An Efficient Algorithm with Distance-based Clustering for Detecting Turning Regions on GPS Trajectories*

Yu-Cheng Lin[1], Chang-Biau Yang[1]†, Kuo-Si Huang[2] and Yung-Hsing Peng[3]
[1]Department of Computer Science and Engineering,
National Sun Yat-sen University, Kaohsiung, Taiwan
[2]Department of Business Computing,
National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan
[3]Digital Transformation Research Institute,
Institute for Information Industry, Kaohsiung, Taiwan

Abstract

In 2023, Peng *et al.* first defined the turning region detection (TRD) problem. Given a GPS trajectory, consisting of a sequence of GPS positions, the TRD problem aims to identify the number of turning regions in the sequence. By analyzing the GPS trajectories, we observe that the number of turns is related to the specified minimum distance, accumulated distance, and angle variation. Based on these parameters, we propose an efficient algorithm for solving the TRD problem. As the experimental results show, the proposed TRD algorithm outperforms the previous TRD algorithms.

Keywords: Turning Region Detection, Clustering, GPS Trajectory, Polar Coordinate System.

1 Introduction

The tea industry in Taiwan has a long and rich history, tracing back to the 1700s, with a legacy spanning over centuries. It constitutes a vital part of Taiwanese agriculture. Evolving from manual harvesting in the early days to comprehensive mechanization in modern times, the harvesting and production techniques in the Taiwanese tea industry continue to undergo continuous innovation.

According to the statistics from 2022 to 2024, the government annually invests 23 billion New Taiwan Dollars to support the development of intelligent and energy-efficient agricultural machinery, aiming to enhance Taiwan's agricultural competitiveness on the global stage [1]. Agricultural machinery plays a pivotal role in the tea plantations, making the effective utilization of these machines a crucial concern.

We observe that the development of agricultural machinery software in Taiwan has not yet been widely discussed. In other words, research on agricultural machinery software remains a topic worthy of exploration [6, 11]. Some researchers attempted to propose solutions for tracking agricultural machinery usage using the global positioning system (GPS) [5, 7], with most involving techniques related to data processing and computational algorithms. Recently, Peng *et al.* [9] introduced an agricultural database aimed at sharing data essential for future agricultural machinery services.

In the practical application of GPS trajectories, the *turning region detection* (TRD) problem, defined by Peng *et al.* [9], tries to determine whether each point in a GPS trajectory is a turning state or not. One of the applications of the TRD is to estimate working efficiency of agricultural machinery by calculating straight and turning speeds, contributing to effective machinery management. Peng *et al.* [9] designed a TRD algorithm with linear time that is easy to implement for building the cloud service of Chang-Shen Tea Factory (CSTF). After that, Huang *et al.* [4] designed a TRD algorithm incorporating the concept of azimuth clustering. They mentioned that the accuracy might be improved based on their experimental results.

In this paper, we propose an efficient TRD algorithm for improving the results of previous al-

*This research work was partially supported by the National Science and Technology Council of Taiwan under contract NSTC 112-2221-E-110-026-MY2.

†Corresponding author. E-mail:
cbyang@cse.nsysu.edu.tw (Chang-Biau Yang).

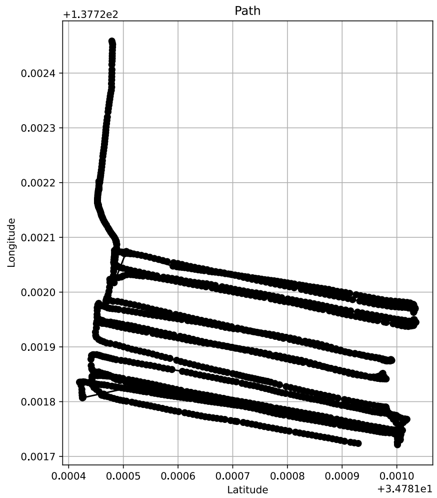

Figure 1: An example of GPS trajectory of 1493 points.

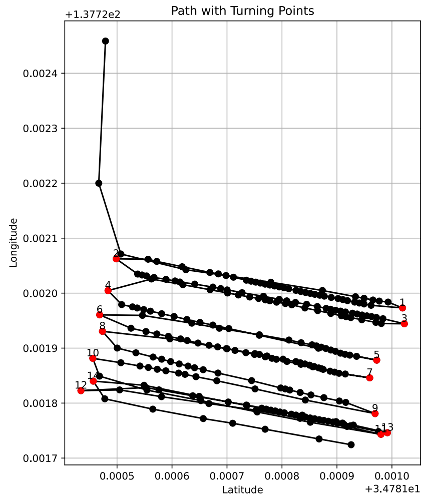

Figure 2: The result of applying turning region detection (TRD) for Figure 1.

gorithms [4, 9]. Our TRD algorithm adopts the concept of polar coordinate system, allowing for a more precise analysis of GPS trajectory. Experimental results show that our TRD algorithm outperforms previous methods, with average accuracy improved from 70% to 90%.

2 Preliminaries

2.1 GPS Trajectory

For agricultural management, the efficiency of automated harvesting is crucial as it significantly impacts production costs. Therefore, Peng *et al.* [9] utilized a GPS tracker to record and report the movement of agricultural machinery, specifically for documenting the harvesting process.

The GPS tracker collects data in the format of $(ID, time, lat, lon)$ and automatically uploads the records to the server. For a record $(ID, time, lat, lon)$, ID represents one point of a GPS tracker, $time$ denotes the recorded time of this entry, and lat and lon are the latitude and longitude, respectively.

Figure 1 illustrates a visual example of a sample trajectory of 1493 GPS points. It is not difficult for humans to observe the positions and quantities of turns that occur with the naked eye. However, as the path becomes increasingly complex, manual calculation becomes more difficult.

2.2 Turning Region Detection

Definition 1. (turning region detection, TRD) *Given a sequence of GPS points* $P = \langle p_1, p_2, \cdots, p_n \rangle$*, where each* $p_i = (lat_i, lon_i)$ *represents its latitude and longitude, respectively, the* turning region detection *(TRD) problem aims to determine the number of turns in* P *and corresponding GPS positions.*

For detecting the turning points, we may analyze the angular variation of each vector in the trajectory. Based on the angular variation of this quantity, we determine whether it satisfies the threshold θ for turning.

In case of a GPS trajectory containing n points, the TRD problem aims to output a number t, representing the number of turns. Figure 2 illustrates the turning points in Figure 1 by applying turning points detection (TRD). It is clear to see the location where the turn occurs as well as its corresponding label.

2.3 The Polar Coordinate System

The *polar coordinate system* [3] is a two-dimensional coordinate system commonly used in mathematics and physics. Unlike the Cartesian coordinate system, which uses x and y axes, the polar system represents points in terms of *radius* (distance) and *angle*. In the polar coordinates systems as shown in Figure 3, a point is denoted by

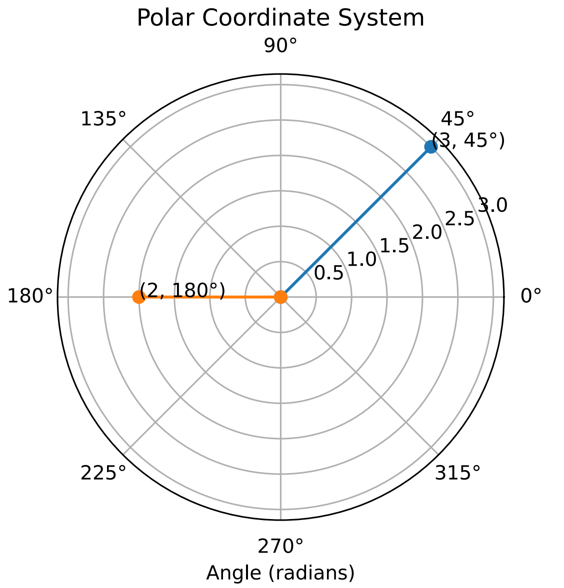

Figure 3: An illustration of the polar coordinate system [3].

its distance from a reference point (usually the origin) and the angle formed by the line connecting the point to the reference point with respect to a reference direction. The distance is known as the radial coordinate, denoted by r, and the angle is called the angular coordinate.

This paper employs the polar coordinate system for more effective path analysis. In our approach, it calculates vectors between neighboring points, utilizing the angle of polar coordinate system to express the direction of these vectors. It simplifies the analysis of path variations, particularly when focusing on the number of turns and turning points.

3 Our Algorithm

3.1 Distance-based Clustering

Clustering [2, 8] is a fundamental concept in data analysis for grouping data points with similar characteristics. In this paper, we focus on the spatial domain, specifically clustering coordinates that are close to each other in terms of distance.

Let $\text{dis}(p_i, p_{i-1})$ represent the Euclidean distance between p_i and p_{i-1}. In Algorithm 1, the *haversine formula* [10] is utilized to calculate the spherical distance $\text{dis}(p_i, p_{i-1})$ between neighboring points. It begins by scanning coordinates and calculating distances between consecutive points.

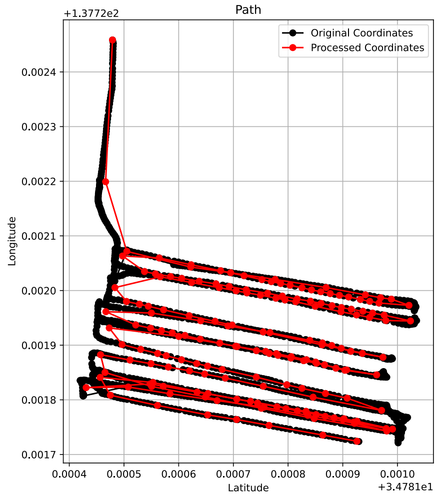

Figure 4: The result of distance-based clustering for Figure 1.

A distance threshold γ and an accumulated distance threshold δ are used to determine when to form a new cluster. When the distance between two points surpasses γ or the accumulated distance of a cluster exceeds δ, a new cluster is initiated. Subsequently, we calculate the average coordinates for each cluster, serving as the representative of the group.

Figure 4 illustrates coordinates clustered by distance-based clustering for Figure 1. It can be observed that although the number of points has decreased significantly, and the overall structure of the graph remains.

3.2 Turning Region Detection

Next, we use a vector to express the direction of two neighboring clusters. The direction change between two consecutive vector can be represented by an angle in the polar coordinate system. If the angle exceeds a predefined threshold θ, the center cluster (3 clusters form two consecutive vectors) is decided as a turning point.

Our algorithm for calculating the number of turning regions is presented in Algorithm 2. If we desire to output the coordinates of turning regions, we can retrieve the coordinates from the clusters which are determined to be turning regions.

By using the Matplotlib, we create a coordinate path plot that highlights turning points, offering a more intuitive visualization. This approach contributes to a deeper understanding of spatial dy-

Algorithm 1 Distance-based clustering (DBC).

Input: A chronologically ordered GPS trajectory $P = \langle p_1, p_2, p_3, \cdots, p_n \rangle$, a minimum distance threshold γ, an accumulated distance threshold δ, where $p_i=(lat_i, lon_i)$.
Output: $C=DBC(P, \gamma, \delta)$, a cluster sequence of coordinates.

for $i \leftarrow 1$ to $|P|-1$ **do** ▷ Remove anomalous or duplicated points.
  if $p_i = (0, 0)$ or $p_i = p_{i+1}$ **then**
    remove p_i from P
$C \leftarrow \{p_1\}$, $G \leftarrow \{p_2\}$, $d \leftarrow 0$
for $i \leftarrow 3$ to $|P|$ **do**
  if $\mathrm{dis}(p_{i-1}, p_i) < \gamma$ and $d < \delta$ **then**
    insert p_i to G
    $d \leftarrow d + \mathrm{dis}(p_i, p_{i-1})$ ▷ Accumulated distance from the first point in G.
  else ▷ A cluster is formed.
    $\bar{p} \leftarrow \mathrm{average}(G)$ ▷ $\bar{p}$: average position
    insert $\bar{p}$ to C
    $G \leftarrow \{p_i\}$, $d \leftarrow 0$
return C

namics between coordinates, particularly in path analysis, laying a valuable foundation for further research.

Algorithm 2 Calculation of the number of turning regions.

Input: A chronologically ordered GPS trajectory $P = \langle p_1, p_2, p_3, \cdots, p_n \rangle$ and an angle threshold θ.
Output: The number of turning regions, t.

$C = \langle c_1, c_2, \cdots, c_m \rangle \leftarrow DBC(P, \gamma, \delta)$
$t \leftarrow 0$
for $i \leftarrow 2$ to $m-1$ **do**
  $u \leftarrow c_i - c_{i-1}$ ▷ u: vector 1
  $v \leftarrow c_{i+1} - c_i$ ▷ v: vector 2
  $A \leftarrow \mathrm{atan2}(u) - \mathrm{atan2}(v)$ ▷ angle diff.
  if A is in range of $\pm$ (360 - θ) **then**
    $t \leftarrow t + 1$ ▷ a turning point
return t

4 Experimental Results

The experimental dataset of twenty GPS trajectories is provided by the Chang Shen Tea Factory (CSTF). We analyze these trajectories collected by agricultural machines in tea gardens.

Table 1: Numbers of turning regions predicted by our TRD algorithm for various values of θ with $\gamma = 0.9$ and $\delta = 50$. Here, T is the correct number of turns by manual identification.

D_i \ θ	60	65	70	75	80	85	T
D_1	37	34	**33**	**33**	**33**	**33**	32
D_2	25	24	24	23	**22**	21	22
D_3	**37**	**37**	**37**	**37**	34	34	36
D_4	143	132	119	107	98	**91**	91
D_5	55	51	50	48	**46**	45	46
D_6	23	23	23	22	**21**	**21**	19
D_7	82	76	71	69	65	**60**	59
D_8	55	50	47	44	41	**40**	40
D_9	12	12	12	**11**	**11**	**11**	10
D_{10}	59	55	53	49	46	**41**	42
D_{11}	**8**	7	7	7	7	7	8
D_{12}	**22**	**22**	**22**	**22**	**22**	**22**	21
D_{13}	**47**	**47**	46	44	43	42	50
D_{14}	71	68	68	**66**	65	62	66
D_{15}	17	**16**	14	14	14	14	16
D_{16}	**14**	**14**	**14**	**14**	**14**	**14**	15
D_{17}	**46**	43	41	37	35	34	46
D_{18}	41	**40**	38	34	32	32	40
D_{19}	**36**	**32**	31	31	29	28	34
D_{20}	46	44	44	**40**	38	37	41

According to the definition of the TRD problem, it outputs the numbers of turns and coordinates of turning points for a specific trajectory. In our experiments, we compare the number of turns obtained by our TRD algorithm for each case with the number of turns manually identified by humans. We can evaluate the performance of the TRD algorithm and understand gaps in practical applications.

In our TRD algorithm, the setting of parameters γ, δ and θ significantly affect the results. According to the preliminary experiment, $\gamma = 0.9$ and $\delta = 50$ results in smaller discrepancies between experimental results and the expected outcomes. Therefore, we apply this configuration to further examine parameter θ.

Table 1 presents the outcomes corresponding to various threshold settings of θ with $\gamma = 0.9$ and $\delta = 50$, where T represent the target number of turns by manual identification. The red markings in Table 1 represent the corresponding θ whose results are closest to target T for each case.

Figure 5 shows the average errors of various values {60, 65, 70, 75, 80, 85} of parameter θ. It shows that our proposed algorithm has the lowest error rate when $\theta = 80$.

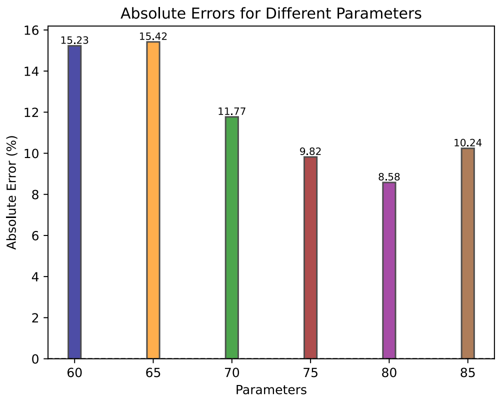

Figure 5: Average errors of various values of $\theta \in \{60, 65, 70, 75, 80, 85\}$.

Table 2: Comparison of two previous TRD algorithms [4, 9] and our algorithm with $\theta = 80$ on datasets D_1 to D_{12}. Here, T is the correct number of turns by manual identification.

D_i \ Alg.	Peng [9]	Huang [4]	Ours	T
D_1	26	**32**	33	32
D_2	**22**	24	**22**	22
D_3	63	30	**34**	36
D_4	100	**86**	98	91
D_5	33	53	**46**	46
D_6	50	30	**21**	19
D_7	**58**	61	65	59
D_8	45	45	**41**	40
D_9	33	**11**	**11**	10
D_{10}	56	**41**	46	42
D_{11}	28	**9**	**7**	8
D_{12}	46	23	**22**	21

Table 2 represents the results of each algorithm on datasets D_1 through D_{12}. According to the table, it is evident that our proposed algorithm has significantly higher accuracy compared to the other two.

5 Conclusions

This paper discusses the turning region detection (TRD) problem proposed by Peng *et al.* in 2023, detecting turning regions of a GPS trajectory to further evaluate the efficiency of agricultural machinery and the proficiency of operators. Through the utilization of distance-based clustering and the polar coordinate system, our approach provides a robust framework for analyzing trajectory paths with enhanced accuracy. It contributes to the field of trajectory analysis by presenting an improved algorithm for detecting turning points. The TRD algorithm offers a valuable tool for applications in agricultural machinery management, transportation logistics, and other domains reliant on accurate trajectory analysis. It may involve further optimization of the algorithm and exploration of its applications in diverse real-world scenarios.

Acknowledgements

The authors would like to thank Chang Shen Tea Factory and Myshine Technology Co., Ltd. in Taiwan for sharing the real GPS trajectories of agricultural machinery and providing technical advice on GPS data processing, which help us publish this paper.

References

[1] "Taiwan agricultural statistics website, Ministry of Agriculture, Taiwan." `https://agrstat.moa.gov.tw/sdweb/public/inquiry/InquireAdvance.aspx`. Accessed: 2024-03-14.

[2] P. C. Besse, B. Guillouet, J. M. Loubes, and F. Royer, "Review and perspective for distance-based clustering of vehicle trajectories," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 11, pp. 3306–3317, 2016.

[3] S. F. Catalono, "Trajectory equations of motion in radar polar coordinates," tech. rep., MIT Lincoln Laboratory, 1967.

[4] K.-S. Huang, Y.-C. Lin, C.-B. Yang, H.-C. Lin, Y.-H. Peng, and S.-H. Wang, "An improved algorithm with azimuth clustering for detecting turning regions on GPS trajectories," *The 28th International Conference on Technologies and Applications of Artificial Intelligence*, Yunlin, Taiwan, 2023.

[5] W. Y. Huang and T. L. Liu, "The development of a new cutting machine for tea harvester," *The Professional Report of Tea Industry in Taiwan*, Vol. 122, pp. 13–14, Dec 2022.

[6] M. M. Kurdi and I. A. Elzein, "Trajectory and motion for agricultural robot," *2022 In-*

ternational Conference on Decision Aid Sciences and Applications (DASA), IEEE, 2022.

[7] J. Li, X. Pei, X. Wang, D. Yao, Y. Zhang, and Y. Yue, "Transportation mode identification with gps trajectory data and gis information," *Tsinghua Science and Technology*, Vol. 26, No. 4, pp. 403–416, 2021.

[8] J. Muckell, J. H. wang, C. T. Lawson, and S. S. Ravi, "Algorithms for compressing GPS trajectory data: an empirical evaluation," *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010.

[9] Y.-H. Peng, Y.-H. Chen, H.-C. Lin, P.-C. Huang, and C.-P. Lee, "The development of agricultural machinery data bank with GPS as an application study," *The 40th Symposium on Combinatorial Mathematics and Computational Theory*, Taoyuan, Taiwan, 2023.

[10] R. W. Sinnott, "Virtues of the haversine," *Sky and Telescope*, Vol. 68, No. 2, p. 158, 1984.

[11] W. Zhuang, W. Chun, and H. Jing, "Development of agriculture machinery aided guidance system based on GPS and GIS," *2010 World Automation Congress*, IEEE, 2010.