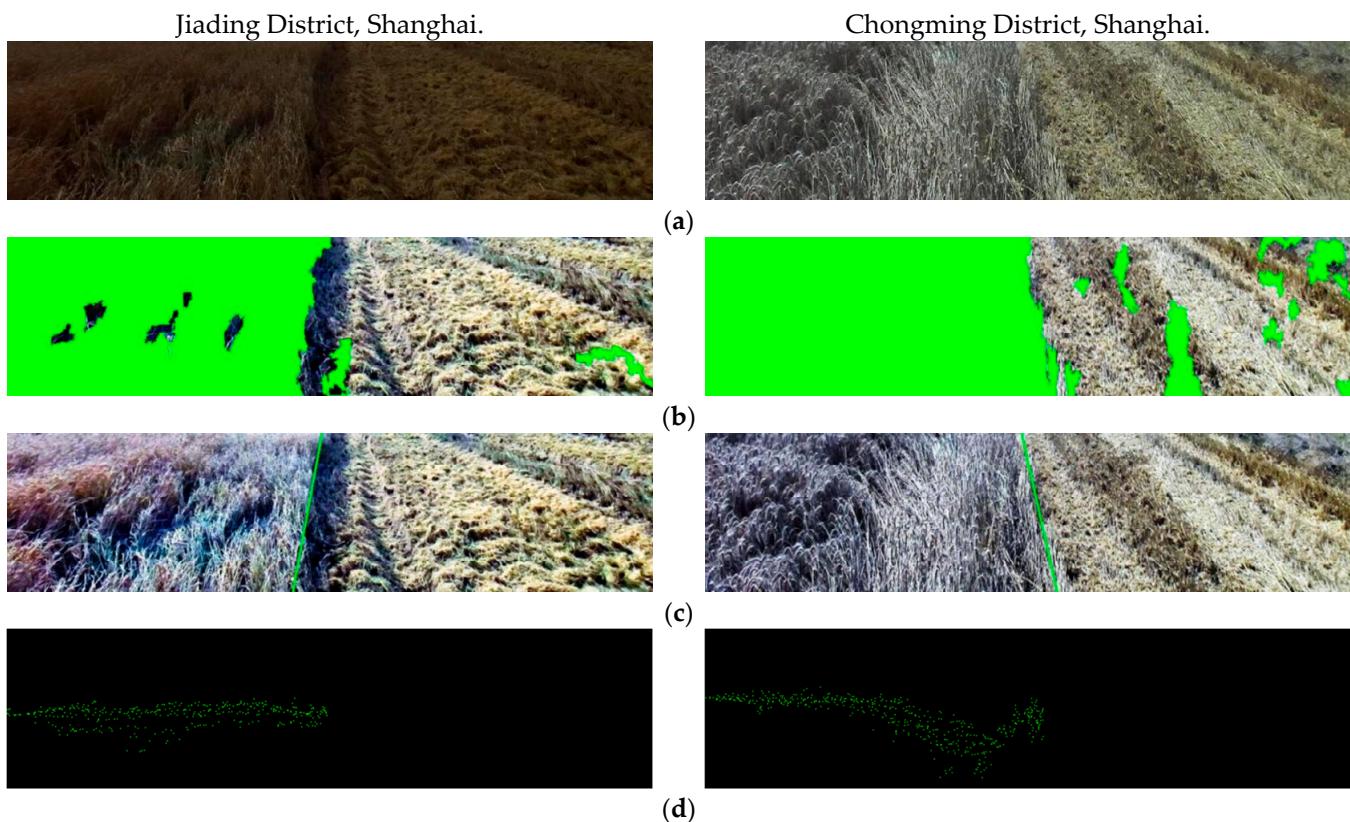


Figure 12. (a) The original image of the harvesting farmland; (b) Superpixel binary classification results diagram; (c) Route extraction graph; (d) Crop height point cloud map.

In addition, the average computation time of each part of the algorithm is shown in Table 2, where the larger part of the crop height estimation step is processed in parallel with the other two steps, so the calculation of one frame of a picture can be completed in about 437 ms, which meets the real-time requirements.

Table 2. The running schedule of each step of the algorithm.

Processing Steps	Time Required (ms)
Image Segmentation Based on AdaBoost	209
Route Detection Based on RANSAC Algorithm	104
Crop Height Estimation Based on 3D Point Cloud	256
Total Time Required	437

5. Discussion

In this paper, we address the perception problem of harvesters operating in fields with lodging, which involves two primary sub-problems: route detection during the harvesting process and crop height estimation in the harvest area.

To achieve more accurate differentiation between the harvested and unharvested areas in farmland with fallen crops, we proposed a classification method based on the AdaBoost model. The SLIC superpixel segmentation preprocessing satisfied the machine learning sample requirements and reduced computational costs. The 16-dimensional feature vector, comprising a 6-dimensional color feature and a 10-dimensional texture feature, was thoughtfully designed to fully utilize image information. This design enhanced the model's robustness compared to single-feature-based training.

For real-time crop height estimation in the harvest area, we converted the depth map obtained from the binocular camera into a 3D point cloud and use IMU data for position and pose correction, ensuring the accuracy of point cloud data during dynamic

operations. Additionally, we constrained the range of unharvested regions based on practical application data to enhance calculation speed and facilitate subsequent processing. The mean height of the regions of interest was then calculated using a grid-based approach.

In the experimental phase, we tested the algorithm in Jiading district and Chongming area farmlands. The results demonstrate that training the AdaBoost classification model accurately segments the farmland in both regions, distinguishing between harvested and unharvested areas. The average error for detecting harvest routes is approximately 1.97° , and the average error for detecting crop height is about 0.054 m, effectively meeting the perception requirements for harvester operations. The total algorithm running time is approximately 437 ms, which fulfills real-time requirements at a harvester speed of about 1.2 m/s.

Future research efforts will focus on integrating existing perception methods with agricultural machinery control processes to enhance the unmanned operation capabilities of rice and wheat harvesters. Additionally, we plan to expand the dataset to optimize the detection model, collecting more fallen farmland data to further improve detection accuracy.

6. Conclusions

This paper addresses the challenges posed by reduced reliability in existing harvester route detection methods and a lack of perception regarding the degree of crop lodging in the field scene. The study focuses on route detection and crop height estimation methods for lodging-affected farmland harvesters during harvesting operations.

The novelty of this approach lies in utilizing image segmentation based on AdaBoost ensemble learning, which proves effective for accurately distinguishing lodging areas and unharvested areas where edge features may not be distinct. Subsequently, two-dimensional images are employed for route detection, while three-dimensional point clouds are used for crop height estimation. The method fully leverages sensor information, enabling the simultaneous accomplishment of two perception tasks based on the same image segmentation results.

Experimental data verifies the effectiveness of the proposed method, with the following key findings:

1. The image segmentation method based on the AdaBoost model enables precise division of the harvested area and the unharvested area in lodging-affected farmland.
2. Utilizing the RANSAC straight line fitting algorithm to obtain the route yields superior results, with an average angle error of approximately 1.97° .
3. By using the IMU for pose correction and calculating the average height of the area of interest through gridding, the detected crop height in the unharvested area exhibits an average error of about 0.054 m, meeting the required accuracy levels.
4. The total running time of the algorithm is approximately 437 ms, effectively meeting the perception requirements during the harvester's operation at a speed of about 1.2 m/s.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, L.G.; investigation, Y.L., Y.G. and L.G.; resources, Y.L.; visualization, Y.G. and L.G.; validation, Y.G. and L.G., writing—original draft preparation, Y.G. and L.G.; writing—review and editing, Y.L., Y.G. and L.G.; supervision, Y.L., L.G. and C.L.; project administration, Y.L. and C.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (grant number 2021YFD2000602) and the National Key Research and Development Program of China (grant number 2022YFD2001305).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Pierce, Q.Z.; Francis, J. *Agricultural Automation: Fundamentals and Practices*; CRC Press: Boca Raton, FL, USA, 2013; ISBN 978-0-429-11217-1.
- Abdullahi, H.S.; Mahieddine, F.; Sheriff, R.E. Technology Impact on Agricultural Productivity: A Review of Precision Agriculture Using Unmanned Aerial Vehicles. In Proceedings of the Wireless and Satellite Systems, Bradford, UK, 6–7 July 2015; Pillai, P., Hu, Y.F., Otung, I., Giambene, G., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 388–400.
- Shah, L.; Yahya, M.; Shah, S.M.A.; Nadeem, M.; Ali, A.; Ali, A.; Wang, J.; Riaz, M.W.; Rehman, S.; Wu, W.; et al. Improving Lodging Resistance: Using Wheat and Rice as Classical Examples. *Int. J. Mol. Sci.* **2019**, *20*, 4211. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xie, Y.; Alleyne, A.G.; Greer, A.; Deneault, D. Fundamental Limits in Combine Harvester Header Height Control. *J. Dyn. Syst. Meas. Control* **2013**, *135*, 0345031–0345038. [\[CrossRef\]](#)
- Hu, P.; Chapman, S.C.; Wang, X.; Potgieter, A.; Duan, T.; Jordan, D.; Guo, Y.; Zheng, B. Estimation of Plant Height Using a High Throughput Phenotyping Platform Based on Unmanned Aerial Vehicle and Self-Calibration: Example for Sorghum Breeding. *Eur. J. Agron.* **2018**, *95*, 24–32. [\[CrossRef\]](#)
- Zhang, Z.; Cao, R.; Peng, C.; Liu, R.; Sun, Y.; Zhang, M.; Li, H. Cut-Edge Detection Method for Rice Harvesting Based on Machine Vision. *Agronomy* **2020**, *10*, 590. [\[CrossRef\]](#)
- Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [\[CrossRef\]](#) [\[PubMed\]](#)
- Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [\[CrossRef\]](#)
- Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Netherlands, 1987; pp. 726–740, ISBN 978-0-08-051581-6.
- Cho, W.; Iida, M.; Suguri, M.; Masuda, R.; Kurita, H. Vision-Based Uncut Crop Edge Detection for Automated Guidance of Head-Feeding Combine. *Eng. Agric. Environ. Food* **2014**, *7*, 97–102. [\[CrossRef\]](#)
- Wen, J.; Yin, Y.; Zhang, Y.; Pan, Z.; Fan, Y. Detection of Wheat Lodging by Binocular Cameras during Harvesting Operation. *Agriculture* **2023**, *13*, 120. [\[CrossRef\]](#)
- Masuda, R.; Fujimoto, S.; Iida, M.; Suguri, M. A Method to Detect the Occurrence of Rice Plant Lodging Using Wavelet Transform. *IFAC Proc. Vol.* **2013**, *46*, 75–80. [\[CrossRef\]](#)
- Luo, Y.; Wei, L.; Xu, L.; Zhang, Q.; Liu, J.; Cai, Q.; Zhang, W. Stereo-Vision-Based Multi-Crop Harvesting Edge Detection for Precise Automatic Steering of Combine Harvester. *Biosyst. Eng.* **2022**, *215*, 115–128. [\[CrossRef\]](#)
- Chateau, T.; Debain, C.; Collange, F.; Trassoudaine, L.; Alizon, J. Automatic Guidance of Agricultural Vehicles Using a Laser Sensor. *Comput. Electron. Agric.* **2000**, *28*, 243–257. [\[CrossRef\]](#)
- Xue, J.; Li, L.; Xie, R.; Wang, K.; Hou, P.; Ming, B.; Zhang, W.; Zhang, G.; Gao, S.; Bai, S.; et al. Effect of lodging on maize grain losing and harvest efficiency in mechanical grain harvest. *Acta Agron. Sin.* **2018**, *44*, 1774–1781. [\[CrossRef\]](#)
- Xie, B.; Wang, J.; Jiang, H.; Zhao, S.; Liu, J.; Jin, Y.; Li, Y. Multi-Feature Detection of in-Field Grain Lodging for Adaptive Low-Loss Control of Combine Harvesters. *Comput. Electron. Agric.* **2023**, *208*, 107772. [\[CrossRef\]](#)
- Wang, H.; Shen, H.; Cao, S.; Xu, X.; Han, T.; Guo, H. Hydraulic System Design of Combined Harvester Header and Simulation of Header Lifting System. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *233*, 032012. [\[CrossRef\]](#)
- Liu, H.; Reibman, A.R.; Ault, A.C.; Krogmeier, J.V. Video-Based Prediction for Header-Height Control of a Combine Harvester. In Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 28–30 March 2019; pp. 310–315.
- Wang, K.; Xie, R.; Ming, B.; Hou, P.; Xue, J.; Li, S. Review of Combine Harvester Losses for Maize and Influencing Factors. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 1–10. [\[CrossRef\]](#)
- Xie, T.; Li, J.; Yang, C.; Jiang, Z.; Chen, Y.; Guo, L.; Zhang, J. Crop Height Estimation Based on UAV Images: Methods, Errors, and Strategies. *Comput. Electron. Agric.* **2021**, *185*, 106155. [\[CrossRef\]](#)
- Zhou, L.; Gu, X.; Cheng, S.; Yang, G.; Shu, M.; Sun, Q. Analysis of Plant Height Changes of Lodged Maize Using UAV-LiDAR Data. *Agriculture* **2020**, *10*, 146. [\[CrossRef\]](#)
- Wang, J.; Ge, H.; Dai, Q.; Ahmed, I.; Dai, Q.; Zhou, G.; Qin, M.; Gu, C. Unsupervised Discrimination between Lodged and Non-Lodged Winter Wheat: A Case Study Using a Low-Cost Unmanned Aerial Vehicle. *Int. J. Remote Sens.* **2018**, *39*, 2079–2088. [\[CrossRef\]](#)
- Cao, W.; Qiao, Z.; Gao, Z.; Lu, S.; Tian, F. Use of Unmanned Aerial Vehicle Imagery and a Hybrid Algorithm Combining a Watershed Algorithm and Adaptive Threshold Segmentation to Extract Wheat Lodging. *Phys. Chem. Earth Parts ABC* **2021**, *123*, 103016. [\[CrossRef\]](#)
- Chu, T.; Starek, M.J.; Brewer, M.J.; Masiane, T.; Murray, S.C. UAS Imaging for Automated Crop Lodging Detection: A Case Study over an Experimental Maize Field. In Proceedings of the SPIE Commercial + Scientific Sensing and Imaging, Anaheim, CA, USA, 9–13 April 2017; Volume 10218. [\[CrossRef\]](#)
- Chang, A.; Jung, J.; Maeda, M.M.; Landivar, J. Crop Height Monitoring with Digital Imagery from Unmanned Aerial System (UAS). *Comput. Electron. Agric.* **2017**, *141*, 232–237. [\[CrossRef\]](#)

26. Liu, T.; Li, R.; Zhong, X.; Jiang, M.; Jin, X.; Zhou, P.; Liu, S.; Sun, C.; Guo, W. Estimates of Rice Lodging Using Indices Derived from UAV Visible and Thermal Infrared Images. *Agric. For. Meteorol.* **2018**, *252*, 144–154. [[CrossRef](#)]
27. Weiss, U.; Biber, P. Plant Detection and Mapping for Agricultural Robots Using a 3D LIDAR Sensor. *Robot. Auton. Syst.* **2011**, *59*, 265–273. [[CrossRef](#)]
28. Moreno, F.; Cielniak, G.; Duckett, T. Evaluation of Laser Range-Finder Mapping for Agricultural Spraying Vehicles. In Proceedings of the Towards Autonomous Robotic Systems, Oxford, UK, 28–30 August 2013; p. 221, ISBN 978-3-662-43644-8.
29. Mausda, R.; Nomura, K.; Iida, M.; Suguri, M. Detection of Rice Plant Lodging Using Camera and Laser Range Finder. In Proceedings of the International Conference of Agricultural Engineering CIGR-AgEng 2012 Agriculture & Engineering for a Healthier Life, Valencia, Spain, 8–12 July 2012.
30. Shinzato, P.Y.; Wolf, D.F. A Road Following Approach Using Artificial Neural Networks Combinations. *J. Intell. Robot. Syst.* **2011**, *62*, 527–546. [[CrossRef](#)]
31. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
32. Khan, J.; Fayaz, M.; Hussain, A.; Khalid, S.; Mashwani, W.K.; Gwak, J. An Improved Alpha Beta Filter Using a Deep Extreme Learning Machine. *IEEE Access* **2021**, *9*, 61548–61564. [[CrossRef](#)]
33. Khan, J.; Kim, K. A Performance Evaluation of the Alpha-Beta (α - β) Filter Algorithm with Different Learning Models: DBN, DELM, and SVM. *Appl. Sci.* **2022**, *12*, 9429. [[CrossRef](#)]
34. Khan, J.; Lee, E.; Kim, K. A Higher Prediction Accuracy-Based Alpha–Beta Filter Algorithm Using the Feedforward Artificial Neural Network. *CAAI Trans. Intell. Technol.* **2022**. [[CrossRef](#)]
35. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.