

Article

Automatic Positioning of Street Objects Based on Self-Adaptive Constrained Line of Bearing from Street-View Images

Guannan Li ^{1,2,3}, Xiu Lu ^{1,2,3}, Bingxian Lin ^{1,2,3}, Liangchen Zhou ^{1,2,3,*} and Guonian Lv ^{1,2,3}

¹ Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China; 191301020@njnu.edu.cn (G.L.); 181301018@njnu.edu.cn (X.L.); linbx@njnu.edu.cn (B.L.); gnlu@njnu.edu.cn (G.L.)

² State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: zhoulich@njnu.edu.cn

Abstract: In order to realize the management of various street objects in smart cities and smart transportation, it is very important to determine their geolocation. Current positioning methods of street-view images based on mobile mapping systems (MMSs) mainly rely on depth data or image feature matching. However, auxiliary data increase the cost of data acquisition, and image features are difficult to apply to MMS data with low overlap. A positioning method based on threshold-constrained line of bearing (LOB) overcomes the above problems, but threshold selection depends on specific data and scenes and is not universal. In this paper, we propose the idea of divide-conquer based on the positioning method of LOB. The area to be calculated is adaptively divided by the driving trajectory of the MMS, which constrains the effective range of LOB and reduces the unnecessary calculation cost. This method achieves reasonable screening of the positioning results within range without introducing other auxiliary data, which improves the computing efficiency and the geographic positioning accuracy. Yincun town, Changzhou City, China, was used as the experimental area, and pole-like objects were used as research objects to test the proposed method. The results show that the 6104 pole-like objects obtained through object detection realized by deep learning are mapped as LOBs, and high-precision geographic positioning of pole-like objects is realized through region division and self-adaptive constraints (recall rate, 93%; accuracy rate, 96%). Compared with the existing positioning methods based on LOB, the positioning accuracy of the proposed method is higher, and the threshold value is self-adaptive to various road scenes.

Keywords: self-adaptive constraint; line of bearing; geolocation; pole-like objects; street-view images; object detection



Citation: Li, G.; Lu, X.; Lin, B.; Zhou, L.; Lv, G. Automatic Positioning of Street Objects Based on Self-Adaptive Constrained Line of Bearing from Street-View Images. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 253. <https://doi.org/10.3390/ijgi11040253>

Academic Editor: Wolfgang Kainz

Received: 19 February 2022

Accepted: 7 April 2022

Published: 12 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Roads are important parts of cities. Street objects on both sides of the road are an important part urban infrastructure management, intelligent transportation system construction, and unmanned high-precision maps [1,2]. Achieving fast and accurate collection of street objects has become an important task for the digital construction of cities and traffic, as well as the realization of automatic driving. The geolocation and attributes of street objects are important indicators for collection, especially location information, which is the basis of the street object and one of the most important factors. The location of street objects can assist in road asset management [3] and calculation of the safety risk index to evaluate road safety [4]. The presence of appropriate street objects in the right places can effectively reduce traffic risks, for example, by placing easy-to-read guide signs on curves [5]. Therefore, it is very important to perform effective geolocation and attribute acquisition of street objects.

In order to obtain geolocation and attribute information of the street objects on both sides of a road, an effective data collection and extraction method is required. Current data collection methods include manual field measurement, aerial remote sensing images, and mobile mapping systems (MMS). Manual field measurement requires a large number of professionals to conduct external measurements, and the labor cost is relatively high. Aerial remote sensing images observe the road surface from a top-down perspective, which can better collect large-area objects, such as road markings [6]. However, the orthographic projection area of narrow and vertical objects on both sides of the road is small and is therefore difficult to capture, lacks local details [7], and is affected by high-rise buildings and trees on both sides of the road. Regarding MMS, which can observe street objects on both sides of the road from a side view, the observation results are more in line with the visual psychology of people observing the objects and are more easily extracted [8]. MMS-based measurement relies on the system being equipped with a high-precision global navigation satellite system (GNSS) to realize positioning, a high-frequency inertial measurement unit (IMU) to realize attitude determination, a high-resolution camera to realize street-view image shooting, and a high-speed laser to realize distance measurement of street objects [9]. As a measurement device, LIDAR can extract the precise location of street objects from point cloud data obtained by scanning, although the cost is high [10,11]. Compared with images, point cloud is more challenging in semantic segmentation, especially for complex scenes where the technology is immature. Therefore, based on a camera mounted on the system, with the help of current, relatively mature image semantic analysis [12] and object-detection methods [13], the localization of street objects based on multi-view street-view images has become a lower-cost alternative [14,15].

The traditional object-positioning method based on multi-view street-view images acquired by MMS relies on visual matching. According to the image features, corresponding points are matched, and positioning is realized by geometric constraints generated by the corresponding points. Chang et al. demonstrated the feasibility of object positioning based on multi-view street-view images by manually matching corresponding points [16]. Nassar takes the camera location, camera spacing, and the heading angle of the target obtained by the MMS system as input parameters, applies geometric soft constraints to the Siamese convolutional neural network, and relies on the matching objects in multiple views to achieve triangulation positioning [17]. Ogawa proposed the use of a map and the location of buildings in images to correct the camera-position parameters of the captured image, thereby improving the object-recognition accuracy and geographic positioning accuracy of the image [18]. Zhu et al. performed street-to-aerial image matching based on an improved Siamese convolutional neural network to estimate the geolocation and orientation of targets in the street-view images [19]. When trying to automatically match corresponding points, due to the similarity of objects and backgrounds in multi-view images, it is difficult to automatically distinguish similar objects in the same background [20]. Multi-view images acquired by MMS have fewer overlapping areas, and it is difficult to achieve satisfactory results in visual matching using image object key points or descriptors [21,22]. The objects in images acquired by MMS have similar visual features of the same type and different instances and different visual features of the same instance and different perspectives, which makes multi-view visual matching difficult.

In order to resolve the difficulty of visual matching of multi-view images acquired by MMS, scholars have tried to transform the visual matching positioning problem into a passive positioning problem [23–25]. Firstly, the object of interest is detected from the multi-view images, and then the orientation of the object of interest relative to the shooting position is calculated according to the pose parameters. The orientation is then represented using line of bearing (LOB), and finally, the possible location of the object of interest is calculated by LOB intersection. Chu et al. proposed a deep-learning-based method for determining the orientation of objects in images [26]. This method provides a new form of LOB orientation acquisition and correction, as well as improvement in the positioning accuracy based on the LOB method. Hazellhoff et al. used the intersection and

aggregation center of object LOBs in multiple views as the positioning result, although the result may contain a large number of ghost nodes [27,28]. Based on the Markov random field optimization method, Krylov et al. introduced the distance data of objects into decision making to reduce ghost nodes, although the object-positioning results often had considerable randomness [21]. Nassar integrates object detection and depth estimation into an end-to-end graph neural network and relies on the estimated depth information to determine the operating distance of the LOB generated by the object-detection result to achieve geographic positioning [20]. Similarly, Lumnitz et al. applied a monocular depth-estimation algorithm and triangulation to Google Street View and Mapillary and gathered adjacent LOB intersections into a cluster to realize meter-level geographic positioning of urban trees [29]. These methods all introduce depth information, the accuracy of which the positioning accuracy depends on. Zhang proposed a method of modified brute-force-based LOB measurement, which reduces the influence of ghost nodes on the positioning results of utility poles without introducing other data and can obtain stable positioning results [30]. Khan realized the geographic positioning of eucalyptus trees on both sides of a road based on the modified brute-force-based LOB measurement, which verified the feasibility of this method once again [31]. However, most of the current LOB-based positioning methods require a large number of threshold constraints, and it is difficult to select an appropriate threshold value. For example, the modified brute-force-based LOB measurement needs to consider the width of the road and other factors, which is difficult to apply to object extraction on both sides of a road with a wide range and different widths.

Aiming at the applicability of threshold selection for different road scenes based on LOB object positioning, in this paper, we propose an automatic positioning method for street objects based on LOB with adaptive constraints. This method can automatically divide the calculation area into grids based on the calibrated effective collection distance and driving trajectory, which constrains the effective range of LOB intersection calculation. According to the relationship between the LOB and the intersection point, two constrained rules independent of the threshold are proposed to further eliminate the ghost nodes generated by LOB intersection. The reserved calculation results are automatic positioning results of street objects. The algorithm presented in this paper is not affected by factors such as road width, and it is suitable for different road scenes. The adaptive selection of the LOB constraint threshold is realized by the effective shooting distance of the MMS images and the driving trajectory, which has universal applicability and generalizability.

The remainder of this paper is arranged as follows. Section 2 introduces the implementation of the proposed method and expounds the basic principle. Section 3 introduces Yincun town, Changzhou City, Jiangsu Province, China, as the experimental area, with street-view images of the area collected by MMS as the experimental data, and we introduce the experimental process and results in detail. Then, we compare and analyze the automatic positioning method of street objects based on the self-adaptive constraint LOB proposed in this paper and the modified brute-force-based LOB positioning method. Different thresholds are illustrated and discussed. Finally, the full text is summarized.

2. Methodology

In this paper, we propose an automatic positioning method for objects in street-view images with self-adaptive constraint LOB, which mainly includes two parts: LOB mapping based on object-detection results and LOB-based geographic positioning. The specific process is shown in Figure 1. LOB mapping based on object-detection results includes object detection marked by a bounding box and simulation of line of sight based on LOBs. LOB-based geographic positioning includes grid division based on driving trajectory, acquisition of intersection points based on LOBs, and elimination of ghost nodes by constrained rules. In the LOB-based geographic positioning process, the grid division based on driving trajectory proposed in this paper improves computational efficiency. It also provides self-adaptive constraints for eliminating ghost nodes. It works with other constraints to improve the positioning accuracy of street objects.

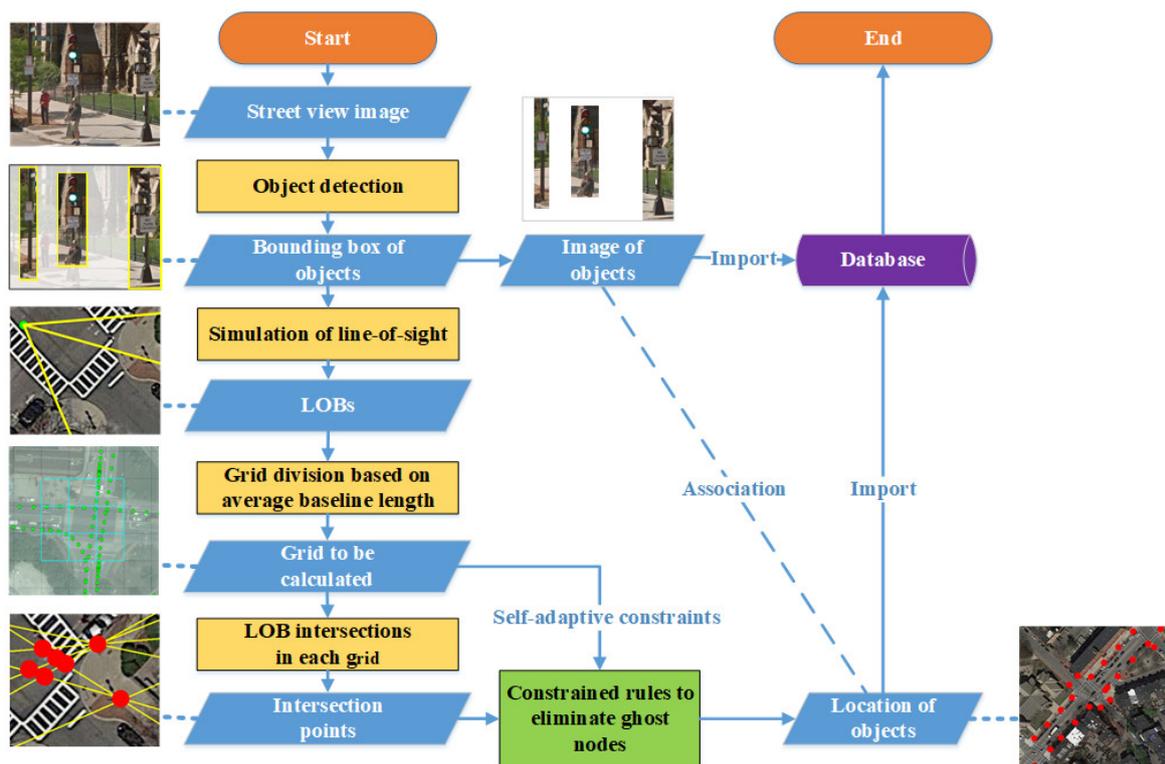


Figure 1. Flow chart of this study.

2.1. LOB Mapping Based on Object-Detection Results

In order to realize image-based object positioning, it is necessary to detect the object in the image. There are relatively mature image object-detection algorithms that have achieved good detection results in street object detection [11,32,33]. In this paper, a cascade region convolutional neural network (cascade R-CNN) is used to realize object detection. As shown in Figure 2, the model extracts the features of the input image through backbone convolutions. It uses the region proposal network (RPN) to obtain a series of rough rectangular proposals of the object. A series of end-to-end subdetectors is cascaded, and the bounding boxes output by the previous stage detector are input into the latter stage detector. This gradually increases the threshold of intersection over union (IoU) between the candidate bound and ground-truth bound to improve detection results [34]. Compared with other R-CNN series models, cascade R-CNN introduces detection sub-networks with different frameworks. It overcomes the overfitting of regression at a specific IoU threshold and can achieve relatively good recognition accuracy [34]. The objects to be detected are marked in the images with bounding boxes, and cascade R-CNN is used for training and learning so that the same type of objects in other images can be automatically marked with bounding boxes.

As shown in Figure 3, the image space coordinates (x_c, y_c, z_c) of the center pixel of the box are obtained according to the bounding box of the recognized object in the image and the imaging law [8,35]. Combined with the spatial coordinates recorded by GNSS and the Eulerian angle recorded by IMU, the absolute location and attitude of the vehicle are obtained, the mapping relationship between the bounding and the observation orientation is calculated, and the LOB is constructed [30,31,36].

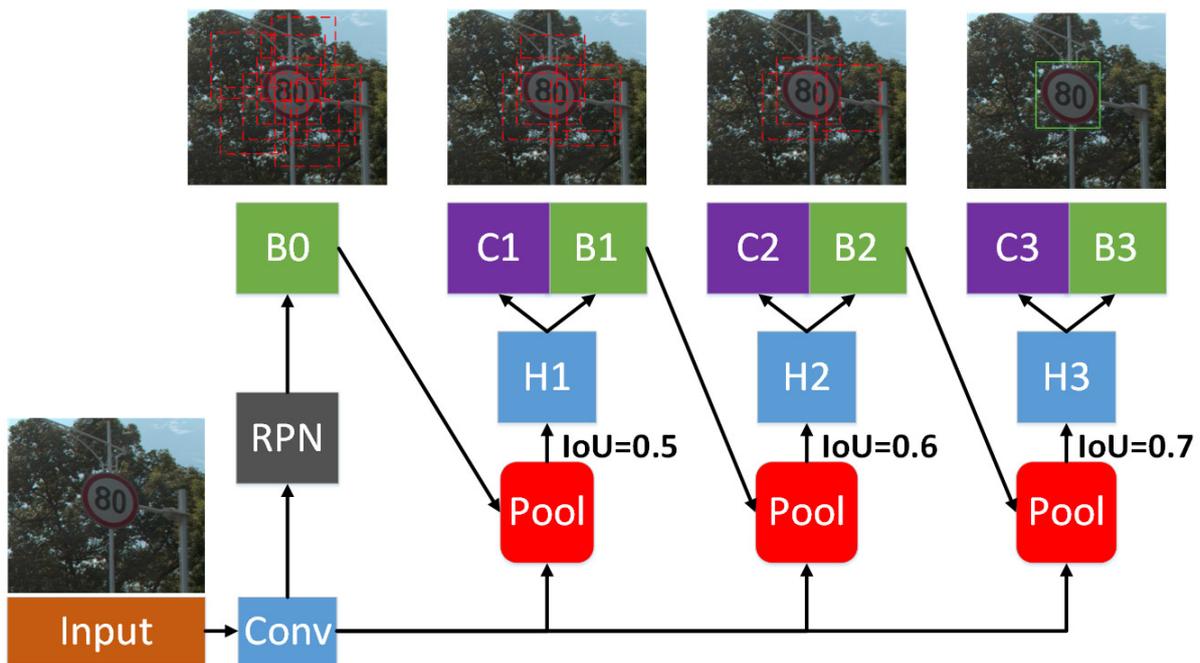


Figure 2. Cascade R-CNN schematic: “Input” represents input image, “Conv” represents backbone convolutions, “RPN” represents region proposal network, “Pool” represents region-wise feature extraction, “H” represents network head, “C” represents classification, and “B” represents bounding box.

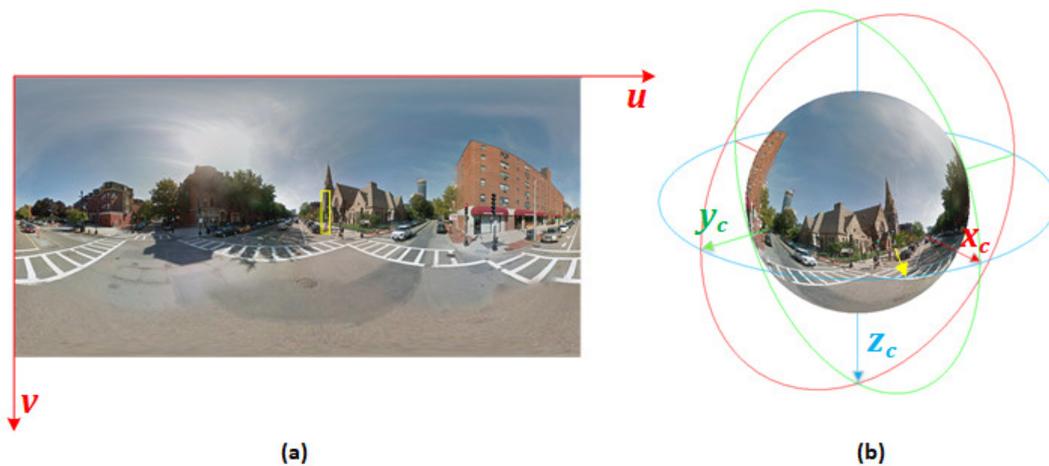


Figure 3. Acquisition of bounding box and LOB: (a) bounding box in street-view image and (b) LOB in the world coordinate.

The projection of the pixel in the world coordinate system (x_w, y_w, z_w) can be obtained through Equation (1):

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = sR \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \begin{bmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{bmatrix}, \tag{1}$$

where s represents the depth coefficient, which will be offset in subsequent calculations; R represents the rotation matrix from the image-space coordinate system estimated from the vehicle attitude parameters to the world coordinate system; and $(x_{cam}, y_{cam}, z_{cam})$ represents the camera world coordinates calculated from the GNSS parameters and calibration parameters.

The orientation, b , corresponding to the pixel can be expressed by Equation (2):

$$b = \arctan\left(\frac{y_w - y_{cam}}{x_w - x_{cam}}\right), \quad (2)$$

LOB is represented by l , as shown in Equation (3):

$$l = (x_{cam}, y_{cam}, z_{cam}, b), \quad (3)$$

2.2. Geographic Positioning Based on Self-Adaptive Constrained LOB

After obtaining the LOB of the detected object mapping in the street-view image, association matching of the detection object in the multi-view images is realized by the spatial aggregation of the LOB. However, a large number of false associations may be generated, forming ghost nodes. To reduce the influence of ghost nodes, the grid division range is first automatically calculated according to the driving trajectory, the effective distance of the LOB is limited, and the ghost nodes are preliminarily eliminated. Then, the intersection of LOBs in each grid is expressed by a relation matrix, and ghost nodes are further eliminated based on the proposed self-adaptive constrained rules. Next, the process of the LOB-based positioning method is described in detail.

2.2.1. LOB Measurement

As shown in Figure 4, when the same object is captured by multi-view images, the bounding box of the object is mapped to an LOB, which will generate a geometric intersection, that is, the LOB intersection. However, due to the existence of observation errors, these intersections often do not overlap completely in space but are aggregated into a cluster within a certain range. The centroid of the intersections within this cluster represents the geolocation of the street object.

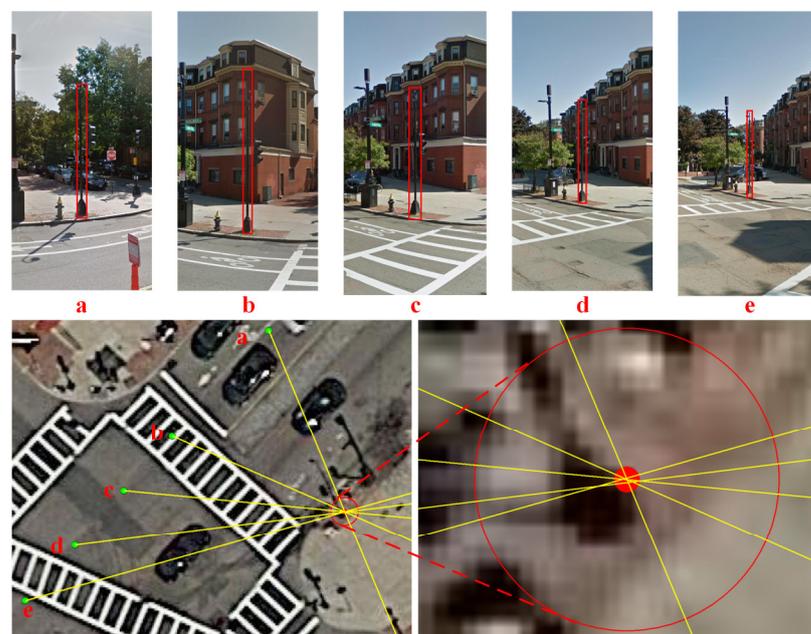


Figure 4. Schematic diagram of bounding box, LOB, and cluster: letters (a) to (e) represent different views.

As shown in Figure 5, the bounding boxes detected from different images in complex scenes are not all the same object. The intersection points can be generated between any two non-parallel LOBs, which includes the real geolocation of the identified object and also

contains a large number of ghost nodes. Therefore, certain constrained rules are required to eliminate ghost nodes.

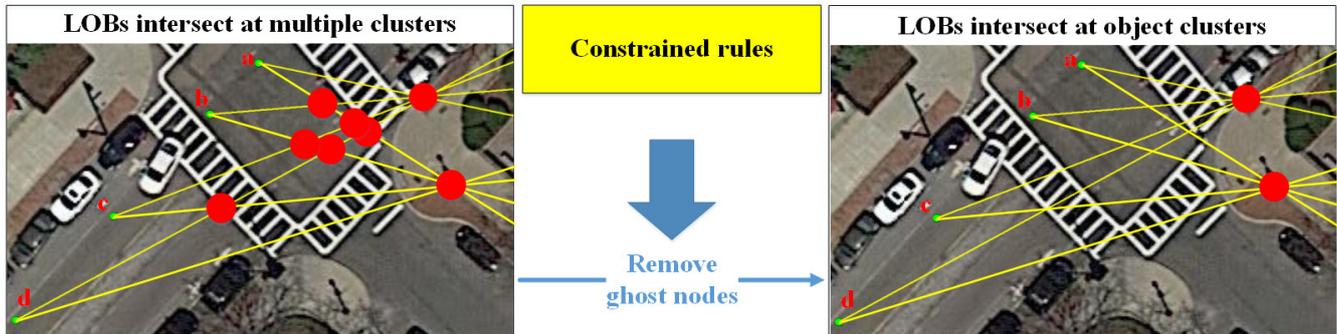


Figure 5. Schematic diagram showing reduction of the influence of ghost nodes on the positioning results based on constrained rules: letters a to d represent different views.

2.2.2. Division of Grid

Due to the large range of the driving trajectory, each trajectory point within the range can generate several LOBs, and only the LOB intersection generated by the trajectory points with a short distance can determine the geolocation of the object. The range of the driving trajectory is divided into grids, and LOB intersection calculation is only performed in the adjacent grid each time, which indirectly restricts the effective length of the LOB, potentially reducing unnecessary calculations and removing the ghost nodes outside the effective range of the LOB.

As shown in Figure 6, a driving trajectory including n records is recorded as $\{T_i(x_i, y_i, z_i) | i = 1, 2, \dots, n\}$, and the average baseline length of triangulations of adjacent views can be estimated by Equation (4), denoted by bl .

$$bl = \frac{1}{n} \sum_{i=1}^{n-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2}, \quad (4)$$

According to the two-dimensional range of the driving trajectory ($min_x, min_y, max_x, max_y$), it is evenly divided into several square grids and takes k times the distance of bl as the size of the unit grid. The number of columns, n_{col} , and the number of rows, n_{row} , can be obtained by Equation (5):

$$\begin{cases} n_{col} = \text{Ceil}\left(\frac{max_x - min_x}{k \times bl}\right) \\ n_{row} = \text{Ceil}\left(\frac{max_y - min_y}{k \times bl}\right) \end{cases}, \quad (5)$$

where Ceil represents the rounding-up function.

In each calculation process, the intersection points are calculated only for the LOB mapped by the captured images within the range of Equation (6), denoted by $\text{Grid}_{calculation}$, and only the geographic positioning results within the range of Equation (7) are recorded, denoted by $\text{Grid}_{recorded}$.

$$\text{Grid}_{calculation} = \begin{cases} x \in [k \times bl \times (col - 1) + min_x, k \times bl \times (col + 2) + min_x] \\ y \in [k \times bl \times (row - 1) + min_y, k \times bl \times (row + 2) + min_y] \end{cases}, \quad (6)$$

$$\text{Grid}_{recorded} = \begin{cases} x \in [k \times bl \times col + min_x, k \times bl \times (col + 1) + min_x] \\ y \in [k \times bl \times row + min_y, k \times bl \times (row + 1) + min_y] \end{cases}, \quad (7)$$

where col represents the column number of the current $\text{Grid}_{recorded}$, and the value is an integer within the range of $[0, n_{col})$; and row represents the row number of the current $\text{Grid}_{recorded}$, which is an integer within the range of $[0, n_{row})$.

During grid division, the threshold, k , is used to ensure that at least two trajectory points ($k > 1$) for mapping LOBs are included in the grid range to be calculated. The grid is a regular square. Because only the results within the central grid are recorded each time, the maximum effective intersection distance of the LOB is $k2\sqrt{2}bl$. However, objects with a long distance tend to have a small number of pixels in the image and poor positioning accuracy. Therefore, the effective shooting distance, V , of the equipment can be estimated according to the equipment conditions. As shown in Equation (8), according to $k2\sqrt{2}bl \leq V$, the value range of the threshold, k , is:

$$k \in \left(1, \frac{V}{2\sqrt{2}bl}\right], \tag{8}$$

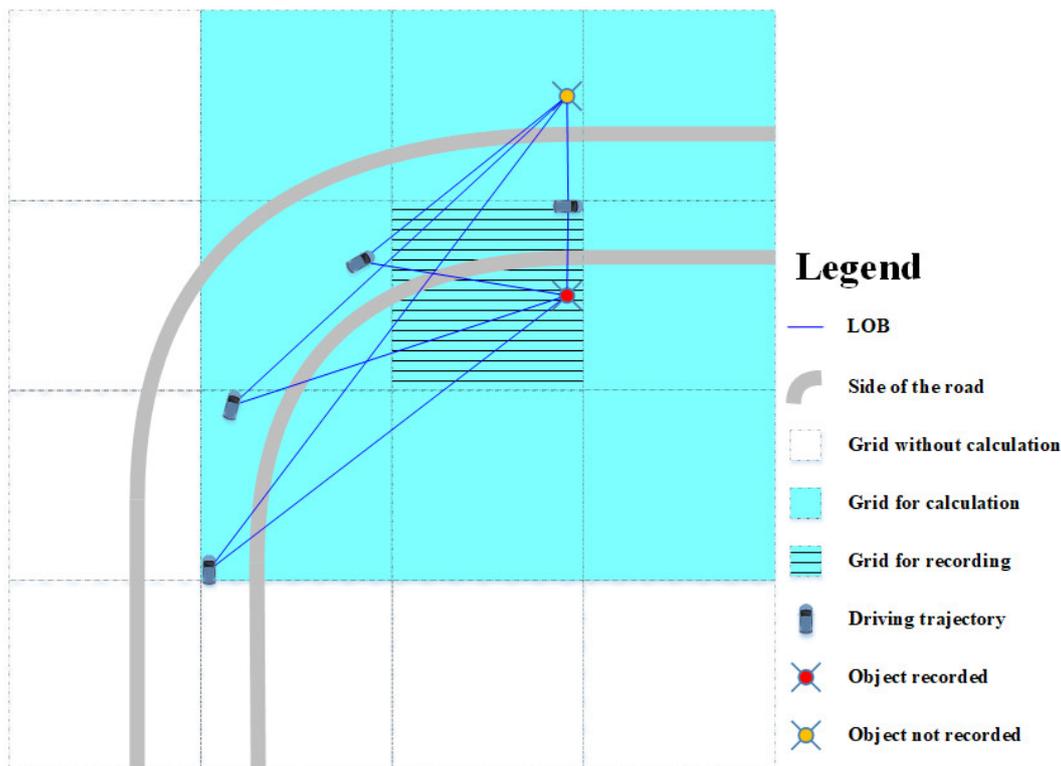


Figure 6. Schematic diagram of grid division and grid-by-grid calculation.

2.2.3. Relationship Matrix Construction

A set, L , is used to represent n LOBs, where the i_{th} LOB is represented by l_i :

$$L = \{l_1, l_2, l_3, \dots, l_i, \dots, l_j, \dots, l_n\}, \tag{9}$$

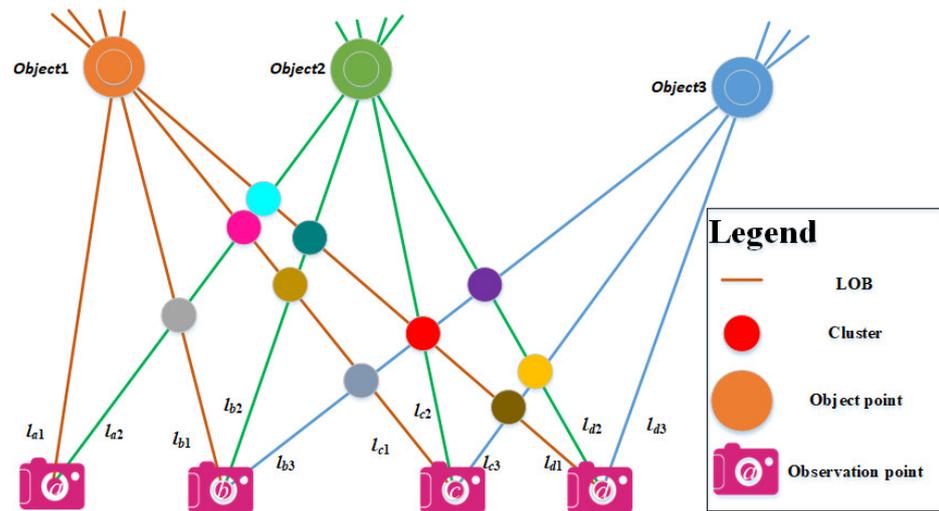
Then, the intersection point is generated by the intersection of LOBs in the set, L , which can be represented by an $n \times n$ intersection matrix:

$$M(P_{i,j}) = (p_{i \times j})_{n \times n}, \tag{10}$$

where $M(P_{i,j})$ represents the intersection matrix of LOBs in the set, L , and $p_{i \times j}$ represents the intersection of l_i and l_j .

As shown in Figure 7a, *Object1*, *Object2* and *Object3* are observed at four observation points: a, b, c, and d, respectively. In order to facilitate understanding, a combination of letters and numbers is used to record the observation position and object of the LOB. For example: l_{a1} is the LOB of *Object1* observed from position a. We use three different colored LOBs to simulate the line of sight when observing three different objects. The object

location is the aggregation location of the LOB intersection of the same color, and LOB intersections of different colors are ghost nodes. As shown in Figure 7b, the intersection matrix is used to describe the intersection relationship between LOBs. Because the matrix has a certain symmetry, it is only necessary to record the upper triangular matrix. Two LOBs that do not have an intersection relationship are recorded as “-”.



(a)

	l_{a1}	l_{a2}	l_{b1}	l_{b2}	l_{b3}	l_{c1}	l_{c2}	l_{c3}	l_{d1}	l_{d2}	l_{d3}
l_{a1}	-	-	$P_{a1 \times b1}$	-	-	$P_{a1 \times c1}$	-	-	$P_{a1 \times d1}$	-	-
l_{a2}	-	-	$P_{a2 \times b1}$	$P_{a2 \times b2}$	-	$P_{a2 \times c1}$	$P_{a2 \times c2}$	-	$P_{a2 \times d1}$	$P_{a2 \times d2}$	-
l_{b1}	-	-	-	-	-	$P_{b1 \times c1}$	-	-	$P_{b1 \times d1}$	-	-
l_{b2}	-	-	-	-	-	$P_{b2 \times c1}$	$P_{b2 \times c2}$	-	$P_{b2 \times d1}$	$P_{b2 \times d2}$	-
l_{b3}	-	-	-	-	-	$P_{b3 \times c1}$	$P_{b3 \times c2}$	$P_{b3 \times c3}$	$P_{b3 \times d1}$	$P_{b3 \times d2}$	$P_{b3 \times d3}$
l_{c1}	-	-	-	-	-	-	-	-	$P_{c1 \times d1}$	-	-
l_{c2}	-	-	-	-	-	-	-	-	$P_{c2 \times d1}$	$P_{c2 \times d2}$	-
l_{c3}	-	-	-	-	-	-	-	-	$P_{c3 \times d1}$	$P_{c3 \times d2}$	$P_{c3 \times d3}$
l_{d1}	-	-	-	-	-	-	-	-	-	-	-
l_{d2}	-	-	-	-	-	-	-	-	-	-	-
l_{d3}	-	-	-	-	-	-	-	-	-	-	-

(b)

Figure 7. Intersection and cluster based on LOBs. (a) Schematic diagram of LOB and cluster point (b). Intersection relationship matrix of LOB: the association relationship between LOB and intersection and the inclusion relationship between cluster and intersection.

As shown in Figure 7a, the intersection point, p_i , is traversed. If the two intersection points are close to each other (the threshold is set to t), they are classified as a cluster and recorded as c_k , and the centroid of the cluster is recorded as $O(c_k)$. If there are no other intersections nearby, the points are recorded as a cluster alone. In Figure 7b, the matrix elements in the same cluster are marked with the same color, and each cluster can be expressed as:

$$c_k = \{p_i | \text{dist}(p_i, O(c_k)) < t\}, \tag{11}$$

where dist represents the function to calculate the distance between two points.

2.2.4. Elimination of Ghost Nodes Based on Constrained Rules

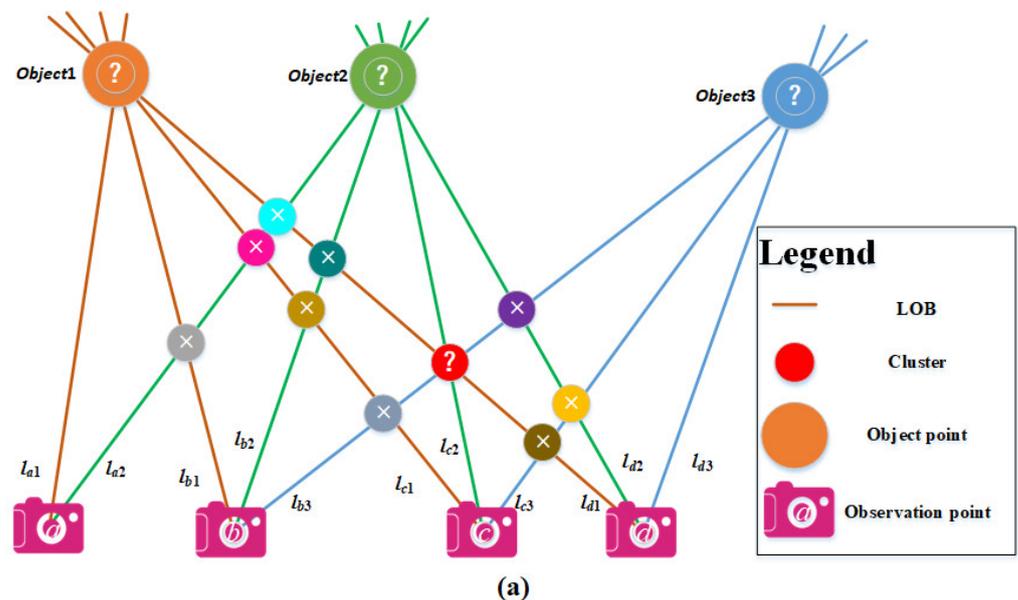
In this study, we introduce two constrained rules without setting dynamic thresholds. By recursively executing the constrained rules until the number of clusters no longer

changes, the effective elimination of ghost nodes is achieved as far as possible, which reduces the impact of ghost nodes on the positioning results.

1. Constrained rules based on the minimum number of intersections in the cluster

When the number of LOBs of the observed object is greater than 2, the number of intersections in the cluster should be greater than 1 [27,28,30,31]. The number of intersections contained in each cluster is counted. If there is only one intersection, the intersection contained in the cluster is determined as a ghost node. As shown in Figure 8a, all clusters (marked with “×”) that are determined to be ghost nodes are deleted, and the intersections within the cluster with LOBs in the intersection matrix are disassociated, as shown in Figure 8b. At this point, most of the ghost nodes have been eliminated, and only the clusters (marked with “?”), the candidate points of which are to be further judged, are retained.

2. Constrained rules based on the uniqueness of LOB association



(a)

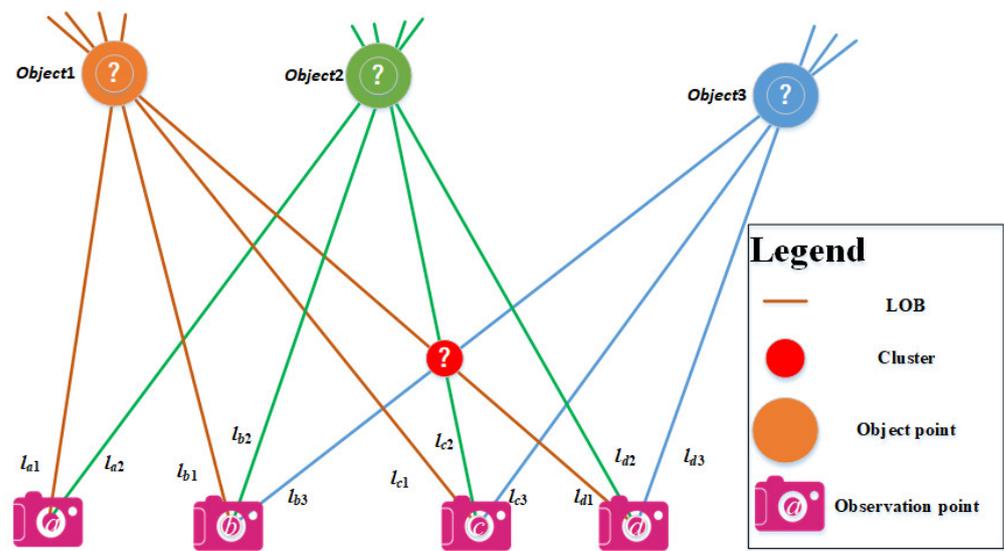
	l_{a1}	l_{a2}	l_{b1}	l_{b2}	l_{b3}	l_{c1}	l_{c2}	l_{c3}	l_{d1}	l_{d2}	l_{d3}
l_{a1}	-	-	$P_{a1 \times b1}$	-	-	$P_{a1 \times c1}$	-	-	$P_{a1 \times d1}$	-	-
l_{a2}	-	-	×	$P_{a2 \times b2}$	-	×	$P_{a2 \times c2}$	-	×	$P_{a2 \times d2}$	-
l_{b1}	-	-	-	-	-	$P_{b1 \times c1}$	-	-	$P_{b1 \times d1}$	-	-
l_{b2}	-	-	-	-	-	×	$P_{b2 \times c2}$	-	×	$P_{b2 \times d2}$	-
l_{b3}	-	-	-	-	-	×	$P_{b3 \times c2}$	$P_{b3 \times c3}$	$P_{b3 \times d1}$	×	$P_{b3 \times d3}$
l_{c1}	-	-	-	-	-	-	-	-	$P_{c1 \times d1}$	-	-
l_{c2}	-	-	-	-	-	-	-	-	$P_{c2 \times d1}$	$P_{c2 \times d2}$	-
l_{c3}	-	-	-	-	-	-	-	-	×	×	$P_{c3 \times d3}$
l_{d1}	-	-	-	-	-	-	-	-	-	-	-
l_{d2}	-	-	-	-	-	-	-	-	-	-	-
l_{d3}	-	-	-	-	-	-	-	-	-	-	-

(b)

Figure 8. Elimination of ghost nodes based on the rule of minimum number of intersections in the cluster: (a) deletion of clusters that do not satisfy this rule; (b) disassociation of relationships that do not satisfy this rule.

Each LOB is a line-of-sight simulation of the observation object. If the LOB is only associated with one intersection within a cluster, the cluster must be the object location. The LOB associated with this cluster should be disassociated from other clusters to ensure the uniqueness of the LOB association.

For example, as shown in Figure 9, l_{a1} has an intersecting relationship only with the intersection points in one cluster, so the cluster must be the object location, that is, *Object1*. Other LOBs associated with *Object1* should also be uniquely associated with *Object1* only. In addition to being associated with *Object1*, l_{d1} is also associated with other clusters. Therefore, it is necessary to disassociate the association between l_{d1} and the intersections in other clusters. At this time, the number of intersections in the red cluster is 1, which will be eliminated as a ghost node in the next iteration.



(a)

	l_{a1}	l_{a2}	l_{b1}	l_{b2}	l_{b3}	l_{c1}	l_{c2}	l_{c3}	l_{d1}	l_{d2}	l_{d3}
l_{a1}	-	-	$P_{a1 \times b1}$	-	-	$P_{a1 \times c1}$	-	-	$P_{a1 \times d1}$	-	-
l_{a2}	-	-	-	$P_{a2 \times b2}$	-	-	$P_{a2 \times c2}$	-	-	$P_{a2 \times d2}$	-
l_{b1}	-	-	-	-	-	$P_{b1 \times c1}$	-	-	$P_{b1 \times d1}$	-	-
l_{b2}	-	-	-	-	-	-	$P_{b2 \times c2}$	-	-	$P_{b2 \times d2}$	-
l_{b3}	-	-	-	-	-	-	$P_{b3 \times c2}$	$P_{b3 \times c3}$	$P_{b3 \times d1}$	-	$P_{b3 \times d3}$
l_{c1}	-	-	-	-	-	-	-	-	$P_{c1 \times d1}$	-	-
l_{c2}	-	-	-	-	-	-	-	-	$P_{c2 \times d1}$	$P_{c2 \times d2}$	-
l_{c3}	-	-	-	-	-	-	-	-	-	-	$P_{c3 \times d3}$
l_{d1}	-	-	-	-	-	-	-	-	-	-	-
l_{d2}	-	-	-	-	-	-	-	-	-	-	-
l_{d3}	-	-	-	-	-	-	-	-	-	-	-

(b)

Figure 9. Elimination of ghost nodes based on the uniqueness of LOB association: (a) clusters filtered by this rule; (b) associations filtered by this rule.

3. Experimental Results and Discussions

3.1. Data Collection and Selection of Research Area

In this study, we used the data collected by the Alpha 3D vehicle-mounted laser-scanning measurement system produced by CHC NAVIGATION for method validation. The system is equipped with a Ladybug panoramic camera, GNSS, IMU, and LIDAR. The original data obtained by the system are a series of binary stream data that cannot be

directly used by users, and a series of preprocessing operations are required. Through CoPre software developed by CHC NAVIGATION, the image stream data output by the Ladybug panoramic camera are read and spliced to form a panoramic image with a 360° viewing angle stored as a general picture format with 8192×4096 -pixel resolution. Through Inertial Explorer software, the IMU and GNSS data are jointly processed to obtain the high-precision driving trajectory (latitude and longitude coordinates, regional projection coordinates, and elevation), speed, attitude (roll, pitch, and heading), and other information in the specified coordinate system, which are output one-by-one to form a structured, readable text. The output trajectory data have a horizontal accuracy of 0.010 m and a vertical accuracy of 0.020 m. For the acquired attitude data, the roll/pitch accuracy is 0.005° and the heading is 0.017° [37]. The high-resolution imagery provides geometric texture and semantic information of street-side objects for the experiments. High-precision position and attitude data provide sufficient measurements for precision support for method verification.

As shown in Figure 10, an Alpha 3D vehicle-mounted laser-scanning measurement system was used to collect data from the Yanziji and Mufushan areas of Nanjing City, Jiangsu Province, China, and the Yincun Town area of Changzhou City, Jiangsu Province, China. Nanjing City and Changzhou City are located in the same province in China, and Yincun area and Mufushan and Yanziji area are only 120 km away, with similar street-layout styles.

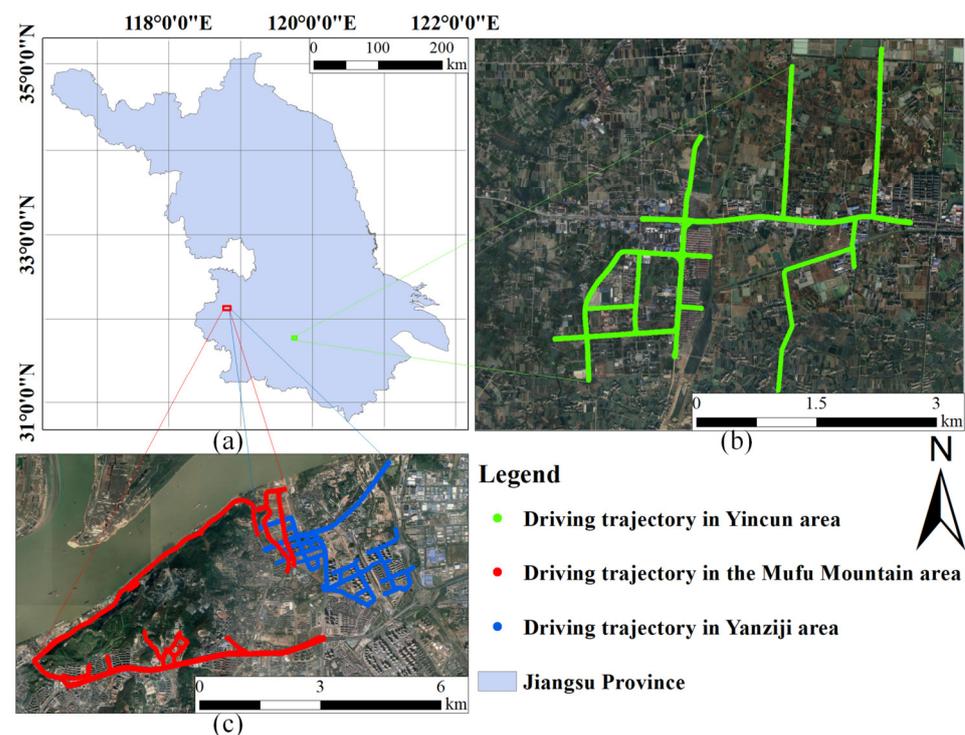


Figure 10. Schematic diagram of data collection and research area: (a) administrative border region of Jiangsu Province; (b) driving trajectory in Yincun area; (c) driving trajectory in Mufu Mountain and Yanziji area.

3.2. Object Detection and LOB Mapping

In this study, 6367 street-view images collected from the Yanziji area and 6920 street-view images collected from the Mufushan area with a resolution of 8192×4096 pixels were used as annotation data. Three classifications of pole-like objects—utility poles, street lamps, and signboards (which were widely and largely distributed)—were used as acquisition objects, and a workstation equipped with an Intel Xeon E5-2698 V4 CPU and a Tesla V100 GPU was used for training.

As shown in Figure 11, because most of the labeled data are distributed in the vertical center area of the image and the image distortion in this area is relatively small, for the convenience of training, the original image was cut into 53,148 2048×2048 subimages that only contain the middle area for object labeling. These pole-like objects were divided into rod parts and top parts for labeling, avoiding overlapping of bounding boxes as much as possible. The labeling results include: 48,162 rod parts, 7435 top parts of utility poles, 26,751 top parts of street lamps, and 5695 top parts of signboards. Taking this as the sample data, the samples were randomly divided into a training set and test set according to a 7:3 ratio and put into the Cascade R-CNN classifier [34] for training. The average precision of training was 0.880, and the recall rate was 0.929 (IoU > 0.5).

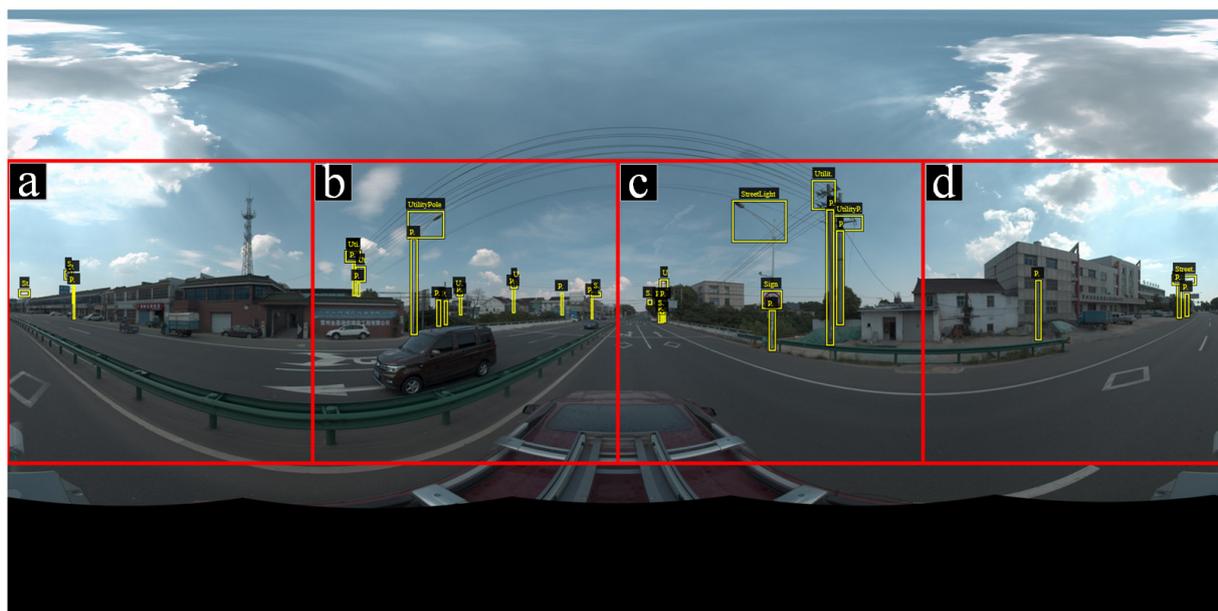


Figure 11. Schematic diagram of cutting and labeling of street-view image: letters (a) to (d) indicate that the street-view image is cut into four subimages.

A total of 4892 street-view images collected in Yincun Town were used for object detection. Because the location of the rod part is used as the object geolocation during positioning, in the detection results, the rod parts, top parts of utility poles, top parts of street lamps, and top parts of signboards were detected separately. The classifications of pole-like objects to which the rod-parts belong is given by the classifications of the closest top parts. This method relies on the orientation of the rod parts to map LOBs, which provides more accurate orientation parameters for the subsequent geographic positioning of the pole-like objects. In order to reduce the impact of recognition errors on the subsequent matching process, the classification results were manually checked, and a total of 6104 LOBs of pole-likes objects were mapped, including 3325 utility poles, 1814 streetlamps, and 965 signboards.

3.3. Geographic Positioning Based on Self-Adaptive Constrained LOB

The region division and geographic positioning algorithms are programmed in C# language and run under the Windows 10 operating system using a personal computer with an Intel Core i7-7700 CPU and 8GB RAM for calculation.

The calculation results of the proposed method were compared with the 1409 pole-like objects collected manually in this area, and the calculation results were evaluated from three indicators: time consumption, recall rate, and precision rate. The closer the recall rate and precision rate are to 1, the better the positioning effect of the algorithm is. The closer the time consumption is to 0, the more efficient the algorithm is. The time consumption is based

on the actual running time (running time is the calculation time of LOB-based geographic positioning). The recall rate and precision rate were calculated using Equation (12):

$$\begin{aligned} \text{Recall rate} &= \frac{N_{rc}}{N_{rf}} \\ \text{Precision rate} &= \frac{N_{co}}{N_{cal}} \end{aligned} \quad (12)$$

where N_{rf} represents the number of reference points, N_{cal} represents the number of calculated results, N_{rc} represents the number of reference points within a 1 m buffer range of all calculation results, and N_{co} represents the number of calculation results within 1 m buffer range of the reference point.

The LOB-based positioning method proposed in this study is affected by two thresholds: the enlargement coefficient, k , for dividing the grid and the distance, t , for aggregating the cluster points. In order to study the influence of the threshold parameter on the method, grid division is carried out according to the threshold parameter, and the LOB intersection and ghost-node elimination are performed grid-by-grid and classification-by-classification. According to the 4892 driving trajectories corresponding to the street-view images, the average baseline length of triangulation of adjacent views is calculated to be 7.21 m. The device can effectively capture pole-like objects within 100 m. The effective shooting distance, V , is set to 100 m. According to Equation (8), the value range of k is calculated as (1, 4.9]. For the convenience of calculation, an integer near this range is taken as the value of the threshold, k . Taking into account the error of recording parameters during acquisition, the distortion of street-view images, and the influence of the curvature of the earth on geodetic triangulation, the clustering distance, t , of the cluster is set to range from 0.1 m to 1 m, and the threshold value, t , is taken every 0.1 m. Table 1 shows the evaluation results of the geographic positioning method based on self-adaptive constrained LOB with different threshold combinations.

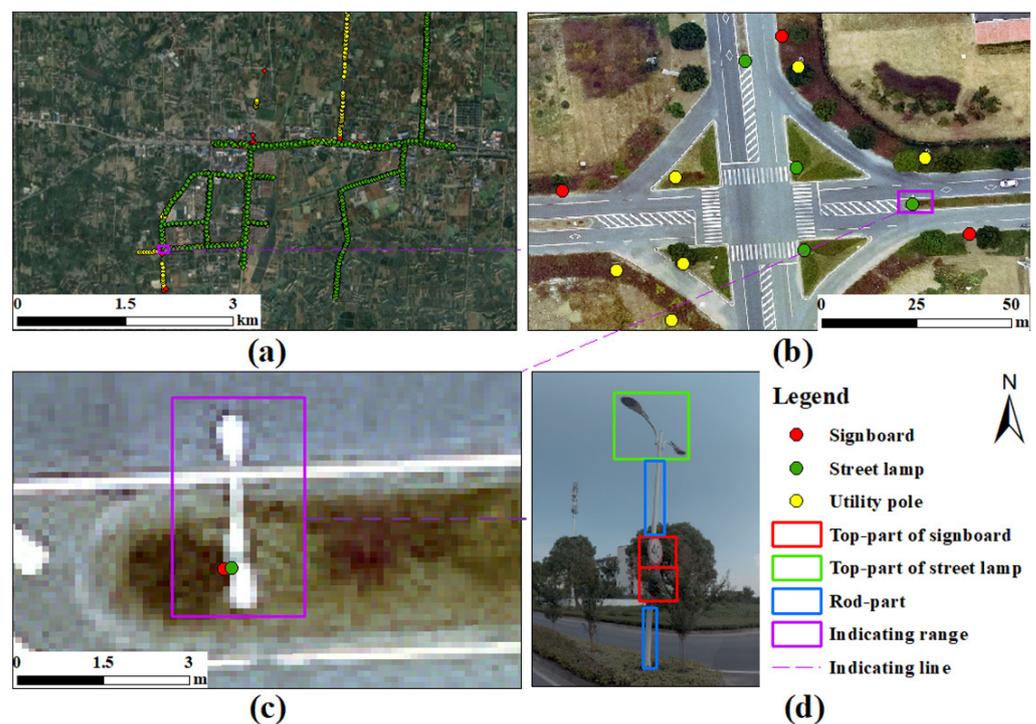


Figure 12. Calculation results: (a) distribution of calculation results of pole-like objects in the study area, (b) scaling of positioning results in the local area, (c) superposition of positioning results and high-resolution remote-sensing images, and (d) object images corresponding to positioning results.

Table 1. Evaluation results of the geographic positioning method based on self-adaptive constrained LOB.

k-Effective Viewing Distance (m)	Threshold of Cluster Distance (m)	Time Consumption (s)	Estimated Number of Poles	Recall Rate	Precision Rate
2–40.81	0.1	6.36	1509	0.88	0.97
	0.2	5.66	1393	0.92	0.97
	0.3	5.18	1374	0.92	0.97
	0.4	5.14	1357	0.91	0.97
	0.5	4.99	1344	0.9	0.97
	0.6	5.05	1342	0.9	0.96
	0.7	5.58	1321	0.89	0.96
	0.8	5.25	1328	0.89	0.96
	0.9	5.65	1319	0.88	0.96
	1	5.36	1316	0.88	0.95
3–61.22	0.1	5.79	1608	0.92	0.97
	0.2	5.57	1420	0.93	0.97
	0.3	6.08	1400	0.93	0.97
	0.4	5.31	1388	0.93	0.96
	0.5	5.22	1367	0.92	0.96
	0.6	5.18	1369	0.91	0.96
	0.7	5.12	1343	0.9	0.96
	0.8	5.09	1350	0.9	0.95
	0.9	5.47	1346	0.89	0.95
	1	5.19	1337	0.89	0.95
4–81.63	0.1	8.51	1617	0.92	0.96
	0.2	8.03	1422	0.93	0.96
	0.3	8.05	1401	0.93	0.96
	0.4	7.50	1393	0.93	0.96
	0.5	7.55	1364	0.91	0.96
	0.6	7.39	1373	0.91	0.95
	0.7	7.19	1356	0.9	0.95
	0.8	7.38	1365	0.9	0.94
	0.9	6.95	1360	0.89	0.94
	1	6.76	1346	0.89	0.94
5–102.04	0.1	13.01	1618	0.91	0.96
	0.2	11.86	1424	0.93	0.96
	0.3	11.43	1402	0.93	0.96
	0.4	11.28	1384	0.92	0.96
	0.5	11.18	1360	0.91	0.96
	0.6	10.74	1355	0.9	0.96
	0.7	10.44	1345	0.9	0.95
	0.8	10.04	1356	0.89	0.94
	0.9	9.95	1354	0.88	0.93
	1	9.74	1359	0.88	0.92

Within the estimated 100 m maximum effective range of the device, k takes 4 to make the maximum line-of-sight range 81.6 m, and the value of t is set to 0.2 m according to the output accuracy of the algorithm of the device. The calculation results obtained by this combination of thresholds are shown in Figure 12. The results prove the effectiveness of the proposed method for the positioning of large-scale pole-like objects. The positioning result has high accuracy and is associated with the corresponding object images, which can be imported into the database as the final result.

3.4. Comparative Analysis and Discussions with Existing Methods

In order to achieve a comparative analysis between the method proposed in this study and existing methods, in this paper, we reproduced the modified brute-force-based

LOB algorithm proposed by Zhang [30,31]. The algorithm is affected by three thresholds: the number of views, the angle, and the distance to the center of the selected road. The threshold is set according to the parameters provided in the paper and the actual situation of the data presented in this paper. The evaluation results are as follows.

Based on the modified brute-force-based LOB algorithm, there is some uncertainty in the selection of the threshold value. It is necessary to make multiple attempts to select the empirical value in combination with the data; especially when the road width is unknown or the road width varies greatly in a large-scale area, the threshold of the distance to the center of the selected road is often difficult to determine. It is easy to see from Table 2 that the modified brute-force-based LOB algorithm relies on expanding the threshold range to increase the number of candidate points, which takes more time. Although this can slightly improve the recall rate, it often leads to a decline in the accuracy of the recall results.

Table 2. Evaluation results of the geographic positioning method based on modified brute-force-based LOB.

Number of Views	Threshold of Angle (°)	Threshold of Distance to Center of Selected Road (m)	Time Consumption (s)	Estimated Number of Poles	Recall Rate	Precision Rate
3	1	10	5.20	1318	0.83	0.9
		15	5.20	1442	0.9	0.89
		20	5.28	1465	0.91	0.88
	2	10	5.82	1452	0.85	0.84
		15	5.79	1604	0.92	0.83
		20	5.57	1637	0.92	0.81
	3	10	5.68	1579	0.85	0.78
		15	5.85	1760	0.92	0.76
		20	7.19	1808	0.93	0.75
4	1	10	6.65	1358	0.83	0.87
		15	7.27	1498	0.9	0.86
		20	7.15	1538	0.91	0.84
	2	10	6.82	1522	0.85	0.81
		15	7.84	1699	0.92	0.78
		20	7.34	1755	0.93	0.76
	3	10	7.33	1696	0.85	0.74
		15	8.36	1917	0.93	0.71
		20	8.61	1994	0.93	0.69
5	1	10	6.86	1387	0.84	0.86
		15	8.08	1547	0.91	0.84
		20	8.03	1593	0.91	0.82
	2	10	8.27	1575	0.85	0.79
		15	9.78	1788	0.92	0.75
		20	9.62	1862	0.93	0.72
	3	10	9.30	1773	0.86	0.71
		15	10.83	2051	0.93	0.67
		20	11.53	2161	0.93	0.64
6	1	10	8.35	1416	0.84	0.85
		15	9.56	1590	0.91	0.82
		20	9.50	1651	0.91	0.79
	2	10	9.87	1625	0.85	0.76
		15	11.16	1872	0.92	0.72
		20	11.69	1981	0.93	0.68
	3	10	11.03	1844	0.86	0.69
		15	13.78	2182	0.93	0.63
		20	14.29	2342	0.94	0.59

Table 2. Cont.

Number of Views	Threshold of Angle (°)	Threshold of Distance to Center of Selected Road (m)	Time Consumption (s)	Estimated Number of Poles	Recall Rate	Precision Rate
7	1	10	8.94	1447	0.84	0.83
		15	10.33	1649	0.91	0.79
		20	11.45	1738	0.91	0.75
	2	10	11.59	1670	0.85	0.74
		15	14.13	1964	0.92	0.68
		20	15.01	2131	0.93	0.63
	3	10	13.89	1914	0.86	0.66
		15	17.95	2321	0.93	0.59
		20	20.04	2564	0.94	0.54
8	1	10	10.70	1467	0.84	0.82
		15	12.41	1709	0.91	0.76
		20	13.46	1838	0.91	0.71
	2	10	13.14	1704	0.85	0.73
		15	17.11	2064	0.92	0.65
		20	18.54	2290	0.93	0.59
	3	10	15.86	1966	0.86	0.64
		15	22.78	2465	0.93	0.56
		20	24.29	2804	0.94	0.49
9	1	10	11.65	1487	0.84	0.81
		15	14.55	1761	0.91	0.74
		20	16.09	1919	0.91	0.68
	2	10	15.05	1736	0.85	0.71
		15	20.37	2150	0.92	0.62
		20	22.99	2422	0.93	0.56
	3	10	19.08	2016	0.86	0.63
		15	26.94	2608	0.93	0.53
		20	31.65	3026	0.94	0.46

Geographic positioning based on the self-adaptive constrained LOB proposed in this study can automatically calculate the range of effective threshold k according to the driving trajectory, and the threshold t is a fixed value under the condition of unchanged equipment. The values of k within the effective shooting range achieved a good recall rate and accuracy rate, with a short calculation time. As the value of k increases, the range of the unit grid becomes larger, the calculation time also gradually increases, and the number of ghost nodes generated when the effective distance of LOB exceeds the actual distance also increases. The recall rate and the accuracy rate both decreased, but they still maintained a high level. Due to the influence of the acquisition equipment, the calculation results often cannot achieve high accuracy and produce offsets. If the cluster aggregation distance, t , is too small, the intersection points near the object location cannot be aggregated into a cluster, which would be used as the result of repeated acquisition. If the value of t is large, it causes adjacent objects of the same type to merge into a cluster, resulting in missing positioning and lowering the recall rate. Because the distance between street lamps and utility poles is large, a high value of t has little impact on them, although it has a considerable impact on signboards that are close to each other.

Compared with other LOB-based positioning methods with which it is difficult to select thresholds, the method proposed in this study is adaptable in threshold selection. The appropriate k value range can be automatically calculated through the driving trajectory, and the t value can take a fixed value according to the output accuracy of the equipment. Compared with the modified brute-force-based LOB algorithm, the proposed method limits the effective range of the LOB by dividing the grid and does not need to rely on thresholds, such as the number of adjacent view points and distance to the center of the selected road, which are affected by changes in road scenes.

4. Conclusions

In this study, we proposed a method for automatic positioning of objects in street-view images based on MMS. Aiming to reduce the difficulty of image feature matching due to a long baseline in street view, a geographic positioning method based on self-adaptive constrained LOB is implemented by referring to the object-matching algorithm based on a combination of object detection and LOB positioning. In order to overcome the time consumption and the difficulties of threshold selection caused by an LOB-based positioning algorithm, the idea of “divide–conquer” is introduced, and the calculation area is divided into grids according to the driving trajectory. The calculations in each grid are independent and do not interfere with each other, which greatly improves computing efficiency. In order to make the algorithm universal, a ghost node elimination algorithm based on self-adaptive constrained rules is proposed according to the line-of-sight rule when observing the object, which realizes the non-image feature matching of the same object in multi-view images.

Taking signboards, utility poles, and street lamps of multiple road sections in Yincun Town, Changzhou City, Jiangsu Province, as the experimental objects, experiments were carried out using multiple thresholds and compared with previous LOB-based object-positioning methods. The results show that the proposed method has higher efficiency and accuracy than previous methods, and the threshold selection range is clear and easy to promote. This method can perform automatic and accurate geographic positioning and image acquisition for a large range of street objects based on high-precision MMS, verifying its feasibility.

The results of this study are applicable in the acquisition of geolocation information for street objects, which can be used to draw high-precision maps required for autonomous driving and to provide data support for autonomous driving positioning, path planning, and traffic warning. Geolocation information on street-side objects can also assist in road safety detection and can help government departments to better manage and maintain urban living facilities and transportation facilities.

Street-side objects are easily blocked by vehicles, resulting in missing detection of targets. When the number of detections of the same object in multi-view images is less than three, the LOB-based positioning method cannot perform effective target positioning, and repeat acquisition is required for the road section in question. The data collection and positioning methods presented in this paper are not synchronized. With the support of software and hardware, the data stream of an MMS system can be converted into image and driving-trajectory data in real time. If the object-detection model proposed in this study is replaced with a lightweight model with higher detection efficiency, combined with the trajectory information obtained at a short distance, it would be possible to realize real-time image-based street-object positioning. This suggests the possibility of real-time online updating and sharing of high-precision maps in the future.

Author Contributions: Conceptualization, Guannan Li; Formal analysis, Guannan Li; Funding acquisition, Liangchen Zhou and Guonian Lv; Investigation, Guannan Li; Methodology, Guannan Li and Liangchen Zhou; Software, Guannan Li; Supervision, Bingxian Lin, Liangchen Zhou and Guonian Lv; Visualization, Guannan Li; Writing—original draft, Guannan Li; Writing—review & editing, Xiu Lu. All authors have read and agreed to the published version of the manuscript.

Funding: This research study was funded by the National Natural Science Foundation of China: 42076203.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

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

References

1. Escalera, A.D.L.; Armingol, J.M.; Mata, M. Traffic sign recognition and analysis for intelligent vehicles. *Image Vis. Comput.* **2003**, *21*, 247–258. [\[CrossRef\]](#)
2. Guo, C.Z.; Kidono, K.; Meguro, J.; Kojima, Y.; Ogawa, M.; Naito, T. A low-cost solution for automatic lane-level map generation using conventional in-car sensors. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2355–2366. [\[CrossRef\]](#)
3. Farhadmanesh, M.; Cross, C.; Mashhadi, A.H.; Rashidi, A.; Wempen, J. Highway Asset and Pavement Condition Management using Mobile Photogrammetry. *Transp. Res. Record.* **2021**, *2675*, 296–307. [\[CrossRef\]](#)
4. Cafiso, S.; Di Graziano, A.; Pappalardo, G. Safety Inspection and Management Tools for Low-Volume Road Network. *Transp. Res. Record.* **2015**, *2472*, 134–141. [\[CrossRef\]](#)
5. Zhao, X.; Ding, Y.; Yao, Y.; Zhang, Y.; Bi, C.; Su, Y. A multinomial logit model: Safety risk analysis of interchange area based on aggregate driving behavior data. *J. Saf. Res.* **2022**, *80*, 27–38. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Kim, J.G.; Han, D.Y.; Yu, K.Y.; Kim, Y.; Rhee, S.M. Efficient extraction of road information for car navigation applications using road pavement markings obtained from aerial images. *Can. J. Civ. Eng.* **2006**, *33*, 1320–1331. [\[CrossRef\]](#)
7. Abdalla, M.; Easa, S.M. Extracting streetlight poles from orthophotos: Methodology and case study in Ontario, Canada. *J. Surv. Eng.* **2007**, *133*, 184–187. [\[CrossRef\]](#)
8. Kim, G.H.; Sohn, H.G.; Song, Y.S. Road infrastructure data acquisition using a vehicle-based mobile mapping system. *Comput. Aided Civ. Infrastruct. Eng.* **2010**, *21*, 346–356. [\[CrossRef\]](#)
9. Puente, I.; González-Jorge, H.; Martínez-Sánchez, J.; Arias, P. Review of mobile mapping and surveying technologies. *Measurement* **2013**, *46*, 2127–2145. [\[CrossRef\]](#)
10. Wegner, J.D.; Branson, S.; Hall, D.; Schindler, K.; Perona, P. Cataloging public objects using aerial and street-level images—Urban trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
11. Liu, L.; Tang, X.; Xie, J.; Gao, X.; Zhao, W.; Mo, F.; Zhang, G. Deep-learning and depth-map based approach for detection and 3D localization of small traffic signs. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 2096–2111. [\[CrossRef\]](#)
12. Dong, G.; Yan, Y.; Shen, C.; Wang, H. Real-time high-performance semantic image segmentation of urban street scenes. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3258–3274. [\[CrossRef\]](#)
13. Gao, Y.; Xiao, G. Real-time chinese traffic warning signs recognition based on cascade and CNN. *J. Real-Time Image Process.* **2020**, *18*, 669–680. [\[CrossRef\]](#)
14. Timofte, R.; Van Gool, L. Multi-view manhole detection, recognition, and 3D localisation. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011.
15. Bayouth, K.; Hamdaoui, F.; Mtibaa, A. Transfer learning based hybrid 2D-3D CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems. *Appl. Intell.* **2021**, *51*, 124–142. [\[CrossRef\]](#)
16. Tsai, V.J.D.; Chang, C.-T. Feature position on Google street view panoramas. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2012**, *4*, 305–309. [\[CrossRef\]](#)
17. Nassar, A.S.; Lang, N.; Lefèvre, S.; Wegner, J.D. Learning geometric soft constraints for multi-view instance matching across street-level panoramas. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019.
18. Ogawa, M.; Aizawa, K. Identification of Buildings In Street Images Using Map Information. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, China, 22–25 September 2019.
19. Zhu, S.; Yang, T.; Chen, C. Revisiting Street-to-Aerial View Image Geo-localization and Orientation Estimation. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021.
20. Nassar, A.S.; D’Aronco, S.; Lefèvre, S.; Wegner, J.D. GeoGraph: Graph-based multi-view object detection with geometric cues End-to-End. In Proceedings of the Computer Vision—ECCV 2020, Glasgow, UK, 23–28 August 2020.
21. Krylov, V.A.; Kenny, E.; Dahyot, R. Automatic discovery and geotagging of objects from street view imagery. *Remote Sens.* **2018**, *10*, 661–671. [\[CrossRef\]](#)
22. Hebbalaguppe, R.; Garg, G.; Hassan, E.; Ghosh, H.; Verma, A. Telecom inventory management via object recognition and localisation on Google street view images. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017.
23. Pattipati, K.R.; Deb, S.; Bar-Shalom, Y.; Washburn, R.B. A new relaxation algorithm and passive sensor data association. *IEEE Trans. Autom. Control* **1992**, *37*, 198–213. [\[CrossRef\]](#)
24. Bishop, A.N.; Pathirana, P.N. Localization of emitters via the intersection of bearing lines: A ghost elimination approach. *IEEE Trans. Veh. Technol.* **2007**, *56*, 3106–3110. [\[CrossRef\]](#)
25. Reed, J.D.; Silva, C.R.C.M.D.; Buehrer, R.M. Multiple-source localization using line-of-bearing measurements: Approaches to the data association problem. In Proceedings of the MILCOM 2008—2008 IEEE Military Communications Conference, San Diego, CA, USA, 16–19 November 2008.
26. Chu, L.; Chen, W. Correction of Mobile Positioning and Direction via CNNs Based on Street View Images. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taichung, Taiwan, China, 19–21 May 2018.
27. Hazelhoff, L.; Creusen, I.; de With, P.H. System for semi-automated surveying of street-lighting poles from street-level panoramic images. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, 24–26 March 2014.

28. Hazelhoff, L.; Creusen, I.; de With, P.H. Robust detection, classification and positioning of traffic signs from street-level panoramic images for inventory purposes. In Proceedings of the 2012 IEEE Workshop on the Applications of Computer Vision (WACV), Breckenridge, CO, USA, 9–11 January 2012.
29. Lumnitz, S.; Devisscher, T.; Mayaud, J.R.; Radic, V.; Coops, N.C.; Griess, V.C. Mapping trees along urban street networks with deep learning and street-level imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 144–157. [[CrossRef](#)]
30. Zhang, W.; Witharana, C.; Li, W.; Zhang, C.; Li, X.; Parent, J. Using deep learning to identify utility poles with crossarms and estimate their locations from google street view images. *Sensors* **2018**, *18*, 2484. [[CrossRef](#)]
31. Khan, A.; Asim, W.; Ulhaq, A.; Ghazi, B.; Robinson, R.W. Health assessment of eucalyptus trees using siamese network from Google street and ground truth images. *Remote Sens.* **2021**, *13*, 2194. [[CrossRef](#)]
32. Soheilian, B.; Paparoditis, N.; Vallet, B. Detection and 3D reconstruction of traffic signs from multiple view color images. *ISPRS-J. Photogramm. Remote Sens.* **2013**, *77*, 1–20. [[CrossRef](#)]
33. Timofte, R.; Zimmermann, K.; Van Gool, L. Multi-view traffic sign detection, recognition, and 3D localisation. *Mach. Vis. Appl.* **2014**, *25*, 633–647. [[CrossRef](#)]
34. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
35. Bruno, N.; Roncella, R. Accuracy Assessment of 3d Models Generated from Google Street View Imagery. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W9*, 181–188. [[CrossRef](#)]
36. Tsai, V.J.D.; Chang, C.T. Three-dimensional positioning from Google street view panoramas. *IET Image Process.* **2013**, *7*, 229–239. [[CrossRef](#)]
37. Alpha 3D Vehicle-mounted Laser Scanning Measurement System. Available online: http://www.huace.cn/product/product_show1/431 (accessed on 19 February 2022).