A Landmark Selection Method for Object-Based Visual Outdoor Localization Approaches of Automated Ground Vehicles

Florian Frank[1], Peter Buckel[2], Ludwig Hoegner[3], Petra Hofstedt[4]

[1]Institute for Continuing Education, Knowledge and Technology Transfer (IWT), Friedrichshafen, Germany – frank@iwt-bodensee.de
[2]Baden-Wuerttemberg Cooperative State University (DHBW), Friedrichshafen, Germany - peter.buckel@dhbw.de
[3]Department of Geoinformatics, Hochschule München University of Applied Sciences, Munic, Germany – ludwig.hoegner@hm.edu
[4]Programming Languages and Compiler Construction, Brandenburg University of Technology Cottbus - Senftenberg, Cottbus, Germany – hofstedt@b-tu.de

Keywords: Outdoor Localization, Object-Based, Camera, Landmark, Autonomous Driving Dataset

Abstract

Autonomous vehicles must navigate independently in an outdoor environment using features or objects. However, some objects may be more or less suitable for localization due to their attributes. Therefore, this work investigates the suitability of landmarks for camera- and object-based outdoor localization methods. First, object attributes are methodically derived from the requirements of object-based localization. The physical representation on the camera image plane, probability of occurrence, and persistence were identified as influencing the object localization suitability. The influence of the object's camera image plane representation regarding object recognition algorithms is not considered or discussed, but advice on the minimum object pixel size is provided. The first milestone was the creation of an equation for object localization suitability calculation by normalizing and multiplying the identified attributes. Simultaneously, potential objects from the outdoor environment were identified, resulting in a structured object catalog. The results of the equation and catalog are a ranked according to the object localization suitability in a comparison table. Our comparison demonstrates that objects such as buildings or trees are more suitable than street lane markings for self-localization. However, most current datasets do not include the proposed instantiated objects. The paper addresses this issue, assists in the object selection for outdoor localization methods and provides input for the creation of future-oriented datasets and autonomous driving maps.

1. Introduction

A significant challenge in Autonomous Driving (AD) and Automated Ground Vehicles (AGV) is the safe, robust, and autonomous localization in different outdoor environments. Georeferenced information is the base of localization which is captured by various sensor systems, fused, processed, and stored. State-of-the-art localization approaches are primarily based on the feature level (Zekavat et al., 2021). Alternatively, object-based localization methods could enhance robustness by storing multiple information about objects.

Object-based localization methods are mainly used indoors due to the defined and standardized environmental objects, which reduce complexity. Outdoor localization has a higher complexity due to many different object classes embedded in different environments and weather conditions. Therefore, outdoor localization methods require high-fidelity object-level maps of the environment as georeferenced information. The selection of objects for map creation affects the detection, pose estimation, and vehicle relocation (Siciliano et al., 2009). Finally, the object selection serves as the foundation for every object-based localization system.

For this purpose, we present a systematic approach to the selection of environment objects for the creation of visual and object-based maps.

2. Related Work

The related work section considers machine vision localization approaches and their datasets. The authors of the paper 'An Overview on Position Location: Past, Present and Future', mention that localization techniques have been integrated into everyday life (Zekavat et al., 2021). They summarize most localization methods and categorize methods into 'visual object-oriented localization' approaches.

Feature-based outdoor localization approaches are mostly based on LiDAR, camera, and RaDAR sensor systems, which enable autarkical localization (Sattler et al., 2018, Schaupp et al., 2020, Qin et al., 2018). These localization approaches are mainly based on algorithms that extract key points from the environment and store these georeferenced landmarks on a map. This idea is extended to edge-based (Ballardini et al., 2021) and pole-based (Schaefer et al., 2021, Yu et al., 2018) approaches.

Ellipsoid approaches use artificial intelligence (AI) to recognize objects within images. The method is abstracting the single object from the image plane to an ellipsoid and using its midpoints for localization (Ok et al., 2019, Tian et al., 2021, Rubino et al., 2018). The mentioned approaches use parked cars as localization objects. As ground truth, two methods use the Kitti dataset (Geiger et al., 2013), and one uses the DARPA Fast Lightweight Autonomy dataset (Paschall and Rose, 2017). The limitation of the localization to parked vehicles is explained by the datasets that do not include other instantiated 3D objects.

Semantic signature localization methods associate classes of semantically segmented georeferenced images with robust features from algorithms, such as ORB or SIFT, for localization (Murali et al., 2018, Weng et al., 2021). Weng Li et al. (Weng et al., 2021) use eleven semantic classes near the roadside in the Open Paris dataset (City of Paris, 2022). Similarly, V. Murali et al. use twelve classes in their method based on the CamVid (Brostow et al., 2009) and the Kitti datasets (Geiger et al., 2013). Both methods describe a potential improvement through better segmented environmental data.

The localization approaches with object recognition and perspective-n-point transformation use key points of objects to determine the relative vehicle position (Qu et al., 2015, Lecrosnier et al., 2019). The approaches often use planar describable objects or object parts, such as traffic signs, buildings, or road markings. Their localization methods require object visibility and description in six degrees of freedom (6DOF). Today's datasets and maps are mostly limited to objects in the road space, such as road markings or traffic signs. As a result, object recognition drops significantly in adverse weather conditions and may lead to failures. A potential improvement could be an object-based digital twin of the environment next to the roadside.

Expanding the datasets to include a detailed, instantiated, segmented environment would be a major enhancement for all object-based localization methods Many well-known datasets provide algorithms and methods for automated driving, such as the Kitti (Geiger et al., 2013), CamVid (Brostow et al., 2009), Mapillary Vistas (Neuhold et al., 2017), Cityscapes (Cordts et al., 2016), A2D2 (Geyer et al., 2020), ApolloScape (Huang et al., 2020), and Toronto Dataset (R. Garnett et al., 1998). They do not provide an instantiated in a detail segmented environment and their digitalization is strictly limited to road spaces. The LoD3 Road Space Model of Ingolstadt by B. Schwab et al. (Schwab and Wysocki, 2021) represents the counterpart of the datasets named above. It presents a high-quality environmental building model, without labeled sensor data for algorithm development.

This paper addresses the question: Which objects should be instantiated and segmented from the external environment so that future object-based localization algorithms can be developed and benchmarked?

3. Object Catalog Creation Methodology

3.1 From feature-based Localization to the object-based Localization

Feature-based and object-based approaches tackle the challenges of pose and position using different methods. So, feature-based approaches are more sensitive to environmental variation, e.g., light conditions, view angles, and wet surfaces. Complex shapes, deformed, and/or various scaled objects may challenge the feature extraction process. The extraction may also need appropriate data quality regarding real-world digitization difficulties such as noise. Additionally, features only have limited contextual information, which can lead to the context being lost in the scene.

In comparison, object-based localization approaches may be more robust regarding object changes and noise problems. For example, a tree with snow can still be interpreted as a tree, whereas extracted features covered with snow may be lost. Objects can provide context information and have spatial relationships, enhancing information for localization approaches. Object-based localization possibly needs more data processing and computation power than feature-based approaches.

In summary, object-based localization may increase safety and robustness but also increase the data processing complexity.

3.2 The basic Idea of the Methodology to create the Object Catalog

Visual localization is generally a controlled process for estimating one's own pose and position in six degrees of freedom (6DoF) with continuous updating based on environmental information. Figure 1 shows the information flow during the process: The environmental information is transported over the air, perceived with sensors, processed, and interpreted by the brain or an electronic control unit.

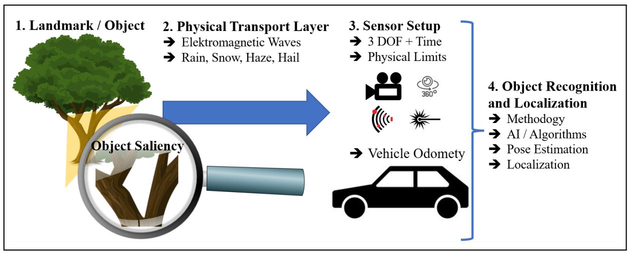

Figure 1. The physical layer of localization information flow and processing from the left to the right.

According to the physical layer, the research for the object catalog starts with requirements of visual-object camera detection, pose, and positioning. With this information, real-world objects are transferred via a camera pinhole model to an image plane, to estimate the median individual object size during drive-by. The median visual information in combination with statistics like their probability of occurrence, geo-information, and lifetime create a formula for ranking the single object localization suitability. In the second step, research of suitable real-world environmental objects and their information for localization is done. Lastly, for each object, the localization suitability is calculated and ranked.

3.3 Localization Requirements to Objects and Object Catalog

The localization requirements for objects and the object catalog are divided into the subcategories of application-specific attributes, object detection, and pose estimation.

3.3.1 Application-Specific Attributes for Localization

These attributes cover the fundamental requirements of functionally safe localization and affect the creation of the object catalog. Starting with functional assurance, T. G. R. Reid et al. (Reid et al., 2019) describe the absolute accuracy of vehicle localization in road space below 10cm in longitudinal and lateral range and an absolute angle of $0.17°$ pointing accuracy. According to these assumptions, the object position accuracy has to be measured more accurately than the localization system inaccuracy. In addition, the Safety Of The Intended Functionality (SOTIF) requires a permanent redundancy of at least three objects (ISO, 2019). The condition of overlapping and disappearances of objects results in a minimum number of four persistently detectable objects. Grounded objects, such as lane markings, should only be used as a support, as they may not be physically detectable by the sensor system due to adverse weather conditions or dirt.

3.3.2 Object Detection and Pose Estimation

Minimum Set of Object Detection Requirements The minimal requirements for possible object detection depend significantly on the raw sensor data. These requirements are explained for the camera to create a correlation between the sensor system design and the objects. Lidar and Radar object detection are not pursued further in this paper. Y. Cai describes in his paper 'How Many Pixels Do We Need to See Things?' (Cai, 2003), the minimum resolution of an object in a camera image so that humans can recognize it. The number of pixels ranges from 18×18 pixels for a face to 47×47 pixels for complex objects. T. Unel et al. (Unel et al., 2019) propose a minimal feature map for person detection of 38×38. This meets with the 35×35 of figure detection (Cai, 2003) and leads to the assumption that computers and humans of abstract pixel-based information are close together. Deriving these assumptions to the minimum representative object size in any pose on the image sensor plane, the size should be similar to the complex object with 47×47 pixels or larger. In order to minimize the uncorrectable errors such as blurring, vignetting, noise, and optical flow, the representative object display may need to be enlarged.

The minimal object representation size assumption can be used for optical path and system design. Figure 2 shows the dependency of the object to sensor distance and the object size. This must be considered as a scaling factor in the sensor system design. Based on the assumptions above and the premise that the object side length and the pixel side length are on one plane, the formula for the camera can be set up according to the intercept theorem to fulfill the minimal object detection requirement:

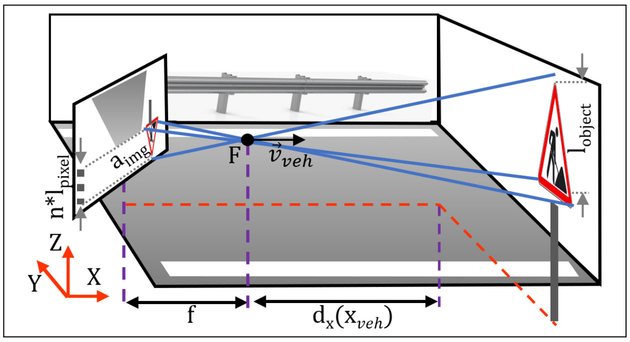

Figure 2. Simplified pinhole model for object representation on camera image plane. Used for proposed minimal object detection requirements and normalization.

$$\frac{l_{obj,img}/d}{l_{pixel}/f} \geq n_{pixel,min} \tag{1}$$

Modern object detection algorithms such as improved MMF-YOLO (Zhang et al., 2022) use feature fusion to detect objects smaller than 32×32 pixels and relate the object-detectability to their size. For simplicity and safety, we propose that projected pixels in line should be greater or equal to the assumed detection requirement of complex scenes with
$n_{pixel,min} = 47$ pixels (Cai, 2003).

Disruptive Influences The essential prerequisite for object-based localization is its detection. Disruptive influences such as bad weather, fading, darkness, and sources of interference are dynamic effects that significantly impact object detection. The influences are dependent on both distance and location and are considered to be complex to describe. Therefore, the attributes are not further considered.

Physical Attributes for Object Description Physical attributes describe objects in their geometric shapes, surfaces, and material attributes. Geometric attributes can be divided into two- or three-dimensional shapes, such as traffic signs or cars. Current datasets annotate objects with segmented 3D bounding boxes. A more accurate annotation with instantiated 3D bounding boxes is a minimum requirement. A further improvement could be the use of a more accurate CAD or mesh model.

The second component that contributes to object detection and pose estimation is surface texture and material behavior. The ideal solution would be to digitize the surface behavior of the objects by capturing the absorption, reflection, and transmission behavior over the electromagnetic spectrum. Further, this would lead to the abstraction of different electromagnetic sensors, viewing angles, and light interference and reduce the gap between the real and simulation world.

Geographical Object Attributes The geographical attributes describe the object on the map. These attributes are categorized into location, environment, and data format.

The derivation of the object location results in the attributes digitizability, statistics, update cycle, and persistence. The digitizability of an object requires its recording, detection, and pose estimation from a distance. To achieve this, objects should be located within a radius of 300m around the vehicle, within sight distance of the sensor system. The radius was chosen to provide a sufficient number of objects for object-based localization methods in rural environments. Leading from this, the attribute **probability of occurrence (PO)** measured in objects per kilometre can be derived. The persistence of objects and object changes correlate with the update cycle of the map. For these two attributes, the key indicators can be infrastructure maintenance and traffic counts.

Objects and object classes should be assigned to environments depending on their appearance-behavior. The goal is to create a cluster of objects that fulfill the object density for functionally safe localization. The suggested environment categories are city, countryside, forest, nature, and highway.

3.4 Perspective Normalization of Object Imaging

The idea of perspective normalization of object imaging is to get an average projected area of the object on the camera image plane by driving along. It is assumed that the average projected area is a factor for the suitability of the object for localization methods and indirectly as well for its recognition. The individual landmark objects are located at different positions in the road space, so they are imaged differently on the camera image plane, depending on their relative distance and 6DOF from the vehicle. The object points are transformed from the 3D world via transformation matrix P to the 2D-pixel coordinates (2) (Hartley and Zisserman, 2011).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2}$$

Assuming that the vehicle with a forward-facing camera travels along a straight road, the longitudinal distance to the object decreases and the object's projected area on the image plane changes. The normalization is done via the integration of the object's projected area along the longitudinal distance $d_x \in [d_{min}, d_{max}]$ to a frustum and its division through the distance (3). An example of a house projected frustum is shown in Fig. 3.

$$a_{img,norm} = \frac{1}{d_{max} - d_{min}} \cdot \int_{d_{min}}^{d_{max}} a_{img}(x_{veh})\, dx \quad (3)$$

In addition, there may be a certain loss of object projection on the camera image plane due to an object being out of the Field of View (FoV) or the projection extending beyond the image plane.

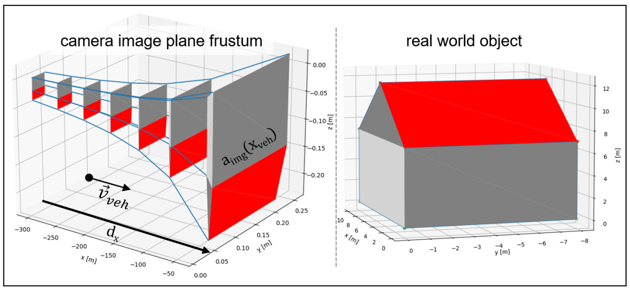

Figure 3. Real-world house-object (right) projections on the camera image plane (left) while driving along in X direction. The connecting lines between the image planes indicate the frustum of integration.

3.5 Object Localization Suitability

The object localization suitability is calculated according to the basic idea in subsection 3.1. As a result, the object suitability was chosen to rely on three factors. The first factor is the normed single object dimension $a_{img,norm}$ on the camera image plane while driving along. The second is the object probability of occurrence **(PO)** in the environment as an essential factor. Additionally, object persistence **(PS)** affects the updating map cycles. All three factors are multiplied without weighting to calculate the object localization suitability in formula (4).

$$suitability = a_{img,norm} \cdot PO \cdot PS \quad (4)$$

The equation was kept simple to allow reproducibility.

3.6 Object Classes for the Creation of an Object Catalog

This section lists potential-spatiotemporally static objects and classes for localization methods in the German road space. Objects or object classes must be replaced, added, or removed for other countries and regions. This paper does not claim completeness of the objects and object classes. It only provides a collection of suggestions.

The referenced datasets in section 2 only annotate the data of individual objects into object classes. Hence, note that this paper addresses the proposal of an instance and panoptic segmented annotation of objects. The explanation of the classification follows the top-down principle and the approach of logical, self-explaining ontology. The presented classification is intended to serve as a template for future related work and to create consistent naming conventions.

At the top level, an object distinction is made between artificial and natural objects.

3.6.1 Artificial Objects

Artificial objects are human-made and can be divided according to their function and geographical locality into the **ground**, **street furniture**, **buildings**, **city furniture**, **supply and communication infrastructure**, **orientation objects**, and **miscellaneous**.

The class **ground** is assigned grounded surface areas with their associated function. Examples are sidewalks, roads, or parking lots.

The **street furniture** class includes all objects for traffic guidance and vehicle routing. For a better overview, the class is subdivided into traffic signs, lightning, restraint systems, obstacles, and mounting systems. Traffic signs can be found in the German traffic sign catalog, including traffic signs, wayfinding, road markings, traffic barriers, and traffic flow control. Lighting equipment is divided into traffic-relevant light signal systems, such as traffic, warning, and street lights. Streetlights are only regulated in their lighting behavior by EN132301 and are not standardized in appearance. Passive protective devices, such as traffic barriers are assigned to traffic restraint systems. Traffic obstructions on the road, such as pollards or big plant pots, impede or stop traffic flow. The last group is mounting systems that distance many objects, like traffic signs, from the ground.

The **building** class is a central part of human civilization and living. Buildings tend to have large dimensions that make object detection in urban areas more challenging due to short distances. This is caused by the requirement that objects must be in the sensor FoV as a whole. Therefore, we propose to divide buildings into granular sub-objects, like masonry, windows, doors, roof ridges, gutters, signs, and lighting.

The class **city furniture** contains objects that may support localization, like trash cans, clocks, telephone booths, billboards, etc. Sporadically appearing objects, like speed cameras or roadside bales, should be classified as **miscellaneous**.

Supply and communication infrastructure is described in the German spatial ordinance. That means power lines, pipelines, and waterways are planned and built parallel to linear route corridors like roads and rails. Therefore, objects of these classes, such as power poles, are primarily found in rural areas beside roads. Maintenance-related buildings, like maintenance holes and fire hydrants, are also assigned to this class. In addition, radio masts for telecommunication are assigned to this category.

Objects that are memorable to people are called **orientation objects** (Gooley, 2020). This object class includes wayside crosses, birdhouses, boundary stones, and oversized symbols with a recognition value.

3.6.2 Natural Objects

Localization in natural environments has always been essential for human survival. Throughout of evolution, the localization instincts of humans decreased. Applied to today, this should be considered as an essential part of the localization methodology, because it can be used in regions with weak infrastructure, such as rural areas or unpaved roads. For object-based localization, natural objects are divided into the classes of **vegetation**, **bodies of water**, **landforms** and **animal -ade buildings**. The **vegetation** includes trees, bushes, meadows, and fields. Small

https://doi.org/10.5194/isprs-annals-X-4-W5-2024-163-2024

plants, such as flowers, are not maintained to due their short lifespan and size. **Bodies of water** make their mark on the landscape in the form of rivers, lakes, or seas. **Landforms** can be cliffs, hills, mountains, forest edges, or silhouettes. **Animal-made buildings** such as anthills or bird nests can point the way.

The number of different classes of natural objects is much smaller than the number of classes of artificial objects. Nevertheless, natural objects are omnipresent.

4. The Object Catalog

The object catalog results from listing potential object classes with attributes according to their suitability for object-based localization methods. An excerpt of the catalog is shown in Table 1. The excerpt contains a subset of objects and attributes to calculate the location **suitability** described in section 3.5 and **instance availability (IA)** in datasets. The complete version of the catalog with 22 objects, ten object attributes, related resources, and an object classes graph, is available on GitHub at `https://github.com/Crackzero/ObjectCatalog`.

object	a_{norm}	**PO**	**PS**	**suitability**	**IA**
unit	[m^2]	[km^{-1}]	[yr]	[m^{2}km^{-1}yr]	
building	0.22222	30.78	60	68.4033	-
tree	0.07667	22.02	200	16.8844	-
lane (road)	0.17980	3.33	30	5.9933	x
edge of the forest	0.24000	1.96	100	4.7005	-
electricity pylon	0.03333	2.46	80	0.8204	-
lane markings	0.00084	3.33	5	0.2997	x
traffic signs	0.00071	37.48	16	0.2665	x
traffic barrier	0.18083	0.10	50	0.1846	~
street lights	0.00064	15.15	40	0.0970	~
anthill	0.00040	14.83	15	0.0593	-
trashcan	0.00030	5.19	12	0.0156	x
delinator	0.00027	9.73	5	0.0130	x
traffic light	0.00067	0.08	10	0.0005	x

Table 1. Excerpt of the object catalog. Probability of occurrence (**PO**); persistence (**PS**); instance availability in datasets (**IA**); '**x**' as instantiated; '**~**' as segmented; '**-**' as unavailable.

4.1 Assumptions and Column Explanation

The object image plane representation area a_{norm} is normalized as described in section 3.4. The focal point $F = [0\,\mathrm{m}, 0\,\mathrm{m}, 1.2\,\mathrm{m}]$ and the distance of the image plane $f = 1\,\mathrm{m}$ has been adopted. The range of $d_x \in [-300\,\mathrm{m}, 0\,\mathrm{m}]$ was determined by driving towards and along the object on the road. The size of the image plane was assumed not to be limited by multi-camera systems. The adopted object dimensions, areas and their associated resources can be found in the online version. The building class represents an assumed, average German house with all its related sub-objects like windows, doors, etc.

The **probability of occurrence (PO)** values were collected from various sources. References of those are in the online version. The values are given per square kilometer and are assumed to be evenly distributed over the area of Germany.

The **persistence (PS)** is based on the mean lifetime. The influence of persistence in the formula was limited to $persistence \in [0, 10]$ years. Object localization **suitability** was calculated according to 3.4. The objects' instance availabilities in datasets were also listed to demonstrate the current situation of instance-segmented datasets.

4.2 Results

To better understand the results, the distribution of the table is discussed first The table was sorted by suitability from good, large, and not-so-good, small values. In this context, suitability should be understood as a normalized object area on an image plane in square metres for years of durability per kilometres driven.

Trees and buildings are located in front of the omnipresent roads. This arrangement is probably based on the assumption that all buildings in Germany are located along streets and that trees are equally distributed over the area. Another critical factor is the minimum distance to the object, which for most objects is the possible minimum of $d_{x,min} = 1.5\mathrm{m}$ to cover a full range of possibilities. A tenfold increase in the average minimum distance to $15\mathrm{m}$ and doubling the lane marking width would improve the suitability of buildings just behind the lane from 6.6 to 12. Persistence is mainly irrelevant in this comparison because of the ten-year limit.

Traffic lights and delineators are omnipresent and therefore at the end of the table. The delineators took these places due to their limited visible area and the traffic lights due to their low occurrence.

5. Discussion

The presented method for the calculation of suitabilities could be described as simple. Its weaknesses will be discussed in this section. The formula for object suitability calculation does not consider a weighting of the factors. Accordingly, to the value range, the suitability is mainly influenced by the normalized representation of object areas a_{norm} and probability (PO). The majority of values were gathered from diverse online sources, accessible through the GitHub online version. The factor a_{norm} depends mainly on the object dimensions and the minimum distance d_{min}. These values are not normed and are estimated based on experience or median values. The suitability calculation does not consider nested objects, features, gradients, textures, or colors, which may directly influence object recognition. The only mention of object recognition is the pixel requirement, which is not taken into account for reasons of generalizability. The resulting suitabilities are based on numerous assumptions and should be viewed as estimated values with potential falsification. The ranking should help to identify additional potentially suitable candidates for object-based localization methods.

Initially, this paper proposes an object selection method and object catalog for outdoor object-based localization methods for automated vehicles. Consequently, no comparable work has been found yet. The IA column in the table indicates the instance availability of objects to datasets mentioned in related work. The missing instance availabilities should be considered in creating new or improving existing datasets in the future.

6. Conclusion

This paper shows a straightforward methodology to calculate the object suitability for object-based localization methods. Object attributes for potential detection, pose determination, description, and geographical attributes were derived based on object-based localization requirements.

The object suitabilities for object-based localization methods include normalized object representation on a camera image plane, its probability of occurrence, and persistence. The normalization is done while driving alongside the objects beside the road and the average projected area on a camera image plane. The statistical values of occurrence probability and persistence were gathered from various sources and can be considered as potential error sources. The method of suitability calculation can be described as simple due to its multiplication and limiting value functions. Further improvements could be made by including additional attributes and weighting the influence of the values.

The main result of this paper is an object catalog in table format that sorts objects according to their suitability for object-based localization methods. Buildings and trees were ranked the highest due to their omnipresent location along roads and their high density in urban areas. In conclusion, it is advisable to include buildings, trees, and other objects beside the road in object-based localization methods. With the presented suitability ranking, the paper prioritizes suitable objects for outdoor camera-object-based localization methods.

The object attributes of instance availabilities present that a large part of the environment is either absent or only partially represented in datasets. Future datasets for object-based localization methods should be developed with appropriate quality. These data sets should include a digital twin on the object level of the environment, as well as instantiated and segmented raw sensor data.

References

Ballardini, A. L., Fontana, S., Cattaneo, D., Matteucci, M., Sorrenti, D. G., 2021. Vehicle Localization Using 3D Building Models and Point Cloud Matching. *Sensors (Basel, Switzerland)*, 21(16).

Brostow, G. J., Fauqueur, J., Cipolla, R., 2009. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2), 88–97.

Cai, Y., 2003. How many pixels do we need to see things? *Computational Science — ICCS 2003*, Lecture Notes in Computer Science, 2659, Springer Berlin Heidelberg, Berlin, Heidelberg, 1064–1073.

City of Paris, 2022. Open data initiative of the city of paris.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 3213–3223.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11), 1231–1237.

Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A. S., Hauswald, L., Pham, V. H., Mühlegg, M., Dorn, S., Fernandez, T., Jänicke, M., Mirashi, S., Savani, C., Sturm, M., Vorobiov, O., Oelker, M., Garreis, S., Schuberth, P., 2020. A2d2: Audi autonomous driving dataset.

Gooley, T., 2020. *The Natural Navigator*. Virgin Books.

Hartley, R., Zisserman, A., 2011. *Multiple view geometry in computer vision*. Second edition edn, Cambridge University Press, Cambridge.

Huang, X., Wang, P., Cheng, X., Zhou, D., Geng, Q., Yang, R., 2020. The ApolloScape Open Dataset for Autonomous Driving and Its Application. *IEEE transactions on pattern analysis and machine intelligence*, 42(10), 2702–2719.

ISO, 2019. Iso/pas 21448:2019: Road vehicles — safety of the intended functionality.

Lecrosnier, L., Boutteau, R., Vasseur, P., Savatier, X., Fraundorfer, F., 2019. Vision based vehicle relocalization in 3d line-feature map using perspective-n-line with a known vertical direction. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, 1263–1269.

Murali, V., Chiu, H.-P., Samarasekera, S., Rakesh, Kumar, 2018. Utilizing semantic visual landmarks for precise vehicle navigation.

Neuhold, G., Ollmann, T., Bulo, S. R., Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 5000–5009.

Ok, K., Liu, K., Frey, K. M., How, J. P., Roy, N., 2019. *Robust Object-based SLAM for High-speed Autonomous Navigation*. IEEE, Piscataway, NJ.

Paschall, S., Rose, J., 2017. Fast, lightweight autonomy through an unknown cluttered environment: Distribution statement: A — approved for public release; distribution unlimited. *2017 IEEE Aerospace Conference*, IEEE, 1–8.

Qin, T., Li, P., Shen, S., 2018. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020. http://arxiv.org/pdf/1708.03852v1.

Qu, X., Soheilian, B., Paparoditis, N., 2015. Vehicle localization using mono-camera and geo-referenced traffic signs. *2015 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 605–610.

R. Garnett, D. Carr, J. Rayes, S. Chowdhury, C. Hernandez, S. Mahmood, Y. Tan, R. Youssef, G. Qi, 1998. City of toronto open data portal.

Reid, T. G. R., Houts, S. E., Cammarata, R., Mills, G., Agarwal, S., Vora, A., Pandey, G., 2019. Localization Requirements for Autonomous Vehicles. *SAE International Journal of Connected and Automated Vehicles*, 2(3). http://arxiv.org/pdf/1906.01061v1.

Rubino, C., Crocco, M., Del Bue, A., 2018. 3D Object Localisation from Multi-View Image Detections. *IEEE transactions on pattern analysis and machine intelligence*, 40(6), 1281–1294.

Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., Pajdla, T., 2018. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. *CVPR 2018*.

Schaefer, A., Büscher, D., Vertens, J., Luft, L., Burgard, W., 2021. Long-term vehicle localization in urban environments based on pole landmarks extracted from 3-D lidar scans. *Robotics and Autonomous Systems*, 136, 103709.

Schaupp, L., Pfreundschuh, P., Burki, M., Cadena, C., Siegwart, R., Nieto, J., 2020. Mozard: Multi-modal localization for autonomous vehicles in urban outdoor environments. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 4828–4833.

Schwab, B., Wysocki, O., 2021. Save - lod3 road space models.

Siciliano, B., Khatib, O., Groen, F., Stachniss, C., 2009. *Robotic Mapping and Exploration*. 55, Springer Berlin Heidelberg, Berlin, Heidelberg.

Tian, R., Zhang, Y., Feng, Y., Yang, L., Cao, Z., Coleman, S., Kerr, D., 2021. Accurate and robust object-oriented slam with 3d quadric landmark construction in outdoor environment.

Unel, F. O., Ozkalayci, B. O., Cigla, C., 2019. The power of tiling for small object detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 582–591.

Weng, L., Gouet-Brunet, V., Soheilian, B., 2021. Semantic Signatures for Large-scale Visual Localization. *Multimedia Tools and Applications*, 80(15), 22347–22372. http://arxiv.org/pdf/2005.03388v1.

Yu, X., Chaturvedi, S., Feng, C., Taguchi, Y., Lee, T.-Y., Fernandes, C., Ramalingam, S., 2018. Vlase: Vehicle localization by aggregating semantic edges. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 3196–3203.

Zekavat, S., Buehrer, R. M., Durgin, G. D., Lovisolo, L., Wang, Z., Goh, S. T., Ghasemi, A., 2021. An Overview on Position Location: Past, Present, Future. *International Journal of Wireless Information Networks*, 28(1), 45–76.

Zhang, Q., Zhang, H., Lu, X., 2022. Adaptive Feature Fusion for Small Object Detection. *Applied Sciences*, 12(22), 11854.