agriculture

MDPI

Article

Identifying Working Trajectories of the Wheat Harvester In-Field Based on K-Means Algorithm

Lili Yang [1,2], Xinxin Wang [1,2], Yuanbo Li [1,2], Zhongxiang Xie [1,2], Yuanyuan Xu [1,2], Rongxin Han [1,2] and Caicong Wu [2,*]

[1] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[2] Key Laboratory of Agricultural Machinery Monitoring and Big Data Application, Ministry of Agriculture and Rural Affairs, Beijing 100083, China
* Correspondence: wucc@cau.edu.cn

Abstract: Identifying the in-field trajectories of harvests is important for the activity analysis of agricultural machinery. This paper presents a K-means-based trajectory identification method that can automatically detect the "turning", "working", and "abnormal working" trajectories for wheat harvester in-field operation scenarios. This method contains two stages: clustering and correction. The clustering stage performs by the two-step K-means iterative clustering method (D-K-means). In the correction stage, the first step (M1) is performed based on the three distance features between the trajectory segments and the cluster center of the trajectory segments. The second step (M2) is based on the direction change of the "turning" and "abnormal working" trajectories. The third correction step (M3) is based on the operating characteristics to specify the start and stop positions of the turning. The developed method was validated by 50 trajectories. The results for the three trajectories and the five time intervals from 1 s to 5 s both have f1-scores above 0.90, and the f1-score using only the clustering method and the method of this paper increased from 0.55 to 0.95. After removing the turning and abnormal operation trajectories, the error of calculating farmland area with distance algorithm is reduced by 17.04% compared with that before processing.

Keywords: agricultural machinery; trajectory recognition; k-means clustering; machine learning; GNSS data

Citation: Yang, L.; Wang, X.; Li, Y.; Xie, Z.; Xu, Y.; Han, R.; Wu, C. Identifying Working Trajectories of the Wheat Harvester In-Field Based on K-Means Algorithm. *Agriculture* **2022**, *12*, 1837. https://doi.org/10.3390/agriculture12111837

Academic Editors: Paul Kwan and Wensheng Wang

Received: 30 September 2022
Accepted: 28 October 2022
Published: 2 November 2022

1. Introduction

Identifying agricultural machinery activities is an important part of a machinery management systems [1], and GNSS has become an important tool for identifying the activities of agricultural machinery [2,3]. Kong Qinghao et al. [4] loaded SEMTECH-DS gaseous pollutant analyzer on the rotary tiller, integrated accessories such as a tachometer, tachymeter, and GPS and designed and collected experimental data based on the time series of driving speed, engine speed, and real-time fuel consumption to conduct a study of pollutant emissions from the tractor under three operating states of idling, moving, and working so as to analyze the pollutant emission trends under different operating states.

Identifying the operational and non-operational behavior of agricultural machinery helps to calculate the operational efficiency of agricultural machinery, and the research on agricultural machinery trajectory identification based on GNSS positioning data usually includes field–road trajectory identification [5–8] and working–turning trajectory identification [9,10]. Jernej Poteko et al. [7] developed the classification detection of field and road trajectories based on two technologies of GNSS devices, EGNOS and RTK, to obtain the velocity, acceleration, curve radius, and angular velocity attributes of tractors and concluded that the model using RTK data is more effective. Wang Pei et al. [11] automatically identified field trajectories, road trajectories, and stopping points based on GPS positioning data and velocity information and analyzed the time utilization rate of agricultural machinery by calculating the number of in-field trajectory points to the total number of running trajectory points of agricultural machinery. Ying Chen et al. [5,8], based on GNSS positioning data and velocity data, used DBSCAN and GCN algorithms to identify field and road trajectories, respectively, and calculated field area based on the identification results for massive trajectories, which can support the tracking of the annual wheat harvest process.

Refining the identification of in-field non-working trajectories can be used to further refine the analysis of agricultural machinery operational efficiency, field area calculation, fuel consumption, and emissions of agricultural machinery. Turning trajectories are an important part of the non-working trajectories that make up the in-field trajectories. Randal K. Taylor et al. [12] used positioning data, speed, time, and mass flow obtained from GPS and yield monitor to identify the three behaviors of harvester harvesting, turning, and unloading the crop so as to analyze the correlation between turning and operation efficiency and concluded that reducing the turning time in the field could improve the operation efficiency. Yulin Li [13] proposed an optimization scheme for harvesting trajectory based on cotton harvester operation trajectory data to optimize the turning mode of harvesters and improve the in-field operation efficiency of agricultural machinery. Weize Tian [14], based on smartphone positioning data, identified the turning, seeding, and sowing trajectories of corn planters and analyzed the fuel consumption of the turning trajectory on this basis.

Wheat is one of the major food crops worldwide [15]. In the context of the global move toward digitalization in agriculture [16,17], research on wheat harvester trajectories is critical to the development of digital agriculture. Wheat harvest trajectories contain typical X-turn trajectories, and the study of wheat harvest trajectories can be extended to the study of harvest trajectories of other crops. Therefore, this paper uses GNSS equipment to obtain the driving trajectory data of real wheat harvesting agricultural machinery, identifying the wheat harvester's turning trajectory, working trajectory, and abnormal working trajectory in a field. Specifically, the preliminary results of "working " and "turning" trajectories are obtained by using the D-K-means method, that is, two K-means iterative clustering, and then a three-step correction is performed:

(1) Reclassification of trajectories based on trajectory segments using three distance algorithms.
(2) Identifying the "abnormal working" trajectory based on the change of direction before and after the agricultural machine's turning trajectory.
(3) Correction of the "turning" trajectory boundary based on the operating characteristics of the harvester.

Finally, it is verified that the algorithm identifies three trajectories well by various indicators. In addition, in order to verify the significance and application value of the algorithm, the harvesting area is calculated by the traditional distance algorithm based on the algorithm identification results, and the accuracy of the area calculation results after removing the turning and abnormal working trajectories from the trajectory data is proved to be improved. In addition, using the rice harvesting trajectory and corn harvesting trajectory to verify the algorithm, it is proved that the recognition algorithm of this paper can be extended to apply to the trajectory data of other crops and many models of agricultural machines. Other researchers can carry out other research on the basis of this study, such as the efficiency of agricultural machinery operation and optimization of agricultural machinery driving routes, which will help the intelligent management of agricultural machinery and promote the development of agriculture to digitalization and intelligence.

2. Materials and Methods

2.1. Overview

The typical trajectory of a harvester in the wheat harvesting operation is shown in Figure 1. This paper aims to identify three types of trajectories of wheat harvesters in the field, which are "turning", "abnormal working", and "working", and the X-turn is often used in wheat harvester operation [18,19]. X-turn refers to the multiple forward and backward behaviors of "forward–reverse–forward" made by agricultural machinery to change the direction of travel during field operation, as shown in part a of Figure 1.

"Abnormal working" occurs when an agricultural machine's trajectory in the "working" area deviates from the initial direction of operation, engages in behaviors such as "going around," and then returns to the original trajectory direction, as shown in part b of Figure 1. Except for the turning trajectories and abnormal working trajectories, the rest are working trajectories, as shown in part c of Figure 1. In this paper, the turning trajectory is only for "X-turn", which is abbreviated as "turning" hereinafter.

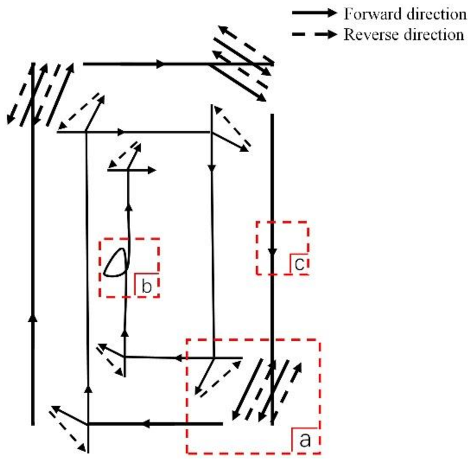

Figure 1. Schematic diagram of the trajectory of the wheat harvester in the field. (**a**) Turning trajectory, (**b**) abnormal working trajectory, and (**c**) working trajectory.

There is at least one inflection point in both the "turning" trajectory and the "abnormal working" trajectory, while there is no inflection point in the "working" trajectory, which is comparatively smooth. These are the data characteristics of each of these three types of trajectory data. The inflection point is the point where the heading angle of one of the consecutive trajectory points is significantly larger, and the heading angle is the angle between two vectors formed by three consecutive trajectory points; detailed definitions are given in Section 2.3. In contrast to the direction of agricultural machinery before and after "abnormal working", the direction of agricultural machinery before and after the "turning" trajectory is different. The angle between the vectors before and after the turning trajectory or the abnormal working trajectory can be used to determine whether the driving direction has changed. If the angle is less than the threshold, the direction has not changed, and vice versa if the angle is greater than the threshold. Refer to Section 2.4.2 for the precise calculating process. For the data characteristics of different trajectories mentioned above, the trajectory recognition algorithm process of the harvester is shown in Figure 2:

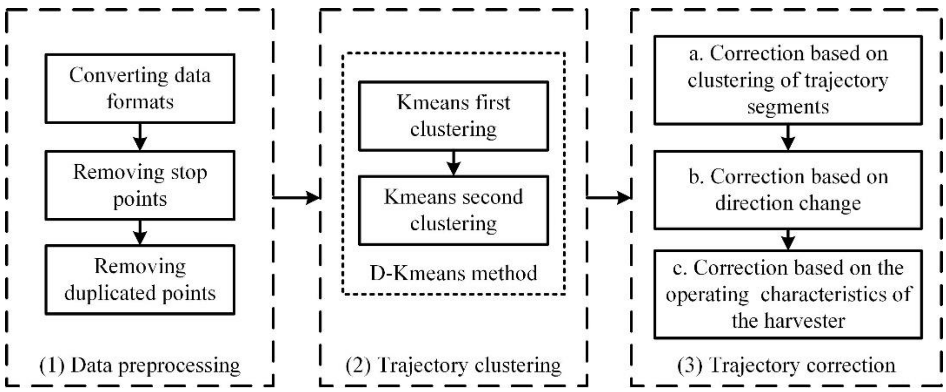

Figure 2. Flowchart of wheat harvester's trajectory identification based on K-means algorithm. a. The first correction can improve the trajectory identifying accuracy of D-K-Means; b. The second correction can identify abnormal working trajectories; c. The third correction can define the begin and end positions of the turning.

(1) Data preprocessing. It includes three operations: converting data formats, removing stop points, and removing duplicated points.
(2) Trajectory clustering. The D-K-means clustering algorithm is used; that is, it is clustered after two K-means iterations. Iterative clustering means using the original data features for the first clustering and later constructing new data features for the second clustering based on the first clustering results. In this paper, this method is named as D-K-means. After iterative clustering, we initially obtain the "working" trajectories and "turning" trajectories, and the results are not good and do not obtain the "abnormal working" trajectories, so further corrections are needed.
(3) Trajectory correction. It consists of three steps to realize three functions, respectively:
    a. Reclassify the current clustering results based on the trajectory fragment using three distance algorithms to correct the situation that the "working" trajectory is mistakenly recognized as the "turning" trajectory;
    b. Distinguish the "abnormal working" trajectory from the "turning" trajectory based on the change of the driving direction of the agricultural machine to correct the "abnormal working" trajectory mistakenly recognized as the "turning" trajectory;
    c. Define the beginning and ending positions of each "turning" trajectory according to the operating characteristics of the harvester.

2.2. Datasets and Data Preprocessing

The wheat harvesting trajectories were collected by harvesters equipped with the Global Navigation Satellite System (GNSS) in the main wheat-producing areas such as Hebei, Henan, and Shandong during the wheat harvest season in June 2022, and the trajectories of agricultural machines in the fields were obtained through a professional website used to label data. The GNSS trajectories were recorded using Model HOMER3E receivers or Model HOMER3 receivers or Model HOMER4S receivers and no correction was applied. Each GNSS record includes tractor ID, timestamp (YYYY-MM-DD hh:mm: ss), geospatial coordinate (including longitude and latitude, WGS84), and speed (m/s). The sampling interval is 1 s to 5 s, and 10 agricultural machine trajectory data points within the field were selected for each time interval, totaling 50 field trajectories and 105,496 data points, with a total area of 51.13 hectares. The included wheat harvesters are self-propelled grain combine harvesters with body lengths of 6.8 m, widths of 2.75 m, and product types of 4LZ-8E2 and 4LZ-7E5.

In addition, each trajectory point in the data was manually labeled as "turning", "working" or "abnormal working", and the labeled results were taken as the real values of the experiment. For the manual labeling of the turning trajectory, referring to the common turning method of wheat harvesting: when the harvester makes an X-turn the machine crosses at least half of the harvested land before turning for quickly clustering [19]. The beginning of the "turning" trajectory is where the first half body distance before the first turning point is located, and the end of the "turning" trajectory is where the final half body distance is located after the last turning point. The trajectory points between 3.4 m before the first inflection point and 3.4 m after the last inflection point are referred to as "turning" trajectory points since the length of the body corresponding to the harvester models concerned is 6.8 m.

The preprocessing operation for the trajectory data consists of three steps:

(1) Data format conversion: The time format is converted to (YYYYMMDDhhmmss); the latitude and longitude of each trajectory point are converted to the coordinates in the geodetic coordinate system using UTM projection (Universal Transverse Mercator Projection), which is recorded as x, y.
(2) Remove the stopping points. A set of stopping points is considered to be any continuous trajectory point with a speed less than 0.5 m/s [9]. Only one trajectory point is kept once the halting point has been removed. The speed is set to zero, and the longitude and latitude of this trajectory point are equal to the average values for this group of stopping locations' longitude and latitude.

(3) Remove duplicate points. When consecutive trajectory points have the same latitude and longitude, the first trajectory point is kept and the other trajectory points are removed.

After preprocessing the trajectory data, there are 99,691 data points left, and the data structure of each trajectory point is represented as $p_i = (\text{time}, \text{x}, \text{y}, \text{speed})$.

2.3. D-K-means-Based Clustering

Because of its straightforward premise, strong stability, and quick convergence, the K-means method is a straightforward and efficient clustering algorithm for processing vast data. It is frequently employed in the field of vehicle driving trajectory recognition study [20–22]. Compared with the DBSCAN clustering algorithm, K-means runs faster and occupies less memory [23].

In order to achieve the initial recognition of the "working" trajectory and the "turning" trajectory, we use the D-K-means method in this paper to cluster the trajectories of the harvester in the field twice. The k-value of both clusters is set to 2, allowing us to obtain the binary classification results of the working and turning trajectory. The input feature of the first clustering is the heading angle of each trajectory point, and the angle is calculated as shown in Formula (1). For the three continuous trajectory points p_{i-1}, p_i, and p_{i+1} in Figure 3, the angle value of the intermediate point p_i is the angle between the vector a and the vector b, where the vector a is formed by the first two trajectory points p_{i-1} and p_i, and vector b is the vectors formed by the last two trajectory points p_i and p_{i+1}. All of the inflection points in the trajectory whose direction has changed can be originally acquired after the first clustering stage.

$$angle = \cos^{-1}\left(\frac{a \cdot b}{|a||b|}\right) \quad (1)$$

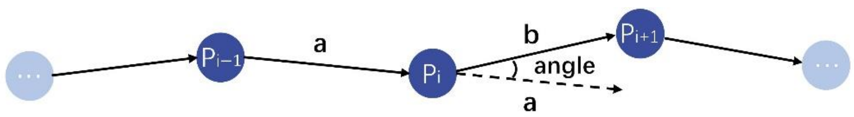

Figure 3. Schematic diagram of angle calculation for trajectory point p_i. a is the vector between p_{i-1} and p_i, b is the vector between p_i and p_{i+1}.

Define dis as the Euclidean distance from each trajectory point to its nearest inflection point, and dis is calculated as in Formula (2):

$$dis = \sqrt{(x_i - \acute{x})^2 + (y_i - \acute{y})^2} \quad (2)$$

where (x_i, y_i) are the coordinates of the geodetic coordinate system for each trajectory point and $(\acute{x}, \acute{y})$ is the geodetic coordinate system coordinates of each inflection point in the first step of clustering results.

The distance feature dis of each trajectory point is used as the input feature in the second step of clustering, which is based on the previous step's clustering results. After clustering, the "working" trajectory (recorded as label0) and "turning" trajectory (recorded as label1) are obtained. The data structure of the trajectory points is $p_i = (\text{time}, \text{x}, \text{y}, \text{speed}, \text{angle}, \text{dis}, \text{label})$.

2.4. Correction Method

2.4.1. M1 Correction Based on Trajectory Segments Cluster

After the D-K-means method, there is a situation that the "working" trajectory is mistakenly identified as the "turning" trajectory, and to correct this situation, the following operation is performed:

(1) Trajectory fragmentation: The trajectory points belonging to the same trajectory category (label0 or label1) and adjacent in the time series are treated as a set of trajectory fragments.
(2) Feature construction of trajectory segments: Calculate the standard deviation, mean, and maximum of the angle at which each group of trajectory segments' trajectory points, and recorded as $X_i = \{std_i, mean_i, max_i\}$ (std_i is the standard deviation of angle of the ith segment, $mean_i$ is the mean of angle of the ith segment, max_i is the maximum of angle of the ith segment), X_i is the feature set of the ith trajectory segment.
(3) Finding the cluster class center: The cluster class center of each category is obtained by calculating the average of the feature set of that category, where the cluster class center of label 0 is denoted as $X_{center0} = \{std_{c0}, mean_{c0}, max_{c0}\}$, and the cluster class center of label 1 is denoted as $X_{center1} = \{std_{c1}, mean_{c1}, max_{c1}\}$.
(4) Trajectory fragment reclassification: Taking the trajectory fragment as the basic unit calculate the distance from each trajectory fragment $X_i(\mathrm{i} = 1, 2, 3\ \ldots\ldots)$ to the cluster class center X_{center0} and X_{center1}, and the distance is calculated by Euclidean distance, Chebyshev distance, and Manhattan distance, which are calculated by Formula (3), Formula (4), and Formula (5). In order to decide which sort of trajectory the group of trajectory belongs to, the voting choice [24] is taken to choose the closest cluster center of each trajectory segment from the calculation results of the three distance methods.

$$ED_{ij} = \sqrt{\left(std_i - std_{cj}\right)^2 + \left(mean_i - mean_{cj}\right)^2 + \left(max_i - max_{cj}\right)^2} \quad (3)$$

$$CD_{ij} = \max\left(\left|std_i - std_{cj}\right|, \left|mean_i - mean_{cj}\right|, \left|max_i - max_{cj}\right|\right) \quad (4)$$

$$MD_{ij} = \left|std_i - std_{cj}\right| + \left|mean_i - mean_{cj}\right| + \left|max_i - max_{cj}\right| \quad (5)$$

where i is the trajectory segment number (i = 0,1,2......), j is the category number (j = 0,1), and $\mathrm{ED_{ij}}$ is the Euclidean distance from the ith trajectory fragment $\mathrm{X_i}$ to the category j center X_{centerj}. $\mathrm{CD_{ij}}$ is the Chebyshev distance from the ith trajectory fragment $\mathrm{X_i}$ to the category j center X_{centerj}, and $\mathrm{MD_{ij}}$ is the Manhattan distance from the ith trajectory fragment $\mathrm{X_i}$ to the category j center X_{centerj}.

2.4.2. M2 Correction Based on Orientation Change

After the D-K-means method and M1 correction, there is still the phenomenon that the "abnormal working" trajectory is misidentified as the "turning" trajectory. The agricultural machine moves differently before and after the "turning" trajectory than it does before and after the "abnormal working" trajectory, which moves in the same direction. Based on the aforementioned data features, the "abnormal working" trajectory and the "turning" trajectory are separated in this section.

The correction method is the "turning" trajectory in the results of the aforementioned D-K-means method and M1 correction method, as shown in Figure 4 triangle trajectory points, calculate the three vectors between the three consecutive points before the "turning" trajectory $a_1,\ a_2, a_3$ and the three vectors after the "turning" trajectory $(b_1,\ b_2, b_3)$ according to the driving direction of the agricultural machinery. The vectors a_i and b_i (I = 1,2,3) are specified as a pair of vectors, and the angle between each pair of vectors is calculated in the same way as Formula (1) in Section 2.3. Because the GNSS positioning accuracy is 5 m, the maximum lateral error of the adjacent positioning trajectory points in a straight line is 5 m, and the median travel speed of agricultural machinery in the "working" trajectory is 4.4 m/s, so the longitudinal distance of adjacent trajectory points from 1 s to 5 s interval is 4.3 m to 21.5 m, from which we can obtain that the direction change of agricultural machinery in the working trajectory is about 13.1 degrees to 49.3 degrees. Therefore, we stipulate that the angle between the front vectors (a_i) and rear vectors (b_i) is less than 10 degrees, which means that the direction of agricultural machinery has no changed. The result of whether the three directions change is obtained from the included angle of three pairs of vectors. The voting decision is adopted to determine whether the target trajectory is "turning" or "abnormal working", and the label of "abnormal working" trajectory points are recorded as 2.

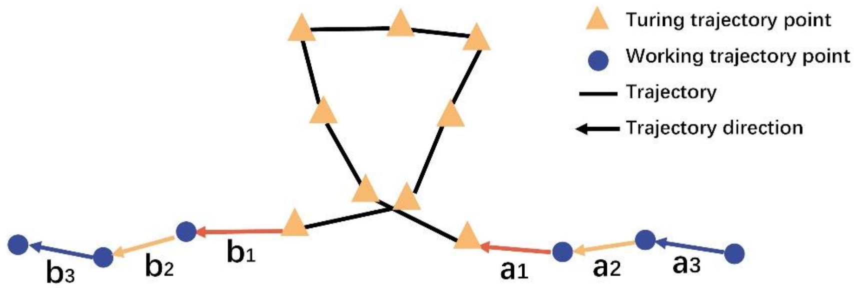

Figure 4. Schematic diagram of the turning trajectory correction method based on direction change. a_1, a_2, a_3 are the first second and third vectors before the target trajectory respectively; b_1 , b_2, b_3 are the first second and third vectors after the target trajectory respectively.

2.4.3. M3 Correction Based on the Working Characteristics of the Harvester

This section defines the begin and end positions of the "turning" trajectory in order to unify the standard since after the D-K-means method, M1 and M2 correction, the begin and end points are inconsistent with the manually marked "turning" trajectory. The algorithmic steps are:

(1) Trajectory fragmentation: The trajectory points that are consecutive in the time series and all belong to "turning" are divided into a set of trajectory fragments.
(2) Constructing features: Features are constructed with angle and dist (Euclidean distance between each trajectory point and its previous trajectory point on the time series) as attributes for each trajectory point.
(3) Locating the inflection points: Search for the inflection points in each "turning" trajectory that belong to the first clustering result and find the inflection points that appear for the first time and the last time in each "turning" trajectory segment.
(4) Correction of "turning" trajectory point: The trajectory point between 3.4 m before the first turning point and 3.4 m after the last turning point is determined as the "turning" trajectory point according to the dist property. For the trajectory containing multiple trajectory points, the length of the trajectory is calculated as the sum of the Euclidean distance of two adjacent trajectory points. In this case, the data structure of the trajectory points is $p_i = (\mathrm{time, x, y, speed, angle, dis, dist, label})$.

3. Results and Discussion

3.1. Performance Metrics

Three evaluations of precision, recall and f1-score, which are defined as Formulas (6)–(8), are used to evaluate the in-field trajectory recognition algorithm of agricultural machinery based on the K-means clustering method, and the evaluation metrics are calculated as follows:

$$precision = \frac{TP}{TP + FP} \tag{6}$$

$$recall = \frac{TP}{TP + FN} \tag{7}$$

$$f1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{8}$$

3.2. Comparisons

3.2.1. Method Comparisons

In this paper, the method of wheat harvester in-field operation trajectory identification is divided into two-step clustering and three-step correction. The results of each technique step are displayed in Figure 5 using the trajectory in one field as an example. After the D-K-means clustering method, the working trajectory and the turning trajectory are obtained as shown in Figure 5A, and the situation that the working trajectory is mistakenly identified as the turning trajectory can be seen in part a (as shown by the red dotted line). After the correction method M1 based on the clustering of trajectory fragments, this situation can be solved as shown in part a' in Figure 5B, and the situation that the abnormal working trajectory is identified as the turning trajectory can be seen in part c. After the correction method M2 based on the direction change, the situation is solved as shown in part c' in Figure 5C; after the correction M3 based on the operating characteristics of the harvester, the beginning and ending positions of the turning trajectory is defined, and the situation is solved as shown in part b' in Figure 5D (compared to part b in Figure 5A).

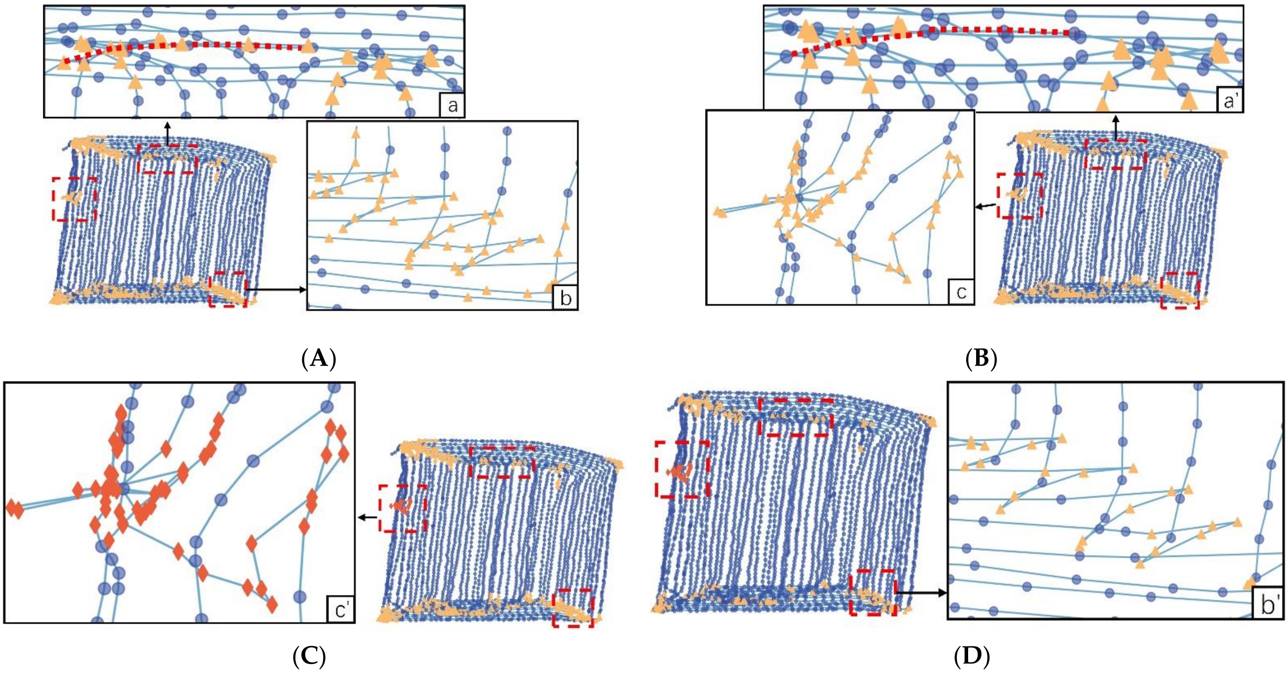

Figure 5. Step-by-step results of in-field trajectory identification of wheat harvester. (**A**) D-K-means results; (**B**) D-K-means + M1 results; (**C**) D-K-means + M1 + M2 results; (**D**) D-K-means + M1 + M2 + M3 results. (**a**) The working trajectory is mistakenly identified as the turning trajectory; (**a′**) The result after correcting the part a with the correction M1; (**b**) Turning trajectory obtained after D-k-means; (**b′**) Turning trajectory obtained after the correction M3; (**c**) The abnormal working trajectory is mistakenly identified as the turning trajectory; (**c′**) The result after correcting the part c with the correction M2. Yellow triangle is the turning trajectory point, blue circle is the working trajectory point, orange diamond is the abnormal working trajectory point.

The comparison effect of using the recognition method (D-K-means + M1 + M2 + M3) in this paper and the D-K-means method without the correction process can be seen in Figure 5A,D, and the comparison results of precision, recall, and f1-score are shown in Table 1. Table 1 shows the statistics of the step-by-step recognition results of wheat harvester trajectories in 50 fields, and the f1-score increases from 0.55 to 0.95 before and after the correction. This shows that the three correction methods can effectively identify different trajectories of harvester in-field operations based on the K-means clustering method.

Table 1. Comparison of step by step results of the trajectory recognition algorithm.

Method	Precision	Recall	F1-Score
D-K-means	0.53	0.62	0.55
D-K-means + M1	0.53	0.63	0.56
D-K-means + M1 + M2	0.85	0.94	0.88
D-K-means + M1 + M2 + M3	0.93	0.96	0.95

3.2.2. Data Comparisons

When the GNSS equipment acquisition frequency is different, the time interval of adjacent localization points in different trajectory data is different. To prove the universality of this identification method, the trajectory data with time intervals from 1 to 5 s were selected, in which 10 trajectories were chosen for each time interval, and a total of 50 trajectories were used for validation. The validation results are shown in Table 2. We can find that the f1-score of abnormal working trajectory is generally lower than working trajectory, because the abnormal working trajectory contains fewer trajectory points (usually containing tens of points) and the working trajectory contains more trajectory points (usually containing thousands of points) in a field; according to the Formulas (6)–(8) in Section 3.1, the number of misidentified same trajectory points have more influence on the recognition results of abnormal working trajectory. Therefore, it shows that the f1-score of abnormal working trajectory is generally lower than that of working trajectory.

Table 2. Comparison of algorithm results for trajectory data at different time intervals.

Time Interval	Turing			Working			Abnormal Working		
	Precision	Recall	f1-Score	Precision	Recall	f1-Score	Precision	Recall	f1-Score
1 s	0.84	0.98	0.90	1.00	0.96	0.98	0.92	0.97	0.94
2 s	0.94	0.95	0.95	0.99	0.99	0.99	0.90	0.94	0.92
3 s	0.94	0.93	0.93	0.98	0.98	0.98	0.84	0.98	0.90
4 s	0.94	0.96	0.95	0.99	0.98	0.99	0.90	0.96	0.93
5 s	0.92	0.97	0.94	0.99	0.98	0.99	0.90	0.94	0.91

Furthermore, the f1-score for the recognition results for turning trajectory, working trajectory, and abnormal working trajectory are all above 0.90, indicating that this method has good performance in the trajectory data of 1 to 5 s time intervals and can better recognize trajectory data of different frequencies. The values of these time intervals can be referred to in the application work.

3.3. Application and Extension

3.3.1. Area calculation Application

The distance algorithm is one of the ways to calculate the area of the field [25,26]. It is based on the principle that the width of the tractor multiplies the distance it travels. However, because the wheat harvester's turning trajectory and abnormal working trajectory in the field overlap with the working trajectory, the calculated area does not match with the real field area. After identifying the trajectory of wheat harvester in-field, the distance algorithm is used to calculate the area after removing the turning trajectory and abnormal working trajectory, which effectively improves the accuracy of the area calculation result. As Figure 6 represents the real area, the calculated area before processing, and the calculated area after processing for the trajectory data of 50 fields, processing means removing the turning trajectory and the abnormal working trajectory, and it can be seen that the calculated value of field area is closer to the real area after processing. Additionally, the 50 fields' real total area S is 51.13 ha, their calculated area without processing S_{before} is 73.10 ha, and their area after processing S_{after} is 64.39 ha. The calculation error is reduced by 17.04%, and the error reduction was calculated as shown in Formula (9).

$$\text{error reduction} = \frac{\left(S_{before} - S\right)}{S} - \frac{\left(S_{after} - S\right)}{S} = \frac{S_{before} - S_{after}}{S} \tag{9}$$

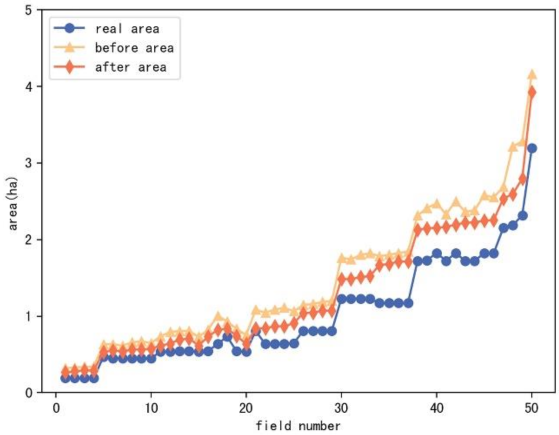

Figure 6. The real area, the calculated area before processing, and the calculated area after processing for the trajectory data of 50 fields.

The 50 field trajectories are sampled at intervals from 1 s to 5 s, and each time interval contains 10 fields. The trajectory data of 10 fields with the same time interval are sorted according to the ratio of the number of turning and abnormal working trajectory points to the total number of field trajectory points in the algorithm results of this paper, from small to large, as 1 to 10, and they calculate the error change of the calculated area of each field before and after treatment.

The results are shown in Figure 7, where a bar represents the sum of the percentages of error reduction in a certain percentage ranking for the five-time interval data. It can be seen that when the ratio of turning and abnormal working trajectory points to the total number of trajectory points is larger, the calculation error of farmland area before and after treatment decreases more. This is because when the distance algorithm is used to calculate the area before processing, the turning trajectory and the abnormal working trajectory belong to the repeated calculation part. The larger the proportion of this part, the larger the calculation error is. In summary, the in-field trajectory recognition algorithm of agricultural machinery in this paper can provide technical support for the calculation of field area.

3.3.2. Data Extension

The recognition algorithm relies on the wheat harvester trajectory to identify the "turning", "working", and "abnormal working" trajectories. The algorithm can also be applied to the trajectory data of other types of crops and other types of farm machinery, as long as these trajectories satisfy the characteristics mentioned in Section 2.1.

We add corn harvesting and rice harvesting data to validate the algorithm. Figure 8A shows the trajectory in the rice harvesting field of model 4LZ-8M6 (the length, width, and height are 7000 mm, 2980 mm, and 3470 mm, respectively), and Figure 8B shows the trajectory in the corn harvesting field of model 4YZ-4B6 (the length, width, and height are 7250 mm, 2750 mm, and 3500 mm, respectively). This shows that the recognition algorithm can be extended to the trajectory data of other types of crops and other types of agricultural machines with good performance.

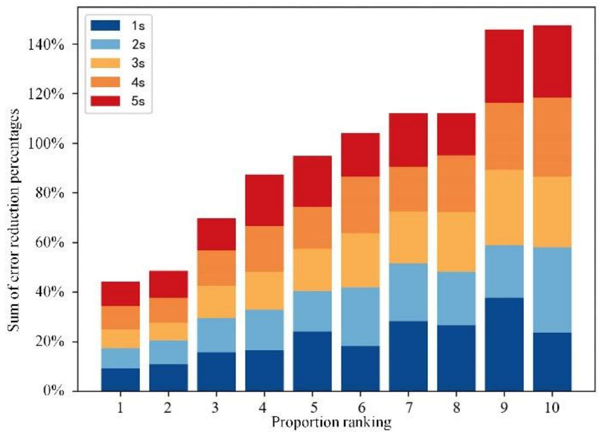

Figure 7. Percentage reduction in area error for each kind of time interval data processed afterward as compared to data processed without processing. The data of the same time interval are considered as a group, the x-axis coordinate represents the ranking of a field in its group, and the y-axis coordinate represents the sum of error reduction percentage of five fields (different time intervals) belonging to the same ranking.

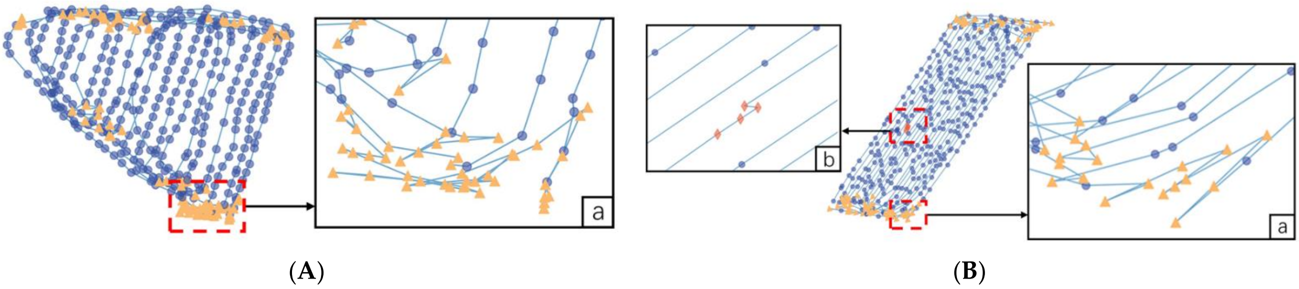

Figure 8. Trajectory identification results of rice and corn. (**A**) Rice harvesting trajectory; (**B**) corn harvesting trajectory. (**a**) Turning trajectory identification result. (**b**) Abnormal working trajectory identification result.

4. Conclusions

Identifying the in-field trajectory of harvester based on the K-means clustering method and three-step correction method, in order to effectively identify the "working" trajectory, "turning" trajectory, and "abnormal working" trajectory of harvesters in the field. The D-K-means method can initially obtain the "working" trajectory and "turning" trajectory, and the final recognition results of the three trajectories can be obtained by the three correction method. After analyzing the recognition results of 50 fields, all the three trajectory recognition results have good performance. It is useful to reduce the error in the application of area calculation, and the verification algorithm using rice harvesting trajectory and corn harvesting trajectory proves that the identification algorithm can be effectively extended to the data of other crops.

The trajectory identification method is for wheat harvester trajectory in-field. Compared with many studies on road vehicles and agricultural machinery trajectory recognition in field, this study provides a simple and effective method to identify different trajectories of harvesters in-field in agricultural scenarios and supports the calculation of field area. The "turning" trajectory is only for the "X-turn" trajectory, and the "abnormal working" trajectory is only for the trajectory that the agricultural machine turns and resumes the direction afterwards. On the basis of this algorithm, future research may take into account other kinds of "turning" trajectories or other kinds of agricultural machinery, as well as the improvement of farm machinery operation effectiveness, the improvement of agricultural machinery operation routes, and the development of a digital management platform for agricultural machinery.

Author Contributions: Conceptualization, L.Y.; Data curation, X.W.; Methodology, X.W.; Re-sources, C.W.; Supervision, L.Y. and Y.X.; Writing—original draft, X.W.; Writing—review and editing, L.Y., X.W., Y.L., Z.X., and R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Precision Agriculture Application Project (JZNYYY001).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reith, S.; Frisch, J.; Winkler, B. Revision of the Working Time Classification to Optimize Work Processes in Modern Agriculture. *Chem. Eng. Trans.* **2017**, *58*, 121–126.
2. Michael, F. Kocher Viacheslav Robert D Grisso, Field Efficiency Determination Using Traffic Pattern Indices. *Biol. Syst. Eng.* **2004**, *20*, 563.
3. Duttmann, R.; Brunotte, J.; Bach, M. Spatial Analyses of Field Traffic Intensity and Modeling of Changes in Wheel Load and Ground Contact Pressure in Individual Fields During a Silage Maize Harvest. *Soil Tillage Res.* **2013**, *126*, 100–111. [CrossRef]
4. Kong, Q.; Maimaiti, T.; Zhao, M. Recognition of tractor working condition based on convolutional neural network. *China Agric. Mach. Chem. News* **2021**, *42*, 144–150.
5. Chen, Y.; Zhang, X.; Wu, C.; Li, G. Field-Road Trajectory Segmentation for Agricultural Machinery Based On Direction Distribution. *Comput. Electron. Agric.* **2021**, *186*, 106180. [CrossRef]
6. Zhang, X.; Chen, Y.; Jia, J.; Kuang, K.; Lan, Y.; Wu, C. Multi-View Density-Based Field-Road Classification for Agricultural Machinery: DBSCAN and Object Detection. *Comput. Electron. Agric.* **2022**, *200*, 107263. [CrossRef]
7. Poteko, J.; Eder, D.; Noack, P.O. Identifying Operation Modes of Agricultural Vehicles Based On GNSS Measurements. *Comput. Electron. Agric.* **2021**, *185*, 106105. [CrossRef]
8. Chen, Y.; Li, G.; Zhang, X.; Jia, J.; Zhou, K.; Wu, C. Identifying Field and Road Modes of Agricultural Machinery Based On GNSS Recordings: A Graph Convolutional Neural Network Approach. *Comput. Electron. Agric.* **2022**, *198*, 107082. [CrossRef]
9. Jensen, M.A.F.; Bochtis, D. Automatic Recognition of Operation Modes of Combines and Transport Units Based On GNSS Trajectories. *IFAC Proc. Vol.* **2013**, *46*, 213–218. [CrossRef]
10. LiYang, l.; Tian, W.; Zhai, W.; Wang, X.; Chen, Z.; Wen, L.; Xu, Y.; Wu, C. Behavior Recognition and Fuel Consumption Prediction of Tractor Sowing Operations Using Smartphone. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 154–162.
11. Wang, P.; Meng, Z.; Yin, Y.; Fu, W.; Chen, J.; Wei, X. Automatic recognition algorithm of field operation status based on spatial track of agricultural machinery and corresponding experiment. *J. Agric. Eng.* **2015**, *31*, 56–61.
12. Taylor, R.K.; Schrock, M.D.; Staggenborg, S.A. *Extracting Machinery Management Information From GPS Data*; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2002.
13. Li, Y. Research on Lean Operation and Key Problems of Field Crop Mechanized Collection and Transportation System. Doctor's Thesis, Chongqing University, Chongqing, China, 2015.
14. Tian, W. Study on Driving Conditions and Fuel Consumption of Agricultural Machinery Based on Machine Learning. Master's Thesis, China Agricultural University, Beijing, China, 2022.
15. Shen, Y.; Mercatoris, B.; Cao, Z.; Kwan, P.; Guo, L.; Yao, H.; Cheng, Q. Improving Wheat Yield Prediction Accuracy Using LSTM-RF Framework Based on UAV Thermal Infrared and Multispectral Imagery. *Agriculture* **2022**, *12*, 892. [CrossRef]
16. Nehrey, M.; Zomchak, L. Digital Technology: Emerging Issue for Agriculture. In *The International Conference on Artificial Intelligence and Logistics Engineering*; Springer: Cham, Switzerland, 2022; pp. 146–156.
17. Qing, Y.; Li, Y.; Yang, Y.; Xu, L.; Ma, Z. Development and experiments on reel with improved tine trajectory for harvesting oilseed rape. *Biosyst. Eng.* **2021**, *206*, 19–31. [CrossRef]
18. Ding, Q.; Ding, W.; Yang, W.; Li, Y.; Tian, Y.; Zhu, Y. Mechanization characteristics of farmland fragmentation conditions—Investigation on field operation behavior of small harvesters. *Zhejiang Agric. J.* **2013**, *25*, 1397–1403.
19. Lovor Heavy Industries Co. *Lovol Gushen Self-Propelled Grain Combine Harvester Manual*; Lovor Heavy Industries Co.: Shandong, China, 2017.
20. Lu, W.; Su, H.; Yang, X.; Yan, X. Subsurface Temperature Estimation From Remote Sensing Data Using a Clustering-Neural Network Method. *Remote Sens. Environ.* **2019**, *229*, 213–222. [CrossRef]
21. Sinaga, K.P.; Yang, M. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]
22. Gu, C.; Qi, J.; Zhao, Y.; Yin, W.; Zhu, S. Estimation of the Mixed Layer Depth in the Indian Ocean from Surface Parameters: A Clustering-Neural Network Method. *Sensors* **2022**, *22*, 5600. [CrossRef] [PubMed]
23. Zhang, Z. Trajectory data mining research based on K-Means and DBSCAN. *China's Strateg. Emerg. Ind.* **2017**, *44*, 113–114.
24. Zhou, X. Research on Big Data Reduction Based on Voting Mechanism and Open Source Framework. Master's Thesis, Hebei University, Baoding, China, 2021.

25. Liu, H.; Meng, Z.; Wang, P.; Wei, X.; Han, Y. Buffer area algorithm based on agricultural machinery space trajectory. *J. Agric. Eng.* **2015**, *31*, 180–184.
26. Wu, C.; Chen, Y.; Yang, W. Construction of big data system of agricultural machinery based on BeiDou. *J. Agric. Eng.* **2022**, *38*, 1–8.