

AN APPROACH TO EXTRACT MOVING OBJECTS FROM MLS DATA USING A VOLUMETRIC BACKGROUND REPRESENTATION

Joachim Gehrung^{ab*}, Marcus Hebel^a, Michael Arens^a, Uwe Stilla^b

^a Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB,
76275 Ettlingen, Germany - (joachim.gehrung, marcus.hebel, michael.arenst)@iosb.fraunhofer.de

^b Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80333 Muenchen, Germany - stilla@tum.de

KEY WORDS: Volumetric Representation, Background Subtraction, Detection and Tracking of Mobile Objects, Change Detection

ABSTRACT:

Data recorded by mobile LiDAR systems (MLS) can be used for the generation and refinement of city models or for the automatic detection of long-term changes in the public road space. Since for this task only static structures are of interest, all mobile objects need to be removed. This work presents a straightforward but powerful approach to remove the subclass of moving objects. A probabilistic volumetric representation is utilized to separate MLS measurements recorded by a Velodyne HDL-64E into mobile objects and static background. The method was subjected to a quantitative and a qualitative examination using multiple datasets recorded by a mobile mapping platform. The results show that depending on the chosen octree resolution 87-95 % of the measurements are labeled correctly.

1. INTRODUCTION

Most of today's mobile mapping systems apply active laser scanners in addition to camera systems to supplement dense 3D environment data. The mobile laser scans generated by such systems are usually of high accuracy and more detailed than airborne laser scans, which makes them suitable for change detection in urban environments, as it is required for the automatic detection of changes in public road space. The task of long-term change detection benefits from the removal of all measurements associated with mobile objects (e.g. cars, persons), so that only the static structures of interest remain. A first step into that direction is to remove moving objects, which are the easiest to identify class of mobile objects. During this step, both moving objects and the remaining background should be preserved. The latter one is of further interest for the modeling process and the former can be utilized to characterize, recognize and remove non-moving instances of these mobile objects in subsequent steps.

Most techniques based on range measurements either distinguish between static structures and moving objects very crudely or extract moving objects - due to real-time constraints - in a short-term sense which may result in a non-accurate background. The approach presented in this paper uses a technique similar to background subtraction to identify range measurements associated with moving objects. Since the intended field of application does not have real-time constraints, the quantity of all measurements and therefore all available information can be utilized. The approach is also model free, so no assumptions about the environment or the moving objects (aside from the ability to be detectable by a LiDAR sensor) are required. The main contributions are the following:

- We perform a detailed analysis of the inherent capability of a probabilistic volumetric representation to discard mobile objects.
- The impact of the system parameters onto the background representation is investigated.

*Corresponding author

- Based on both insights, a straightforward and yet powerful approach with unique properties for the labeling of measurements as moving objects and static structures is proposed.

This paper is organized as follows: Section 2 provides an overview of related work. The analysis of the probabilistic volumetric representation and its system parameters is given in Section 3. The process of labeling range measurements based on the volumetric representation is given in Section 4. The proposed approach is evaluated based on the setup described in Section 5. A discussion of the results is given in Section 6. Finally, Section 7 presents our conclusion.

2. RELATED WORK

2.1 Detection of moving objects

The distinction between moving objects and static structures is a relevant topic in many fields of research. Lots of work regarding this topic is done in computer vision, especially in relation to *Object Tracking* in 2D images. Since the focus of this paper is on MLS measurements, the methodically most related topics are *Detection and Tracking of Moving Objects (DATMO)* and *Change Detection*. The former is investigated especially in robotics and autonomous driving. The latter one deals with the finding of differences between several scans of a scene.

Some approaches for change detection such as the one proposed by Girardeau-Montaut et al. (2005) and Zeibak and Filin (2008) compare point clouds of different epochs, but only work under certain conditions such as stationary sensor platforms.

In Hebel et al. (2013), an approach based on the analysis of evidential conflicts between free and occupied 3D-space is used to detect changes between two airborne LiDAR scans. Xiao et al. (2015) use a similar approach to detect moving objects in mobile LiDAR measurements based on conflict search between a reference and a target epoch. Both approaches require the object to be completely removed from the space it previously occupied and can be misled when one object is replaced by another.

Both techniques can in theory be extended to consider all available scans and therefore be able to generate the same results as the approach presented in this work. However, this would lead to quadratic runtime complexity, since each measurement in the dataset must be cross-referenced with vision rays generated from all other measurements.

Azim and Aycard (2012) detect moving objects based on inconsistencies between scans. An octree-based occupancy grid is utilized to represent the current state of the environment. A list of dynamic voxels is managed and updated whenever a conflict between a scan and the environment representation is detected. Since the approach adapts very quickly to changes in the environment, it is less suitable for the extraction of long-term stable structures. The approach could be modified by tweaking the sensor model to consider long-term stable structures, but not without missing most of the moving objects.

In *Simultaneous Localization and Mapping (SLAM)*, the elimination of moving objects can contribute to increase consistency of the environment's representation. In the approach presented by Litomisky and Bhanu (2013), clusters of point clouds of two different epochs are formed. Correspondences are established between them and based on these, moving objects are detected and removed. A high percentage of the points belonging to static structures is kept, but not all of them. Since that approach is designed for SLAM, more emphasis is put on consistency than on a complete representation of the environment.

The approach presented in this work provides a unique combination of properties. First of all, it prevents short-time stable objects from merging with the background; instead it identifies them correctly as moving objects. Most of the fore- and background is kept intact, meaning no larger chunks of measurements are lost. Furthermore, the proposed algorithm has a linear runtime complexity. All measurements only need to be touched twice; once to build the volumetric representation and a second time for determining their label.

2.2 Volumetric representation

For the representation of the background an occupancy grid has been chosen, since unlike a merely surface-based method the volumetric approach is able to describe free and unexplored space. Moravec and Elfes (1985) generate a probabilistic occupancy grid based on data from ultrasonic sensors. It describes the state of the environment along a plane on the sensor level with a probability that indicates either free or occupied space. Although the ultrasonic signal has a conical shape that covers a larger quantity of space, the resulting representation is considerably accurate. Meagher (1982) presented an approach to model arbitrary geometries that utilizes an octree to store binary occupancy information. Later the concept has been extended to use probabilistic information (Payeur et al., 1997). Hornung et al. (2013) added a nearly lossless compression strategy based on probability clamping. The latter also allows the fast adaption to a changing environment. This approach has gained huge popularity under the name *OctoMap* in the robotic community. Based on this work, we proposed a concept to represent probabilistic volumes on a global scale as well as to handle the efficient storage and retrieval of the associated data (Gehring et al., 2016).

3. VOLUMETRIC BACKGROUND MODELING

3.1 Approach to volumetric modeling

In the following, the approach for volumetric representation used in this work is briefly outlined. It is based on an approach proposed in Gehring et al. (2016) which is founded on the Bayesian framework established by Hornung et al. (2013). It was favored over a Dempster–Shafer based approach since the combination rule of the latter can lead to non-intuitive results. Also updates require more computational efforts than it is the case for the former one, which requires just a single addition.

For the global management of volumetric environment data, a grid over a global Cartesian reference frame such as ECEF is defined. Each grid cell comprises a cubic subspace and is referred to as *tile*. A tile has a unique address and contains a probabilistic occupancy octree based on the work of Hornung et al. (2013). This volumetric representation permits the deduction of information about free, occupied and unexplored space. Each voxel in the octree contains an occupancy probability which corresponds to the logarithm of a fraction that represents a probabilistic two class decision problem. This is also denoted as a *log-odd*. Due to the transformation into logarithmic space, adjustment of the log-odd can be done by simply adding or subtracting the logarithmic equivalent of a probability.

New range measurements are integrated into the system by ray-casting. For each ray the traversed voxels as well as the end voxel are determined. The log-odd of the corresponding octree nodes are adjusted accordingly; for free space l_{free} is subtracted, occupied space is handled by adding l_{occ} . Both are also referred to as sensor parameters and describe the reliability of the sensor system, more precisely its ability to measure the correct state of the observed space. To ensure fast adaption to changes in the environment, the log-odd is confined to an interval defined by the so-called clamping thresholds l_{min} and l_{max} . These thresholds are also required for compressing the octree by condensing nodes with a similar state.

3.2 Automatic removal of mobile objects

The purpose of this work is to investigate the hypothesis that a volumetric representation constructed as stated above converges to a state where it only contains non-moving objects. This state of the environment is assumed to only contain static structures and non-moving instances of mobile objects. For the sake of simplicity both are summarized by the umbrella term *background*. The idea is that a moving object occupies a point along its path only for a short duration. In terms of the volumetric representation this means the following: for all voxels occupied by the object, a probability mass corresponding to l_{occ} is added to their occupancy probability. Multiple measurements accumulate probability mass which leads to the occupancy probability converging against the upper clamping threshold, thus implying solidity. Once the object left the voxels, utilizing a 360° rotating laser scanner it is very likely that these are traversed by other range measurements. In such a case the probability mass is reduced by l_{free} for each ray passing the voxel. This leads to a convergence in direction of the lower clamping threshold. Since the moving object occupies a voxel only for a short duration, the amount of measurements traversing it outweighs the amount of measurements implying solidity. Therefore in the longterm the voxel converges to a free state, removing the former imprint of the moving object. An impression of this process is given in Figure 1.

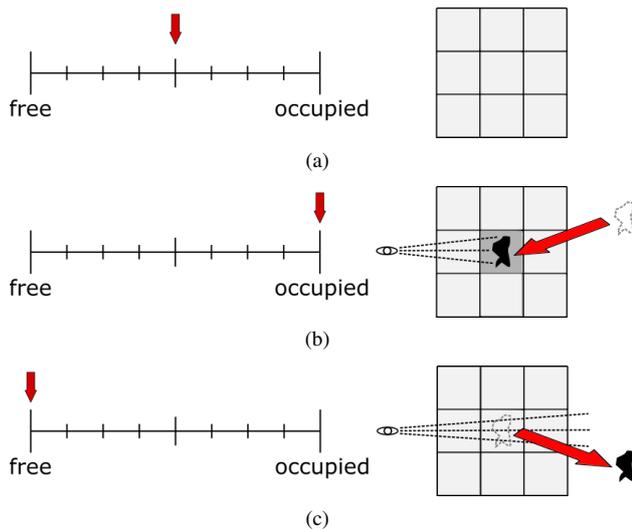


Figure 1. The influence of an object to a voxel's log-odd based occupancy probability. (a) Voxel with zero knowledge. (b) Voxel occupied by an object. (c) Voxel after the object left.

3.3 Clamping thresholds and sensor parameters

In this section the impact of both the clamping thresholds and the sensor parameters on the volumetric representation generation process is explained. The clamping thresholds l_{\min} and l_{\max} define the lower and upper boundary of an interval that limits the occupancy probability. The interval can either be symmetrical or asymmetrical (cf. Figure 2). The former considers convergence to both the states *free* and *occupied* as equal. Setting the interval to asymmetric values has the advantage of a faster convergence to either occupied or free, depending on which threshold is closer to the log-odd of $l = 0$, which implies equal distribution. Regardless of the symmetry aspect it is recommended to select both thresholds in such a way that the behavior of the function is approximately linear.

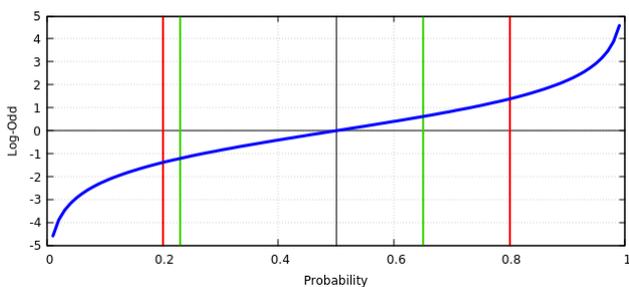


Figure 2. Symmetric (red) and asymmetric (green) clamping thresholds.

The sensor parameters l_{free} and l_{occ} can also be selected either in a symmetrical or an asymmetrical way. In the first case the convergence speed towards both states is identical. The latter case leads to faster convergence towards either free or occupied, depending on which sensor parameter is the larger one.

3.4 Minimal observation durations

Based on the parameters described in the last section and the duration per scan, the approximate time for a state change can be calculated. It should be noted that due to reasons of consistency, occupancy updates are handled based on scans, not on single

measurements. So the exact number of measurements per voxel does not matter. The period of time required to change a voxel's state from zero knowledge to occupied is calculated as

$$t_{\text{init}} = \left\lceil \frac{l_{\max}}{l_{\text{occ}}} t_{\text{dps}} \right\rceil, \quad (1)$$

where t_{dps} is the duration per scan in seconds. For a rotating laser scanner this is reciprocal to its rotation frequency. The duration for the transition from zero knowledge to free is defined as

$$t_{\text{init}} = \left\lceil \frac{l_{\min}}{l_{\text{free}}} t_{\text{dps}} \right\rceil. \quad (2)$$

The period of time required for a full state change from free to occupied is defined as

$$t_{\text{bg}} = \left\lceil \frac{l_{\max} - l_{\min}}{l_{\text{occ}}} t_{\text{dps}} \right\rceil, \quad (3)$$

whereas the inverse transition from occupied to free is calculated as

$$t_{\text{mobile}} = \left\lceil \frac{l_{\max} - l_{\min}}{-l_{\text{free}}} t_{\text{dps}} \right\rceil. \quad (4)$$

3.5 Effects of parametrization

Table 1 shows how different clamping thresholds and sensor parameters are able to influence the convergence time between the states *free*, *occupied* and *zero knowledge*. The durations are calculated using Equations 1-4. Parameter set A applies symmetrical clamping thresholds and sensor parameters. This leads to an unbiased behavior where neither the free nor the occupied state is favored. Parameter set B leads to a faster convergence towards free, both from the initial zero knowledge state as well as the occupied state. This leads to a reset behavior of voxels that contained moving objects.

The parameters defined in scenario C result in a fast initial convergence from zero knowledge towards free, whereas the convergence times between free and occupied are symmetrical. Parameter set D illustrates the effects of using both asymmetrical clamping thresholds and sensor parameters. The initial convergence from zero knowledge towards free is quite fast in contrast to the convergence towards solid. Also the duration between state changes clearly prefers the free state.

A parameter selection such as the one in A may fit most applications. Scenarios may apply asymmetrical parameters like specified in B and C in cases where initial or overall convergence towards one state is preferred. Should both be required, an asymmetrical parameter selection such as in D will be suitable.

Num.	l_{\min}	l_{\max}	l_{free}	l_{occ}	$z \rightarrow f$	$z \rightarrow o$	$f \rightarrow o$	$o \rightarrow f$
A	-2	2	-0.04	0.04	5 s	5 s	10 s	10 s
B	-2	2	-0.08	0.04	2.5 s	5 s	10 s	5 s
C	-0.5	2	-0.04	0.04	1.25 s	5 s	6.25 s	6.25 s
D	-0.5	2	-0.08	0.04	0.625 s	5 s	6.25 s	3.125 s

Table 1. Example parameters and the corresponding convergence time for a 10 Hz scan frequency. z , f and o denote the states *zero knowledge*, *free* and *occupied*.

4. MOVING OBJECT SEGMENTATION BY BACKGROUND SUBTRACTION

4.1 Background subtraction

In this section an algorithm for labeling range measurements utilizing a volumetric environment representation is proposed. It can also be considered as an experiment designed to evaluate the hypothesis presented in Section 3.2. The algorithm's input consists of a set of range measurements acquired by a mobile LiDAR sensor. These are organized as *scans*, where each scan includes multiple range measurements. A common representation for range measurements are point clouds.

The process consists of two steps that are illustrated in Figure 3 and can be compared to background subtraction as known from computer vision. First, a volumetric representation based on all scans is generated. Based on the assumption explained above this representation should converge to a state where only non-moving instances of mobile objects and static structures are left. In a second step, the same scans are labeled using the volumetric representation. For each range measurement the voxel corresponding to its endpoint is determined. Based on the occupancy probability of the voxel in question the following cases can be distinguished:

- If the occupancy probability is larger or equal to an upper clamping threshold, the measurement belongs to a solid object.
- If it is less or equal to a lower clamping threshold, the measurement is part of a moving object.
- If none of the preceding cases occurs, the state of the measurement is undetermined.

4.2 Parameter selection

As explained in Section 3.3, the clamping thresholds and sensor parameters have a direct impact on the convergence of the volumetric representation and therefore on the labeling results. Therefore it is recommended to utilize the equations presented in Section 3.4 to select the parameters based on whatever duration seems plausible for the case at hand. The octree resolution has to be adapted to the accuracy and precision of the sensor. If the voxels are too small, minor deviations of the range measurements endpoint can lead to inaccuracies of the resulting representation. The size of the voxels will also have influence on the time required to generate the background representation and the discretization errors. Restricting the maximum length of range measurements is also worth considering, since in this way negative effects of inaccuracies in geolocalization and sensor calibration can be reduced. Limiting the ray length is a compromise between information loss and the improvement of the background representation and highly depends on the deployed sensor system. It is recommended to use another set of clamping thresholds for the labeling step. If both thresholds are closer to $l = 0$, more points are labeled either as belonging to a moving object or to background and therefore less points will end up as undetermined.

5. EVALUATION

5.1 Experimental setup

The presented approach has been evaluated with MLS measurements taken on the campus of the TU München and along public

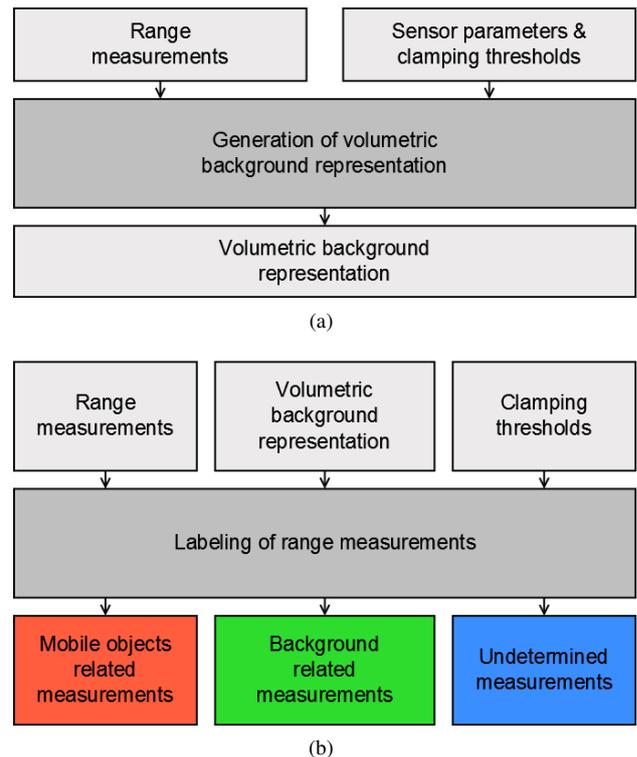


Figure 3. Illustration of the two-stage process for labeling range measurements. (a) Generation of the background representation. (b) Labeling of range measurements.

buildings in the vicinity. The *MLS 1 - TUM City Campus* dataset consists of georeferenced MLS scans and is publicly available¹ under a Creative Commons License. An overview of the dataset is given in Figure 4(a). Measurements have been carried out with the mobile mapping system *MODISSA* of Fraunhofer IOSB. The vehicle itself is equipped with four LiDAR sensors. Two Velodyne HDL-64E are mounted at an angle of 35° on the front roof. This setup makes it possible to cover both a part of the road ahead of the vehicle as well as the facades of the surrounding buildings. Two additional Velodyne VLP-16 Puck are installed on the back of the roof and are inclined by an angle of 15° . Eight cameras mounted at the roof corners can be utilized to texture the LiDAR measurements. Georeferencing of the measurements is based on navigation data provided by an Applanix POS LV navigation system with two GNSS antennas, an inertial measuring unit and a distance measuring indicator. The navigation data has been post-processed to increase accuracy.

In this paper, the provided dataset is divided into two subsets that are used to evaluate different facets of the proposed approach. The first subset is referred to as *Inner Yard* and shows a busy inner courtyard on the campus. The sequence has a duration of about 4 minutes and contains multiple groups of people standing or sitting as well as walking around. Also included are several trees with dense foliage whose leaves are moving in the wind. The second subset denoted as *Crossing* comprises a traffic situation which the mobile mapping platform waits at a crossing. Multiple cars pass by, passengers and a cyclist cross the road. After 22 seconds the platform continues on its way. Both subsets can be seen in Figure 4(b) and 4(c).

¹<http://s.fhg.de/mls1>

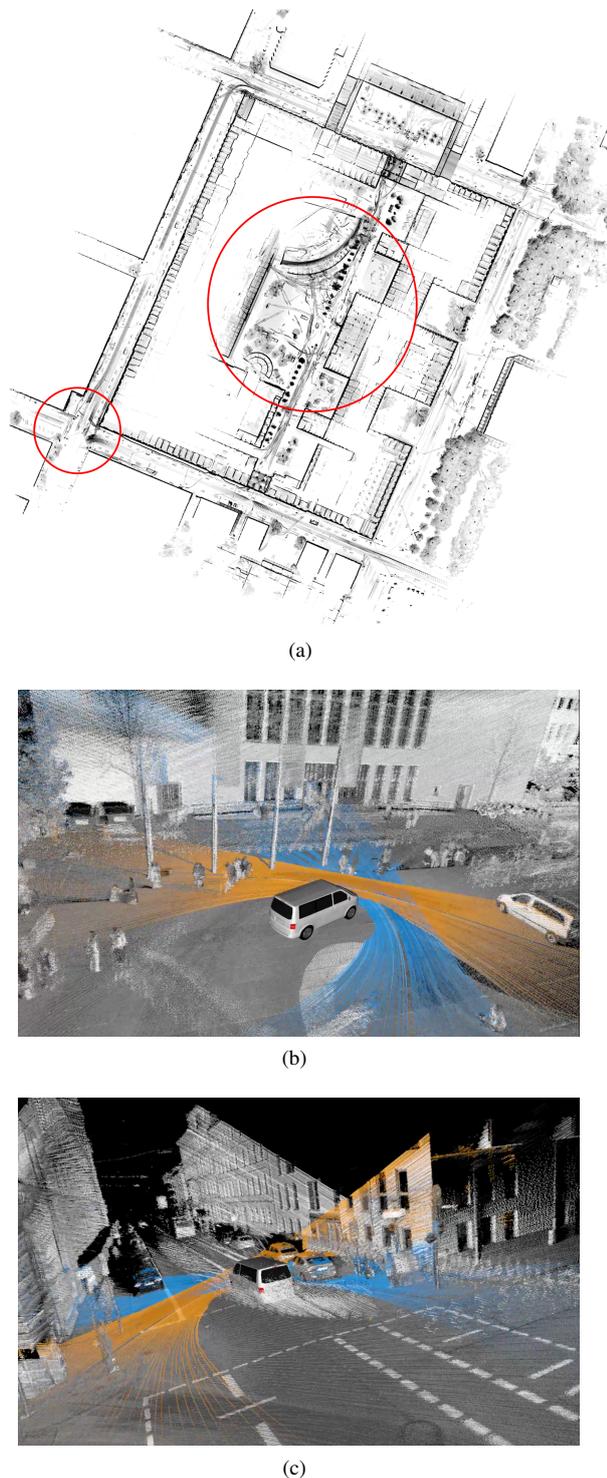


Figure 4. The TUM City Campus dataset. (a) Overview. (b) Inner Yard. (c) Crossing.

5.2 Quantitative examination

The results of the approach are examined by comparing its results with a ground truth that comprises per point labels. Since every LiDAR scan consists of about 100000 points, the process of labeling is very time consuming. Therefore an automated approach has been used for ground truth generation. A bounding box has been placed within the *Crossing* subset; all points within the box are labeled as belonging to a moving object, all outside

are part of the static environment. This easy approach works quite well, since the scene contains both static and moving objects, but structured in a way that both are easy to divide. This resulting sequence of ground truth data has a duration of 22 seconds.

5.3 Qualitative examination

The quantitative examination is supplemented by a visual inspection by a human observer. The results are examined based on the following criteria:

- Has each moving object at least some measurements labeled correctly?
- How high is the percentage of correctly labeled measurements per moving object?
- What is the false positive and false negative rate?
- How many percent of all measurements remain undetermined?

These questions are answered based on the *Inner Yard* subset. The labeled results are inspected scan by scan, whereby color coding allows the easy interpretation of the labeled results. In addition to the questions above, the results are inspected for artefacts and other discrepancies.

5.4 Comparison with other approaches

For a more in-depth evaluation of the proposed approach, it is compared with other techniques for the moving object detection. The goal is to work out the advantages and disadvantages between them. A sliding window approach similar to the one presented by Azim and Aycard (2012) is simulated by building volumetric representations for sequences of 10 second length. The corresponding scans are labeled and visually compared to the labels created within the scope of the *Inner Yard* subset.

To compare our work to a per scan based technique, the approach proposed by Zeibak and Filin (2008) has been chosen. This approach compares two depth images created by a stationary LiDAR sensor. For comparison the *Crossing* subset has been selected, since at the beginning of the sequence the car is stationary for about 20 seconds.

5.5 Runtime

Since the time required for labeling the range measurements highly depends on the environment and the density of the scans, the runtime regarding this step has been further investigated based on both subsets. More information regarding the generation runtime of the volumetric representation can be found in our other work (Gehring et al., 2016).

6. RESULTS AND DISCUSSION

6.1 Results of the quantitative examination

For the generation of the volumetric representation the clamping thresholds and sensor parameters were chosen symmetrically and in a way that the convergence time between states is about 8 seconds; this has turned out to be good practice for the given scene and sensor setup. Since the most critical parameter in regards to the quality of the results is the octree resolution, the

examination has been executed for multiple octree resolutions (caused by selecting multiple depth levels). Table 2 shows the results of quantitative examination - the rates of correct classification, type I and type II errors. It can be seen that a high resolution favors also a high rate of correct classifications, but is also correlated to both type I and type II errors.

Rate	10 cm	20 cm	40 cm	80 cm	160 cm
True Positive Rate	97.2 %	95.2 %	87.1 %	74.3 %	66.9 %
True Negative Rate	71.3 %	86.4 %	94.8 %	98.2 %	99.1 %
False Positive Rate	28.7 %	13.6 %	5.2 %	1.8 %	0.9 %
False Negative Rate	2.8 %	4.8 %	12.9 %	25.7 %	33.1 %

Table 2. The rates of correct classification, type I and type II errors for multiple octree resolutions, calculated based on the results of the quantitative evaluation.

The corresponding ROC curve in Figure 5 indicates a working point at either a 20 cm or a 40 cm resolution, both with a good trade-off between true positive rate and false positive rate. For the qualitative examination in the next section the latter has been chosen.

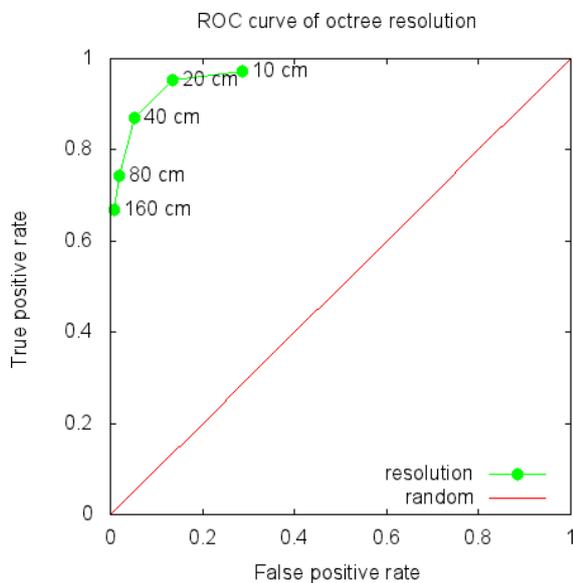


Figure 5. ROC curve describing the evaluation results depending on the resolution of the octree.

6.2 Results of the qualitative examination

The parameters for the generation of the volumetric representation were chosen as suggested in the last section. The clamping parameters for labeling have been set to $l_{\min} = -0.5$ and $l_{\max} = 0.5$ to decrease the number of points labeled as undetermined. An impression of the labeled results can be found in Figure 6.

Coverage of all moving objects – All moving objects were manually identified by comparing multiple scans in chronological order. In summary, it can be stated that surface points have been detected on all moving objects. Several times it has been observed that persons that stopped for a few seconds were still labeled as a moving object. This can be considered as a clear benefit in comparison to the sliding window approach.

Coverage per moving object – All moving objects were closely examined by a human observer. The number of correctly labeled

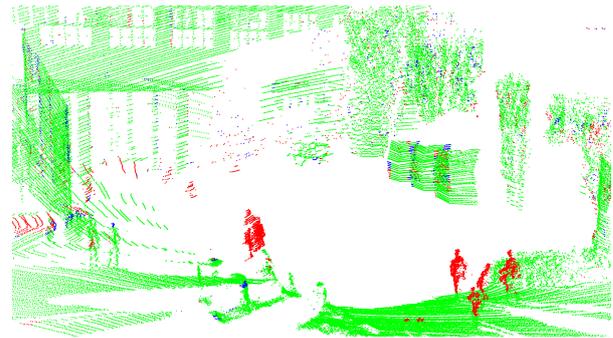


Figure 6. A labeled scan from the *Inner Yard* subset. The point clouds are divided into the categories *moving object* (red), *static structure* (green) and *undetermined* (blue).

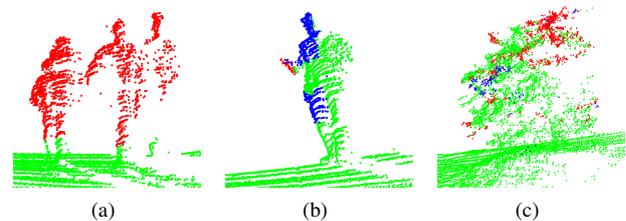


Figure 7. Examples for moving objects and vegetation. Moving object (red), static structure (green) and undetermined (blue). (a) Discretization error. (b) Labeling error due to partial movement. (c) Tree with leaf movement.

measurements corresponds to the rates determined by the quantitative examination. The rest is usually considered to be static and is located within the lower part of the object. The reason for this are discretization errors caused by overlaps of the object with volumes containing the ground. In other, less frequently observed cases larger block-shaped parts were labeled incorrectly as being uncertain or solid. This happens usually when a person stands still for most of the time and only moves occasionally. See also Figure 7(a) – 7(c).

False positive and false negative rate – False negatives are measurements of moving objects that are labeled incorrectly to be part of a static object or structure. This was mostly observed in the case of discretization errors and partly moving objects as discussed above.

False positives include all measurements of static structures and objects that are incorrectly labeled as belonging to a moving object. These appear in multiple cases. Both grid-like structures and leaves are sometimes hit by a laser, sometimes not (see Figure 7(c)). This leads to conflicting measurements which cause the corresponding voxels to converge towards the state of being free. This in turn marks all measurements ending within the voxel as being part of a moving object. False positives can also happen because of sensor noise. The latter leads to incorrect labeled points that are distributed in a non-systematic way along the surfaces of static background. That effect appears to become stronger the smaller the voxels are.

Multiple times in both subsets huge spots of false positives were observed, mostly on the ground or on walls (compare Figure 8). Their size varies between half a meter and multiple meters. Due to the clustering of affected voxels in close proximity to each other a systematic cause for this artefact is assumed. A possible explanation is that this effect is generated by measurement rays

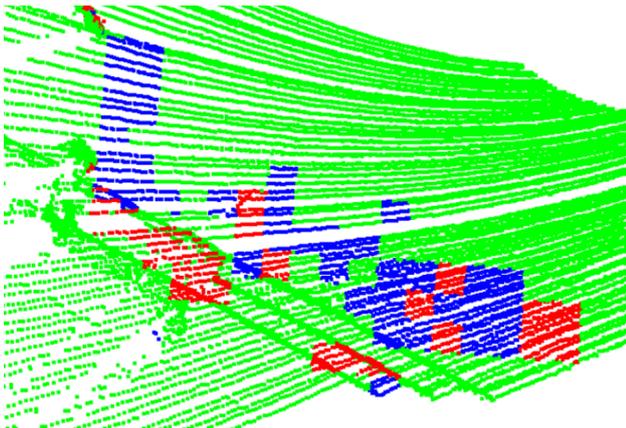


Figure 8. Solid ground (green) labeled as undetermined (blue) or moving (red), possibly due to measurements under unfavorable incidence angle.

that pass along the said plane in a relatively flat angle, thereby traversing voxels belonging to the plane (cf. Figure 9). As a consequence, the state of these voxels is changed step by step until they are considered to be free. A possible explanation for the different degrees of belief - visible in form of different labels in Figure 8 - could be measurement noise or sensor calibration errors.

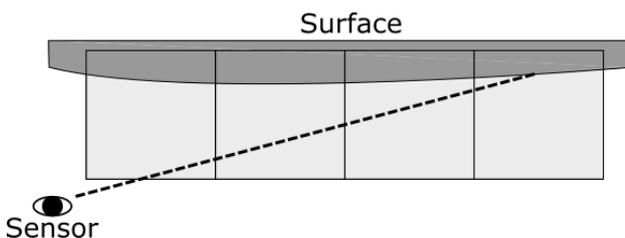


Figure 9. A ray may traverse other voxels assumed to be solid before hitting a surface, thereby successively changing their state to free.

Rate of undetermined measurements – One of the two states can only be assigned to a measurement once its occupancy probability is beyond the corresponding clamping threshold. As described above these have been adapted to reduce the number of undetermined measurements. The result of this action is that only 5 % of all measurements cannot be labeled as either being static or part of a moving object.

6.3 Differences to other approaches

Sliding window – The labeling results of the sliding window based approach are comparable with the ones described in Section 6.2. Also present is the false positive artefact, which has the same characteristics. Since the window has a length of only 10 seconds, surface points of slowly moving or deforming objects like flags are associated mostly with incorrect static labels. It appears also to be less likely to label measurement points on a person that has stopped for a few seconds correctly, especially if the action of stopping extends over several windows. The results show that a sliding window based approach has only a subset of the capability of an approach that considers all available measurements.

Scan based comparison – In case of a stationary sensor with a high scan rate, the successive scan based approach tends to generate false negatives. This is due to the fact that the difference in

the 2.5D depth image is too small to be considered as a movement of the corresponding object. In theory this could be handled by carefully adapting the threshold, in practice the sensor noise limits these efforts to some lower boundary. This is an inherent problem of an approach that is based on the comparison of successive scans; one that the approach presented in this paper does not have.

Another difference between both approaches is the ability to handle sensor movement. Depending on the parametrization of the scan comparison based approach there are either lots of false positives or lots of false negatives, since it only works based on 2.5D information. The approach proposed in this work utilizes geolocated full 3D measurements and is therefore able to handle also a moving platform, although the associated noise due to localization errors effect the results in a negative way.

6.4 Runtime

Subset	Resolution	Runtime
Inner Yard	20 cm	1000-2200 ms
Inner Yard	40 cm	800-1500 ms
Crossing	20 cm	900-1300 ms
Crossing	40 cm	900-1100 ms

Table 3. The runtime for the *Inner Yard* and *Crossing* subset, determined with different voxel resolutions.

The runtime for the labeling process has been determined for both subsets and with different voxel sizes. The results can be seen in Table 3. Since not all scans contain the same number of measurements, the runtimes are stated as an interval. It can be assumed that the upper runtime boundary is also affected by reloading operations of volumetric data from the hard drive if the former one has not been found in the in-memory cache. The results show that while the labeling is executed quite fast it would not be able to process the measurements from a Velodyne LiDAR sensor in realtime, since this would require a runtime of less that 100 ms.

7. CONCLUSION AND FUTURE WORK

We have presented a straightforward, yet powerful approach to distinguish measurements of a mobile LiDAR sensor into moving objects and static background. It relies on the hypothesis that a probabilistic volumetric representation converges against a point where it is free of moving objects and therefore can be used for background subtraction. An evaluation of the approach consisting of both a quantitative and a qualitative part based on a real-world dataset proved the hypothesis, but also showed special cases such as vegetation that need to be handled separately. For most moving objects 87-95 % of all measurements are labeled correctly; all moving objects were provided with labels. The amount of measurements which can not be associated with either background or moving objects is less than 5 %. It has been shown that even moving objects stopping for some time are still labeled correctly, since the full time horizon available is considered.

In the future, we plan to further investigate solutions for the artefact that causes larger spots of static background to be labeled as moving object. It is also planned to apply the approach for point cloud labeling within the context of change detection.

References

- Azim, A. and Aycard, O., 2012. Detection, classification and tracking of moving objects in a 3d environment. In: *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, pp. 802–807.
- Gehring, J., Hebel, M., Arens, M. and Stilla, U., 2016. A framework for voxel-based global scale modeling of urban environments. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W1, pp. 45–51.
- Girardeau-Montaut, D., Roux, M., Marc, R. and Thibault, G., 2005. Change detection on point cloud data acquired with a ground laser scanner. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36-3/W19, pp. 30–35.
- Hebel, M., Arens, M. and Stilla, U., 2013. Change detection in urban areas by object-based analysis and on-the-fly comparison of multi-view ALS data. *ISPRS Journal of Photogrammetry and Remote Sensing* 86, pp. 52–64.
- Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C. and Burgard, W., 2013. Octomap: An efficient probabilistic 3d mapping framework based on octrees. *Autonomous Robots* 34(3), pp. 189–206.
- Litomisky, K. and Bhanu, B., 2013. *Removing Moving Objects from Point Cloud Scenes*. Vol. 7854, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 50–58.
- Meagher, D., 1982. Geometric modeling using octree encoding. *Computer Graphics and Image Processing* 19(2), pp. 129–147.
- Moravec, H. and Elfes, A., 1985. High resolution maps from wide angle sonar. In: *Proceedings of the 1985 IEEE International Conference on Robotics and Automation*, pp. 116 – 121.
- Payeur, P., Hebert, P., Laurendeau, D. and Gosselin, C. M., 1997. Probabilistic octree modeling of a 3d dynamic environment. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 2, pp. 1289–1296.
- Xiao, W., Vallet, B., Brédif, M. and Paparoditis, N., 2015. Street environment change detection from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 38, pp. 38–49.
- Zeibak, R. and Filin, S., 2008. Change detection via terrestrial laser scanning. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, pp. 430–435.