The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-G-2025
ISPRS Geospatial Week 2025 "Photogrammetry & Remote Sensing for a Better Tomorrow…", 6–11 April 2025, Dubai, UAE

Incremental Crowd-Source Data Fusion and Map Update Method Based on Driving Data for Traffic Signs

Haopeng Hu[1], Hangbin Wu[1,*], Shengke Huang[1], Wei Huang[1], Chun Liu[1]
[1]College of Surveying and Geo-informatics, Tongji University, Shanghai, China – (haopenghu, hb, peanuts, wei_huang, liuchun)@tongji.edu.cn

Keywords: High-Definition Map, Crowd-source Data Fusion, Traffic Sign, Map Update.

Abstract

Traffic signs provide important traffic information for automatic driving, and accurate and complete traffic sign data of HD (High Definition) map provides important data support for intelligent transportation, automatic driving and other emerging service industries. Driving record data fills the data gap of crowd-source updating in HD maps, and the crowd-source updating method of road traffic facilities in HD maps using massive driving record data has become a new research hotspot. In this paper, an incremental HD map traffic sign crowd-source update method is proposed based on the driving record data. The traffic sign detection results are matched with the existing traffic signs in the HD map for traffic sign change detection, and the added results are optimized and fused for position, and the new sign positions are optimized using the unchanged signs to obtain the optimized new traffic sign positions. The experiments in Shanghai show that the matching method can meet the matching requirements of crowd-source updating; the accuracy of the traffic sign positions after position optimization and crowd-source fusion is obviously improved, with an average plane error of 3.69 m and a standard deviation of error of 3.29 m, which can provide data support for crowd-source updating of the HD map.

1. Introduction

HD (High Definition) maps play a crucial role in industries such as autonomous driving, transportation, and geographic information. While the construction of HD maps is advancing smoothly, maintaining the updating of HD maps and the consistency between the information on HD maps and reality has become a new challenge. If HD maps cannot ensure continuous and efficient updates, the safety risks caused by lagging information will continue to increase, which will ultimately affect driving safety. Road traffic signs, as one of the elements of HD maps, will change with changes in the road environment, such as temporary traffic control, lane closure, or long-term road maintenance and rerouting. How to update traffic signs of HD maps has also become one of the new research directions in the field of HD maps.

The current HD map data collection methods are divided into two types: professional collection and crowd-source collection; the former uses professional equipment to collect road data with high precision, while the latter is derived from crowd-source data with low precision but wide collection range. However, the high cost and low update efficiency of professional equipment make it difficult to adapt to the demand of rapid update of HD maps (Lai et al., 2019; Qin et al., 2020; Yang et al., 2019; Yang, 2018). Therefore, crowd-source data with wide coverage, large data volume and high timeliness become an important data source in the study of HD map updating, such as traffic record data.

In this paper, a method of traffic sign crowd-source fusion and map updating based on driving record data is proposed. This method can use vehicles equipped with different equipment to make multiple repeated observations of road traffic signs and detect changes in the observation results with the existing HD maps in order to realize crowd-source updating of HD map change areas.

2. Related Work

Compared with professional data, crowd-source data, although less accurate, is characterized by large data volume and wide coverage, which is suitable for low-cost updating work of HD maps (Li, 2022). In some algorithms, the step of element position calculation is skipped, and change detection is performed directly based on images, and the changes of objects in the scene are judged by image alignment, feature extraction, and similarity determination (Yang et al., 2023). Alcantarilla (Alcantarilla et al., 2018) used the street video from a vehicle-mounted monocular camera to coarsely match the images of the same area based on the process of SLAM and 3D reconstruction using a Deep Inverse Convolutional Network for pixel-level change detection; Zohar (Zohar et al., 2012) performed change detection of the scene based on the changes of 3D line elements in the image; Park (Park et al., 2022) proposed a vehicle-cloud collaborative map updating system that uses the YOLO model to extract road traffic signs in the driving record scene and compares the extraction results with the traffic signs within the HD map to confirm whether the elements need to be updated; Zhang (Zhang et al., 2021) use the extraction results of semantic segmentation, and using SLAM algorithms can calculate the camera position and attitude between sequential images to obtain the lane marking point cloud, vectorize the point cloud results and compare them with the existing HD maps, and update the changed elements; and Kim (Kim et al., 2018) address the need to establish new layers in the HD maps by using the data of intelligent vehicles, the GraphSLAM is utilized to create new element layers. However, although these methods can acutely sense the changes occurring in the environment, they are unable to compute the position of the added elements. In order to realize crowd-source updating of HD maps, using the results of position calculation of map elements is still an important source of update information.

Considering that there is still a large deviation from the true value in the position calculation results of map elements based on crowd-source data, even if crowd-source data collected by devices such as commercial-grade vehicle-mounted cameras and GNSS equipment, the 3D reconstruction accuracy is still at the sub-meter level (Das et al., 2020; Jo et al., 2018; Kim et al., n.d.; Yang et al., 2023), and the 3D reconstruction error of vehicle recording data collected by cell phones or vehicle recording devices will also be at the meter level. Many researches take the massive data at the vehicle end, store, compute, and fuse them in the cloud, and then publish the update results to form a complete set of updating framework and updating strategy (Kim et al., 2021, 2018; Tchuente et al., 2021; Yang et al., 2023). A set of updating architecture for traffic signs proposed by Tchuente (Tchuente et al., 2021) forms a framework for sensing and capturing the traffic signs at the vehicle end, statistically

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-641-2025 | © Author(s) 2025. CC BY 4.0 License.

analysing the massive data in the cloud and matching it with the HD maps, and then the update results are returned to the vehicle end in a crowd-source updating and fusion model. These studies provide a solid theoretical foundation for the development of this project.

3. Method

3.1 Overview

Compared with the existing traffic signs in the HD map, the relationship between the detection results of individual traffic signs and the actual signs may exist in the following five cases: no change signs, only semantic change signs, only position change signs, new signs, and removed signs. Based on the characteristics of driving record data, this study proposes an incremental HD map traffic sign crowd-source fusion update method, which can realize independent processing of each segment of driving record data and crowd-source fusion of multi-segment trajectory traffic sign detection results, and its processing flow is shown in Figure 1.

For a section of driving record data, the traffic sign detection results are first filtered according to the confidence level, and then matched with the HD map traffic signs to distinguish between existing signs and new signs. For the existing signs, the semantic fusion results are used to determine the semantic update; for the new signs, the coordinates are first optimized using the matching results of the existing signs, and then combined with the detection results of other traffic record data for new fusion. Considering that the position and semantics of the traffic signs in the detection results have greater uncertainty compared to the HD map layer, the results of the new traffic signs and the change information of the traffic signs should not be directly changed to the traffic sign layer, this study proposes to establish a traffic sign change layer between the detection results layer and the HD map layer, and to fuse the detection results of the multi-segment trajectories from crowd-source data. During the addition fusion process, the added signs are again matched with the signs in the change layer to merge the added results of the multi-segment trajectory data detection.

Therefore, the incremental HD map traffic sign crowd-source updating method proposed in this study mainly involves the following steps, which are introduced in the subsequent contents of this section.

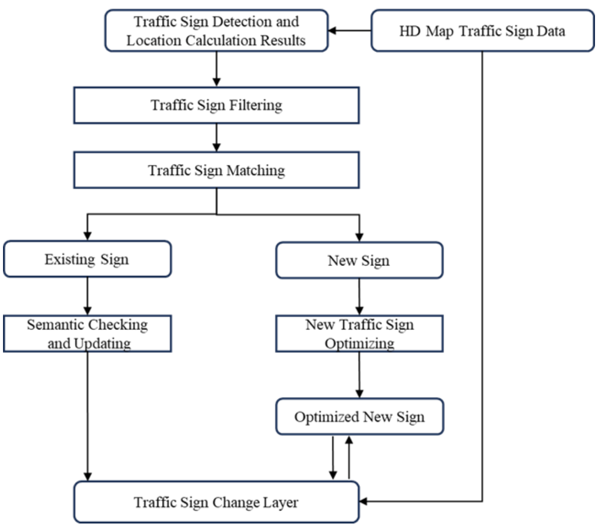

Figure 1. Proposed method.

3.2 Data Collection

In this study, we developed a driving record data collection software based on the Android smartphone platform. In addition to recording the basic positioning information, we recorded the gravity sensor data, acceleration sensor data and other cell phone sensor information in the cell phone at the same time as we recorded the images of the sampling points, and the data structure and the software interface are shown in Table 1 and Figure 2, respectively.

Data item	Data type	Meaning
Longitude	floating point	Sampling point position longitude
Latitude	floating point	Sampling point position latitude
Timestamp	Time	Sampling point record time stamp
Azimuth	Float	Direction of the sampling point
Speed	Float	Vehicle speed at the sampling point
Image	String	Image path of the sampling point
Gravity	3×1 vector	Gravity sensor value
Orientation	3×1 vector	Orientation value of cell phone
Acceleration	3×1 vector	Accelerometer value of cell phone

Table 1. Data structure recorded by self-developed driving record data collection software

Figure 2. Schematic diagram of the interface between the collection device and the software.

3.3 Traffic Sign Detection and Position Calculation

Among the existing traffic sign detection methods, the TsingNet model (Liu et al., 2021) has achieved better results in various types of traffic sign datasets. By constructing an attention-driven bilateral feature pyramid network, this model has a strong recognition effect on small-scale and occluded traffic signs in outdoor environments. However, since the traffic sign enclosing frame extracted by the TsingNet model is drawn based on the hotspot region of the detection results, the enclosing frame has a poor fit with the traffic sign, which has an impact on the subsequent calculation of the traffic sign position.

In order to solve the above problems, this study is based on the two models of YOLOv5 (Jocher et al., 2020) target detection and classification, which deal with the detection process and the classification process respectively. The model is trained with the TT100K dataset, and the detection model is used to detect and coarsely classify the traffic signs in the driving record images, to obtain their image positions and the broad categories (prohibitions, indications, warnings) to which the traffic signs belong, and then to subdivide the detection results.

In addition, the detection results of the monocular camera on the same target at different positions are used to obtain the matching relationship of traffic signs between consecutive frames, and then each camera is used as a measuring station to calculate the spatial position of traffic signs using the Structure from Motion (SfM) and Bundle Adjustment (BA) technique.

In order to evaluate the accuracy of the extraction results, this study uses the confidence level to evaluate the semantic information and position information of the extraction results. The confidence level of a traffic sign extraction result consists of semantic confidence level and position confidence level, as shown in Equation (1).

$$P_{confidence} = [P_{semantic} \quad P_{coordinate}] \qquad (1)$$

$P_{semantic}$ is represented by the average of the classification correctness of the method. $P_{coordinate}$ is generated by the combination of the coordinate error of the camera position itself and the error of the calculation technique.

3.4 Traffic Sign Filtering

Traffic sign detection results usually contain semantic classification errors or position calculation errors, so it is necessary to filter the traffic sign detection results before the fusion of HD map crowd-source data. In this study, the semantic confidence and position confidence of traffic signs are used to set different confidence thresholds, and the detection results that do not meet the threshold requirements are excluded. Set the semantic confidence threshold $T_{semantic}$, the positional confidence thresholds $T^{x}_{coordinate}$ and $T^{y}_{coordinate}$, and the relative error of the camera position fitting threshold $T^{\delta}_{coordinate}$. Assuming that the collection of traffic signs before selection is TS_{before}, then the collection of traffic signs after selection is TS_{after}:

$$TS_{after} = \{TS_{before} | (TS_{semantic} \geq T_{semantic}) \wedge (TS_{xcoordinate} \leq T^{x}_{coordinate}) \wedge (TS_{ycoordinate} \leq T^{y}_{coordinate}) \wedge (TS_{\delta coordinate} \leq T^{\delta}_{coordinate})\} \qquad (2)$$

3.5 Traffic Sign Matching

Traffic sign matching plays a key role in the determination of existing signs and the incremental position fusion of new signs. In the determination process of existing signs, the traffic sign detection results are matched with the HD map, and the matching target is searched for in the traffic sign layer of the HD map to determine its corresponding truth value, and if there is no corresponding match, it is determined as a newly added sign; in the position fusion process of the newly added sign, it is necessary to match the extracted results with the existing temporary results in the change layer to carry out the position fusion.

For each traffic sign detection result, a buffer with radius R_{search} is constructed to select the traffic signs in the true value layer of the HD map and obtain candidate traffic signs. The buffer radius R_{search} is related to the accuracy of the traffic sign detection result from its matching target, which can be evaluated by labeling samples. In order to evaluate the degree of matching between the extraction results and the candidate signs, this study proposes a variety of indicators to evaluate the degree of matching, and the indicators are calculated as follows.

3.5.1 Visibility Indicator: Signs with their backs to the direction of vehicle travel cannot be observed; at the same time, in the overlapping area between elevated and ground level, traffic signs selected through the planar buffer may be located in a different road from the current road of the vehicle travelled on, so there is a visibility problem for locating traffic signs. The visibility index $P_{visibility}$ is calculated as follows:

$$P_{visibility} = cos(\theta) \qquad (3)$$

Where θ is the angle between the observation direction and the traffic sign facing direction.

3.5.2 Classification Similarity Indicator: Due to the wide variety of traffic sign targets, it is difficult for the traffic sign detection and classification model to cover all classifications, and traffic sign classification is prone to mistakes. Some traffic sign classifications with sub-semantics usually change their sub-semantics content with the change of road environment, for example, the speed value in various types of speed limit signs is usually changed. In order to match the traffic sign detection results with their corresponding HD map layers as much as possible, and to refine the impact of the error caused by the misclassification of traffic signs, this study summarizes the correspondence between the classifications of the traffic sign detection results and the classifications of the candidate signs in different cases: (1) the classifications are identical; (2) the classifications belong to "other classifications"; (3) the classifications are consistent but the sub-semantics are changed; (4) the classifications are changed but still belong to the same sign category; and (5) the classifications are completely changed. For the results of category comparison in different cases, this study set up different category similarity indicators to differentiate. Use P_{type} to denote the category similarity indicator.

3.5.3 Matching Result Calculation: In the set of candidate traffic signs, the matching indexes between the detection results and each candidate sign are calculated separately, and the summed matching probability of each candidate traffic sign is calculated by combining the results of semantic confidence and positional confidence according to the use of joint probability, as in Equation (4).

$$P_{match} = P_{semantic} * P_{visibility} * P_{type} * (1 - \delta) \qquad (4)$$

Since there may be new traffic signs, it is not always possible to obtain the corresponding traffic sign for each detection result, so it is necessary to define the matching threshold T_{match}. when $P_{match} \geq T_{match}$, then the current detection result is considered to match this candidate result and the judgment of semantic change is carried out; otherwise, the extracted result will be treated as a new sign and the crowd-source fusion will be carried out.

3.6 Traffic Sign Addition

Newly added signs lack corresponding records in the traffic sign layer of the HD map, so they usually do not obtain matching results in the matching process with the traffic sign layer of the HD map. In the process of fusion of new signs, a new sign may be detected in multiple driving record trajectories. Therefore, for a single new sign detection result, the result needs to be fused with the existing new signs in the change layer. The traffic sign matching process is repeated again in the traffic sign change layer to obtain the corresponding traffic sign. If there is a matching

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-641-2025 | © Author(s) 2025. CC BY 4.0 License.

result, the traffic sign detection result is merged with the matching sign in the change layer, the coordinate position of the fused sign is the weighted average of the multiple detection results of the sign, and the semantics of the fused sign is the type of traffic sign that occurs with the highest frequency in the multiple classification results; if there is still no matching result, the new sign can be added in the change layer directly.
In addition, there is a difference between the detected traffic signs and the real traffic signs. In the same consecutive frame, the offsets of the detected traffic signs are usually the same, so the matching result of the unchanged sign can be used to correct the offset of the added sign position to obtain a more accurate position.

4. Experimental Results

4.1 Study Area and Data

In this study, an experimental area is set up in the northeastern part of Shanghai city to utilize the measured traffic record data for traffic sign fusion updating. The experimental area contains various types of roads and intersections in the urban road environment, so that the effect of the proposed method can be fully investigated. The specific scope of the experimental area is shown in Fig. 4.
The HD map traffic sign data used in the experiment is provided by the Shanghai Municipal Institute of Surveying and Mapping (updated in July 2022), and is supplemented by high-resolution DSMs collected and generated by DJI's Jingwei M300 RTK professional drone. There are a total of 91 traffic signs in 32 categories in the HD map, and the distribution is shown in Figure 3.

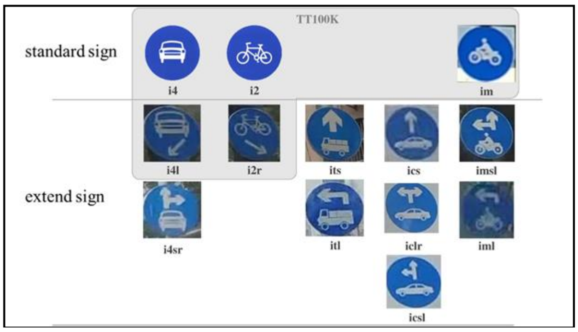

Figure 3. Examples of different traffic signs.

The driving record data set contains 15 trajectories of driving record data, total 79.858km, and the repeated observation area is concentrated in the trajectory of The North Zhongshan Road Number Two and the trajectory of The Huangxing Road. The trajectory collection area and trajectory distribution are shown in Figure 4, and other information of each trajectory is shown in Table 2.

ID	Date	Time	Mileage (km)	Device number
1	2023/9/11	11:55:28-12:02:58	4.209	1
2	2023/9/12	11:37:30-11:41:43	3.786	1
3	2023/9/12	11:47:43-11:53:01	3.108	1
4	2023/9/12	11:55:59-12:07:59	3.745	1
5	2024/1/30	08:13:19-08:54:09	11.788	1
6	2024/1/30	08:55:00-09:02:07	1.936	1
7	2024/1/30	09:32:39-09:49:34	6.461	1
8	2024/1/30	09:04:10-09:19:06	4.521	1
9	2024/1/30	09:57:29-10:24:35	3.699	1
10	2024/1/30	08:13:17-08:44:00	8.116	2
11	2024/1/30	08:44:00-08:55:07	3.699	2
12	2024/1/30	08:55:08-09:13:44	5.359	2
13	2024/1/30	09:13:48-09:33:36	9.606	2
14	2024/1/30	09:33:44-09:49:32	6.386	2
15	2024/1/30	09:58:08-10:17:09	3.439	2

Table 2. Details of traffic sign position calculation and crowd-source fusion updating experimental traffic record data

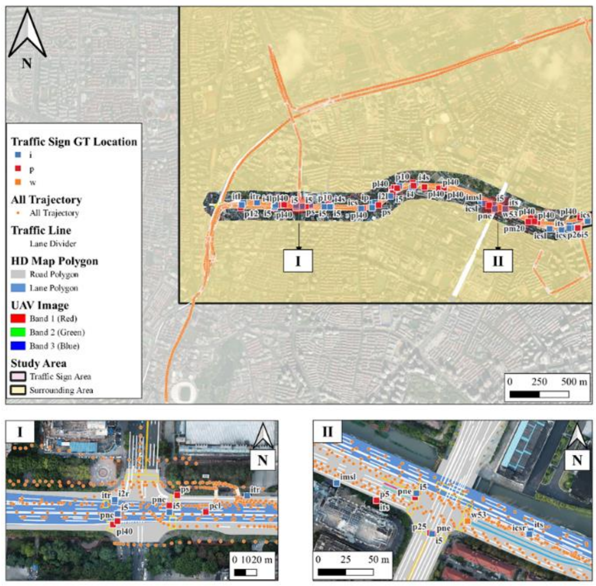

Figure 4. Description of the experimental area.

4.2 Experiment Setting and Accuracy Evaluation

4.2.1 Experiment Setting and Data Processing: In the set of candidate traffic signs, the matching indexes between the detection results and each candidate sign are calculated separately, and the summed matching probability of each candidate traffic sign is calculated by combining the results of semantic confidence and positional confidence according to the use of joint probability, as in Equation (4).
In section 3.4, the position confidence thresholds $T^{x}_{coordinate}$ and $T^{y}_{coordinate}$ in the X direction and Y direction are both set to 5 meters, the relative error threshold $T^{\delta}_{coordinate}$ for camera position fitting is 20%.
According to the category correspondence between the traffic sign classification results and the candidate signs, the category similarity indicator P_{type} is set as follows:

(1) The categories are identical: $P_{type} = 1$;
(2) Category belongs to "other category": $P_{type} = 0.75$;
(3) The categories are consistent but the sub-semantics are changed: $P_{type} = 0.6$, and there are only speed limit signs (pl*), weight limit signs (pm*), height limit signs (ph*), minimum speed indicator signs (il*), motorcycle indicator signs (im*), bicycle indicator signs (i2*), motor vehicle indicator signs (i4*), width restriction signs (pw*), and signs for lifting the maximum speed limit (pr*). pr*) 9 major categories of signs contain sub-semantics;
(4) Category changes but still belongs to the same major sign category: $P_{type} = 0.1$;
(5) Category changes completely: $P_{type} = 0$.

The parameters used in traffic sign matching and addition will be discussed and illustrated in Section 4.3.

4.2.2 Accuracy Evaluation Metrics: For each traffic sign position calculation result, the error E_{sign} is calculated as Equation (5). The accuracy of the traffic sign position calculation method is evaluated using the mean error ($\mu_{E_{sign}}$), and the standard deviation error ($\sigma_{E_{sign}}$) of all sign position calculation results.

$$E_{sign} = distance\left(p_{sign}^{HD\ map}, p_{sign}\right) \tag{5}$$

Where the function $distance(p_1, p_2)$ denotes the straight line distance between point p_1 and point p_2, p_{sign} denotes the coordinates of the result of the traffic sign position calculation, and $p_{sign}^{HD\ map}$ denotes the coordinates of the corresponding HD map traffic sign.

The experiments conduct matching tests on different search radius R_{search} and matching thresholds T_{match}, using the loss rate ($Rate_{lost}$) and matching error rate ($Rate_{mismatch}$) of the matching results of each detected target as metrics, and the metrics are calculated as in Equation (6) and Equation (7).

$$Rate_{lost} = \frac{Num_{lost}}{Num_{all}} \times 100\% \tag{6}$$

$$Rate_{mismatch} = \frac{Num_{mismatch}}{Num_{all}} \times 100\% \tag{7}$$

Where Num_{lost} denotes the number of no matching results, Num_{all} denotes the number of all extracted results, and $Num_{mismatch}$ denotes the number of incorrect matching results.

4.3 Results and Analysis

4.3.1 Traffic Sign Detection and Position Calculation Results: Figure 5 demonstrates the results of the TsingNet and YOLO methods for sign detection in the two experimental regions. Compared with the TsingNet model, the detection and classification of targets are performed separately in the YOLO method, which supports more classification categories, more sensitive target detection results, and better overall average classification accuracy and recall of the model. In some of the images in Region I, the TsingNet model has difficulty in directly detecting traffic signs that are far away, and fewer valid observations can be obtained.

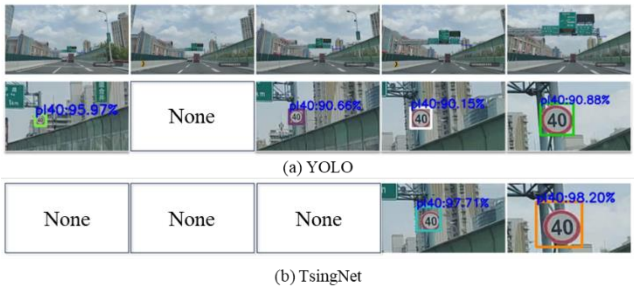

Figure 5. Sign detection results of region one.

In order to evaluate the accuracy of the calculation results of traffic sign centroids, the manually label center coordinates of the encircling frame are used as the true value of the traffic sign center position, and the difference between the center coordinates obtained by the traffic sign center fitting calculation method and the center coordinates of the encircling frame relative to the true value of the center position are calculated separately, and the results are shown in Table 3.

It can be found that compared with TsingNet, the difference between the center coordinates of the extracted frame of the traffic sign acquired by the YOLO method relative to the true value of the center position of the traffic sign is smaller, indicating that the encircling frame of the traffic sign acquired by the YOLO method is more complete and more accurate.

Model	Method	Δx **(pixel)**	Δy **(pixel)**	X relative error	Y relative error
Tsing Net	Fitting method	2.789	2.882	4.326%	4.325%
	Center	4.248	4.974	8.969%	8.828%
YOLO	Fitting method	2.144	2.181	4.135%	4.191%
	Center	0.879	0.924	2.751%	2.667%

Table 3. Comparison results of different traffic sign center position calculation methods and label truth values

4.3.2 Traffic Sign Matching Results: In order to reasonably compare the matching error rate, this experiment was conducted on the search radius of 0 to 50 m interval, matching threshold between 0 -0.5, the experimental results are shown in Figure 6.

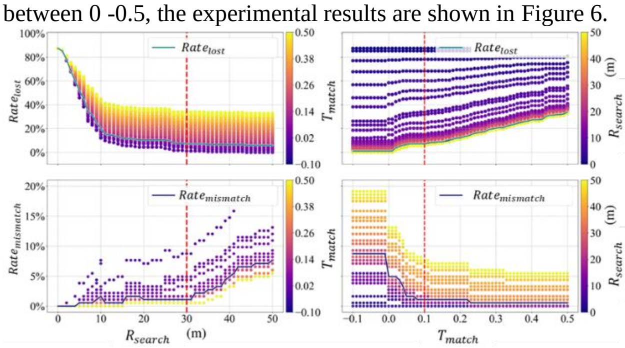

Figure 6. Match loss rate and match error rate with search radius and matching threshold.

From the above figure, it can be seen that as the loss rate decreases, the number of matched flags gradually increases, resulting in an increasing error rate. As the matching threshold rises, some of the candidate flags are discarded because they do not meet the matching threshold requirement even though they

have obtained the highest matching metrics results, resulting in an increasing loss rate but decreasing error rate. Combining the results of loss rate and error rate, this study uses a search radius of 30 meters and a matching threshold of 0.1 as the parameter values in the matching process.
For the traffic sign detection results in the experimental area, after excluding the new signs, there are 13 lost match results, accounting for 7.10% of the number of all extracted results; and 2 incorrect match results, accounting for 1.18% of the number of all extracted results. In summary, the matching effect is better.

4.3.3 Traffic Sign Addition results: Some traffic signs are deleted from the HD map layer, and the detection results corresponding to signs with true values in the experimental area are used as inputs to simulate the experimental results in the case of added signs. As the search radius gradually increases, the recall rate gradually increases. The search radius of the change layer is set to 40 m due to the effects of search radius. The recall rate of traffic signs in the experimental area is 58.24%, and the precision rate is 88.33%. Figure 7 gives the schematic diagram of a traffic sign change.

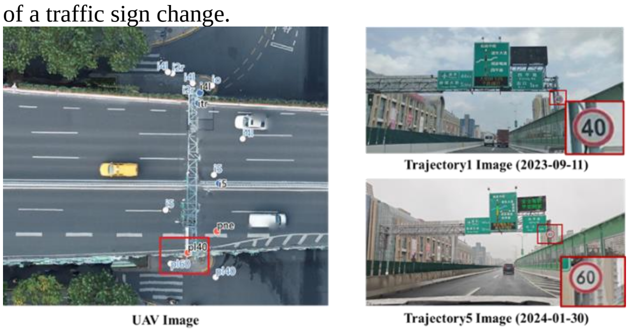

Figure 7. Schematic diagram of changed sign.

When the same sign is repeatedly detected by multiple trajectories, the added method will perform position fusion on the multiple detection results, and this study also uses the existing signs to optimize the coordinates of the added results. In order to verify the optimization effect, half of the traffic sign data is randomly deleted from the HD map traffic sign data, and the other half of the data is retained to optimize the detection results. The detection results with HD map traffic sign annotations are used as input to verify the optimization effect of the added signs under the condition of having HD map matching results. The accuracy comparison is shown in Table 4.

Method	Mean Error (m)	Standard Deviation Error (m)
Correlation original data	7.50	4.86
After fusion (without optimization)	5.66	4.08
After fusion (with optimization)	**3.69**	**3.29**

Table 4. Coordinate accuracy results of fusion and optimization

With the support of coordinate fusion and optimization, the mean value of the fused traffic sign error is 3.69 m, the standard deviation is 3.29 m, and the median is 2.80 m, which basically achieves similar computational results to those using more accurate GNSS positioning equipment, indicating that it is important to optimize the detection results by using the results of matching the traffic sign layers of the HD map.

5. Conclusion

In this paper, based on the driving record data, the position calculation of traffic signs in the road environment and the crowd-source updating method of traffic sign layer in the HD map are investigated. In this paper, an incremental HD map traffic sign crowd-source update method is proposed. The traffic sign detection results are matched with the existing traffic signs in the HD map for traffic sign change detection; a position optimization and fusion method for the added results is proposed to optimize the position of the added signs by using the unchanged signs, and the multiple extraction results are fused to obtain the optimized position of the added traffic signs.
Experiments show that the traffic sign position accuracy after crowd-source fusion and optimization is improved obviously, the average plane error is 3.69 m, and the standard deviation of error is 3.29 m, which can provide data support for the crowd-source update of HD map.

Acknowledgements

This work was supported by the National Key Research and Development Program of China [2021YFB2501103].
We thank the Shanghai Municipal Institute of Surveying and Mapping for providing the HD map data. Further, we thank Yiren Technology Co. (Shanghai, China) for providing the experimental dataset. We gratefully acknowledge the comments from the editor and the reviewers.

References

Alcantarilla, P.F., Simon, S., Germán, R., Roberto, A., Riccardo, G., 2018. Street-view change detection with deconvolutional networks. *Autonomous Robots*, 42, 1–22.

Das, A., Ijsselmuiden, J., Dubbelman, G., 2020. Pose-graph based Crowdsourced Mapping Framework. *2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS)*, Victoria, BC, Canada, 2020, 1-7

Jo, K., Kim, C., Sunwoo, M., 2018. Simultaneous Localization and Map Change Update for the High-Definition Map Based Autonomous Driving Car. *Sensors*, 18(9): 3415.

Jocher, G., and Stoken, A., 2020. ultralytics/yolov5: v6.0. https://doi.org/10.5281/zenodo.4154370.

Kim, C., Cho, S., Sunwoo, M., Jo, K., 2018. Crowd-Sourced Mapping of New Feature Layer for High-Definition Map. *Sensors*, 18(12): 4172.

Kim, C., Cho, S., Sunwoo, M., Resende, P., Jo, K., 2021. Updating Point Cloud Layer of High Definition (HD) Map Based on Crowd-Sourcing of Multiple Vehicles Installed LiDAR. *IEEE Access*, 9, 8028–8046.

Kim, C., Jo, K., Bradai, B., Sunwoo, M., n.d. Multiple vehicles based new landmark feature mapping for highly autonomous driving map. *2017 14th Workshop on Positioning, Navigation and Communications (WPNC)*, Bremen, Germany, 2017, 1-6.

Lai, C., Zhang, M., Zheng, D., 2019. A Secure and Efficient Map Update Scheme for Autonomous Vehicles. *Journal of Computer Research and Development*, 56, 2277.

Li, Y., 2022. Crowdsourcing map update technology for autonomous driving. *Beijing Surveying and Mapping*, 36.

Liu, Y., Peng, J., Xue, J., Chen, Y., Fu, Z., 2021. TSingNet: Scale-aware and Context-rich Feature Learning for Traffic Sign Detection and Recognition in the Wild. *Neurocomputing*, 447(4).

Park, Y.-K., Park, H., Woo, Y.-S., Choi, I.-G., Han, S.-S., 2022. Traffic Landmark Matching Framework for HD-Map Update: Dataset Training Case Study. *Electronics*, 11(6): 863.

Qin, Z., Li, Y., Liu, Y., 2020. Preliminary exploration of compliance issues in high-precision map crowdsourcing updates. *Automobile and Parts*, 3.

Tchuente, D., Senninger, D., Pietsch, H., Gasdzik, D., 2021. Providing more regular road signs infrastructure updates for connected driving: A crowdsourced approach with clustering and confidence level. *Decision Support Systems*, 141: 113443.

Yang, D., Li, Q., Peng, W., 2019. Several Issues Concerning the Development of Autonomous Driving Maps and Positioning Industry. China GNSS and LBs the 8th Annual Conference & China BDS Applications Forum.

Yang, D., 2018. Many directions that smart cars need to break through. *China industry & information technology*, 2, 21-27.

Yang, M., Jiang, K., Wen, T., Chen, H., Huang, J., Zhang, H., Huang, J., Tang, X., Yang, D., 2023. Review on Status and Challenges of Crowdsourced Updating of Highly Automated Driving Maps. *China Journal of Highway and Transport*, 36, 244-259.

Zhang, P., Zhang, M., Liu, J., 2021. Real-Time HD Map Change Detection for Crowdsourcing Update Based on Mid-to-High-End Sensors. *Sensors*, 21, 2477.

Zohar, T., Ariav, I., Bar-Zohar, M., 2012. Robust and efficient change detection algorithm based on 3D line segments, *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, 2012, 1-5.