# REAL-TIME FOREGROUND SEGMENTATION FOR SURVEILLANCE APPLICATIONS IN NRCS LIDAR SEQUENCES

Lóránt Kovács[1,2,*], Marcell Kégl[1], and Csaba Benedek[1,2]

[1] Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network, 1111 Budapest, Kende utca 13-17.
[2] Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, 1083 Budapest, Práter utca 50/A.
(kovacs.lorant, keglmarcell, benedek.csaba)@sztaki.hu

**Commission I, WG I/2**

**KEY WORDS:** Lidar, non-repetitive circular scanning, foreground segmentation, background model, surveillance

## ABSTRACT:

In this paper, we propose a point-level foreground-background separation technique for the segmentation of measurement sequences of a Non-repetitive Circular Scanning (NRCS) Lidar sensor, which is used as a 3D surveillance camera mounted in a fixed position. We show that by applying the NRCS Lidar technology, we can overcome various limitations of rotating multi-beam Lidar sensors, such as low vertical measurement resolution, which is disadvantageous in surveillance applications. As the main challenge, we need to efficiently balance between the spatial and the temporal resolution of the recorded range data. For this reason, we automatically generate and maintain a very high-resolution background model of the sensor's Field of View, while for enabling real-time analysis of dynamic objects we use low integration time to extract the consecutive time frames. As a result, the laser reflections from foreground objects reflect sparse, but geometrically accurate samples of the silhouettes providing valuable input for higher-level shape description or event analysis steps. We demonstrate the efficiency of the new approach in different realistic NRCS Lidar measurements sequences, obtaining a 0.76 overall F1-score on the measured dataset.

## 1. INTRODUCTION

Accurate and real-time foreground-background separation is a critical task in surveillance applications. As alternative solutions of conventional optical video cameras, range sensors offer significant advantages for scene analysis, since direct geometrical information is provided by them (Börcs et al., 2017). The use of infrared light based Time-of-Flight (ToF) cameras (Schiller and Koch, 2011) or laser-based Light Detection and Ranging (Lidar) sensors (Kaestner et al., 2010) enables recording directly measured range images, where we can avoid artefacts of the stereo vision based depth map calculation.

From the point of view of data analysis, ToF cameras record depth image sequences over a regular 2D pixel lattice, where established image processing approaches, such as morphological filters or Markov Random Fields (MRFs) can be adopted for smooth and observation consistent segmentation and recognition (Benedek et al., 2013). However, such cameras can only be reliably used indoors, due to the limitations of current infra-based sensing technologies, and they may have a narrow Field of View (FoV), which fact can be a drawback for surveillance and monitoring applications.

By extracting accurate 2D or 3D object silhouettes, one can obtain various sorts of valuable scene information which can be directly exploited in among others people detection, tracking, biometric recognition, or activity analysis.

Prior existing Lidar-based surveillance solutions utilize mainly Rotating Multi-Beam (RMB) Lidar sensors (Alkhalili et al., 2019). These systems can capture point cloud sequences of the full 360°view with a recording frequency of $15-30$ fps. The RMB Lidar's vertical resolution is determined and fixed by the number of the laser beams, while the horizontal resolution depends on the speed of the sensor rotation. Each laser point of the output point cloud is associated with 3D spatial coordinates, and possibly with auxiliary channels such as reflection number or an intensity value of laser reflection. RMB Lidars can produce high frame-rate *point cloud videos* enabling dynamic event analysis in the 3D space. On the other hand, the measurements have low spatial density, which quickly decreases as a function of the distance from the sensor, and the point clouds may exhibit particular ring patterns typical of the sensor characteristics.

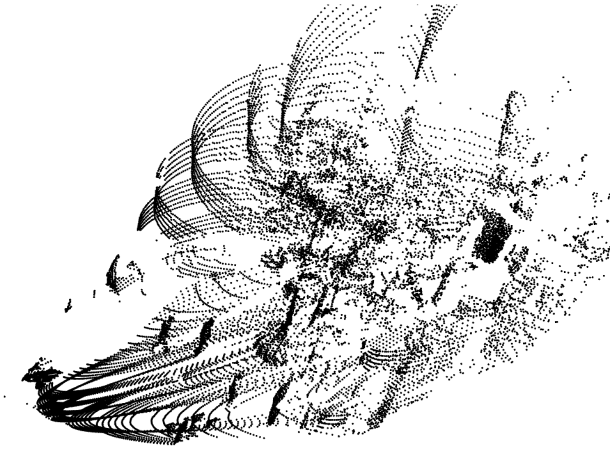


Figure 1. Point cloud recording from the Courtyard dataset, recorded using the Livox Avia sensor with its Non-repetitive Circular Scanning (NRCS) technique

While previous works have shown (Benedek et al., 2018), that RMB Lidar measurements can be used for certain dynamic scene analysis tasks, such as object separation, tracking and

[*] Corresponding author

https://doi.org/10.5194/isprs-archives-XLIII-B1-2022-45-2022

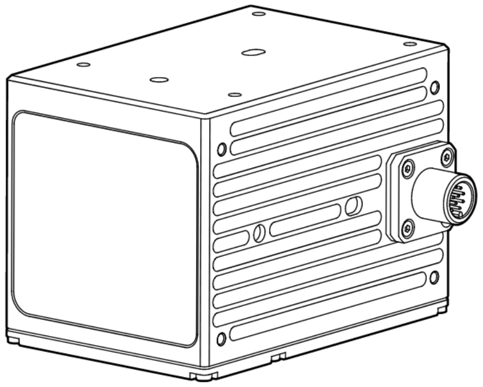


Figure 2. Schematic of the used Livox Avia Lidar sensor. The window on the left covers the sensor. (Avia User Manual, https://www.livoxtech.com, 2021)

even gait-based biometric person recognition and activity analysis, the constant and low vertical resolution of the measurements that is physically constrained by the number of vertically fixed laser emitters and receivers (typically 32 or 64), means a clear limitation by applying them in a static sensor configuration. Moreover, the RMB Lidar sensors are generally expensive, thus their application is not widespread for surveillance tasks.

An alternative to the RMB Lidars is a new type of Lidar sensor shown in Fig. 2, which is called Livox Avia (Li et al., 2021) and it implements a unique Non-repetitive Circular Scanning (NRCS) technique. This sensor uses a multi-line laser combined with high-speed scanning on a circular path, which results in a point cloud data capturing rate of up to 240,000 points/s.

The Livox Avia sensor has six Lidar beams organized in a linear beam array, which is moved and rotated inside the sensor to cover its Field of View (FoV): horizontal $70.4^\circ$ and vertical $77.2^\circ$ FoV (See in Fig. 3), and $0.05^\circ$ angular precision. The non-repetitive scanning technology is used to improve the static scanning effect, and it also increases the coverage area ratio and improves the detection of objects and details within the FoV. As demonstrated in (Lin and Zhang, 2020, Wang et al., 2021), this NRCS approach is suitable for the majority of use case scenarios including traditional mapping and low-speed autonomous driving.

The NRCS Lidar sensor is capable of providing measurements for real-time scene analysis, while the sensor is available on the market at affordable prices compared to the other Lidar technologies (Glennie and Hartzell, 2020). The sensor continuously records distance measurements with corresponding timestamps following its non-repetitive circular pattern in its field of view. Here, by setting fixed integration time, the consecutively collected points can be grouped into separate Lidar time frames. The main challenge is to efficiently balance between the spatial and the temporal resolution of the recorded range data. While allowing larger integration time, the laser beams cover a higher proportion of the FoV yielding high spatial measurement resolution of the measurement frame, the object movements of dynamic objects in the observation area induce various artefacts (e.g., blurred pedestrian silhouettes), which do not allow efficient dynamic event analysis. For example, the Livox Avia sensor collects $240,000$ points within a time-window of $1s$, while $720,000$ points in a $3s$-window. On the other hand, if the measurements are collected in a narrow time window (e.g., in 100 ms) the resulting point clouds are very sparse, which phenomenon yields a loss of details across the spatial dimension of the FoV: a sample frame of 24,000 points is shown in Fig. 1.

For the above reasons, in the proposed approach we generate and maintain a very high-resolution (HR) background model of the scene fully automatically in the range image domain of the sensor's FoV, while for enabling real-time analysis of dynamic objects we use low integration time to extract the consecutive time frames. The measured points are matched to the high-resolution background model components in the closest matching positions. This process ensures that the spatial accuracy of the native measurements is largely maintained, instead of applying a rough spatial downscaling technique. As a result, we can obtain sparse, but geometrically accurate point cloud segments representing the moving objects, which can be used in higher-level scene analysis steps of surveillance systems.

This paper presents a new point-level foreground-background separation method by processing measurement sequences of a (NRCS) Lidar sensor, which is used as a surveillance sensor mounted in a fixed position. The outline of the paper is as follows. The steps of the proposed approach are detailed in Sec. 2. Sec. 3 describes our new annotated dataset created for testing the algorithm. For this purpose, two different measurement sequences have been recorded by a Livox Avia sensor, in a controlled *Courtyard* environment, and in a more challenging *City Center* scene, respectively. In Sec. 4 we describe and analyse the quantitative and qualitative evaluation results. Finally, concluding remarks and future plans are given in Sec.5.

## 2. PROPOSED METHOD

The goal of the proposed method is to separat foreground and background regions in Lidar frames extracted with a narrow integration window (used 100ms) from a measurement sequence of a static NRCS Lidar sensor.

Formally, in a given time frame $t$, we assign to each point $p \in \mathcal{L}^t$ a label $\omega(p) \in \{\mathrm{fg}, \mathrm{bg}\}$ corresponding to the moving object (i.e. foreground, fg) or background classes (bg), respectively.

The sensor's non-repetitive circular scanning approach implies a critical challenge to be handled: the moving laser beams cannot densely cover the whole field of view within the considered data collection window, which results in several sparse/empty regions in the individual Lidar frames. Moreover, we can observe strongly inhomogeneous point density as shown in Fig. 1.

Surveillance applications demand real-time solutions. To avoid computationally expensive algorithmic steps in the 3D point cloud domain, and to enable the efficient and robust utilization of the sparse data, we map the problem to the 2D range image domain, by transforming the 3D Euclidean point coordinates into a polar representation.

The proposed method consists of three main steps, as follows:

1. Incoming Lidar measurements are collected within a 100ms time window for composing the next point cloud frame of the sequence. Thereafter, the distances of the 3D measurement points from the sensor are assigned to corresponding pixels in a high-resolution range image.

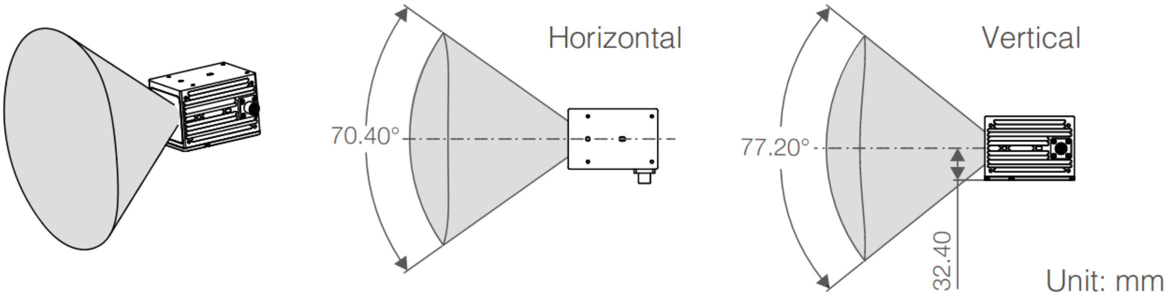


Figure 3. Field of view of the Livox Avia Lidar sensor (Avia User Manual, https://www.livoxtech.com, 2021)

2. A local background (Bg) model is assigned and maintained for each pixel of the range image lattice, following the Mixture of Gaussians (MoG) approach (Stauffer and Grimson, 2000) applied for the range values. Considering the sparseness of the captured point clouds, in a given time frame, only the MoG Bg model components of range image pixels linked to the actual measurement points are updated. The incoming measurement points are classified either as foreground or as background -based on matching the measured range values to the local MoG distributions.
3. False foreground points in dynamic background regions (e.g. by moving vegetation) are filtered out by using an extension of the original MoG approach. To ensure compact shapes for the extracted moving objects fast spatial filters are adopted for segmentation refinement.

The above steps are detailed in the following subsections one after another.

### 2.1 Range image formation

The point cloud's representation is transformed from the 3D Descartes to a spherical polar coordinate system. A 2D pixel lattice is generated by quantizing the horizontal and vertical FoV-s, and each 3D point's distance from the sensor is stored in a pixel determined by the corresponding azimuth and elevation values. The polar direction and azimuth angles correspond to the horizontal and vertical pixel coordinates, and the distance is encoded in the corresponding pixel's 'gray' value. As a result, the upcoming steps of the proposed foreground segmentation method can be developed in the 2D range image domain.

Using a narrow timing window the range image of a certain frame contains several pixels with undefined range values as a consequence of the NRCS scanning technology. The number of undefined pixels depends on both the timing window and the predefined size of the range image. In our experiments, exploiting the precision parameters of the used Livox AVIA sensor, its FoV is mapped onto a $600 \times 660$ sized pixel lattice, resulting in an $8.5\text{px}/^\circ$ spatial resolution. We also have to consider that the density of the recorded valid range values is decreasing towards the peripheral regions of the range image due to the applied scanning technique: the scanning pattern crosses the optical center of the sensor more frequently than covering the regions of the FoV's perimeter. The sparseness of the range image makes it significantly more difficult to perform e.g. object-based foreground-background segmentation.

### 2.2 Background model

The scene's estimated background is represented in the 2D range image domain defined in Sec. 2.1.

Our background modeling technique is based on (Benedek et al., 2013), which extends the Mixture of Gaussians (MoG) approach (Stauffer and Grimson, 2000) to the range image domain. A fitness term $f_{\text{bg}}(p)$ is assigned to each point $p \in \mathcal{L}^t$ of the cloud, which measures the quality of the hypothesis that $p$ is a background point. As explained in Sec. 2.1, we map the points to the range image pixels, where we use the predefined and fixed sized 2D pixel lattice. For each $s$ cell of $S^{\text{bg}}$, we calculate an MoG approximation of the $d(p)$ distance histogram of $p$ points being projected to $s$. Following the approach of (Kaestner et al., 2010), we use a fixed 5 number of components with weight $w_s^i$, mean $\mu_s^i$, and standard deviation $\sigma_s^i$ parameters, $i = 1 \dots 5$. Thereafter the weights are sorted in decreasing order, and the minimal $k_s$ number is determined, which satisfies

$$\sum_{i=1}^{k_s} w_s^i > T_{\text{bg}}, \tag{1}$$

where we used $T_{\text{bg}} = 0.89$.

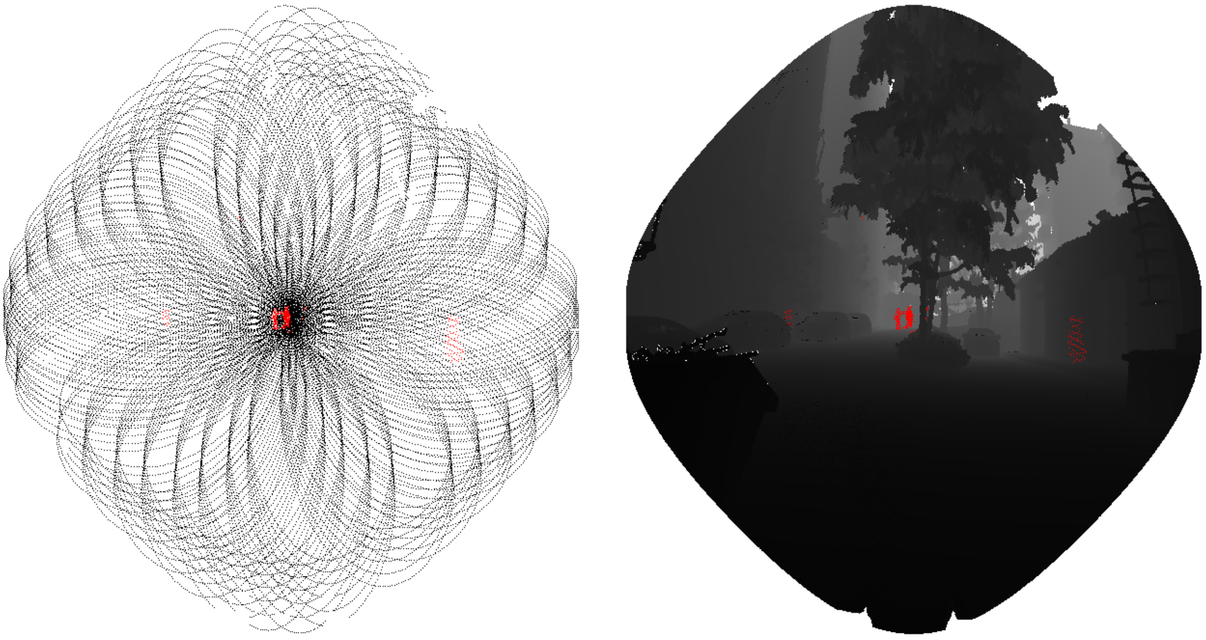
We consider the components with the $k_s$ largest weights as the background components. Then, denoting by $\eta()$ a Gaussian density function, and by $\mathcal{P}^{\text{bg}}$ the projection transform onto $S^{\text{bg}}$, the $f_{\text{bg}}(p)$ background evidence term is obtained as:

$$f_{\text{bg}}(p) = \sum_{i=1}^{k_s} w_s^i \cdot \eta\left(d(p), \mu_s^i, \sigma_s^i\right), \text{ where } s = \mathcal{P}^{\text{bg}}(p). \tag{2}$$

The Gaussian mixture parameters are calculated and refreshed based on (Stauffer and Grimson, 2000). By thresholding $f_{\text{bg}}(p)$, we can get a dense foreground/background labeling of the point cloud (Kaestner et al., 2010, Stauffer and Grimson, 2000).

As the incoming points from the consecutive sparse NRCS Lidar frames are processed one after another, each pixel of the HR background range image lattice becomes covered by valid range measurement several times, thus the associated MoG distribution can learn the appropriate parameters. The used background model is adaptive, thus it automatically updates itself when the background scene changes: for example, a static object is relocated, or a parking car departs. Besides updating the high-resolution background nap, the method also classifies the incoming frame's individual points whether they belong to the foreground or the background classes.

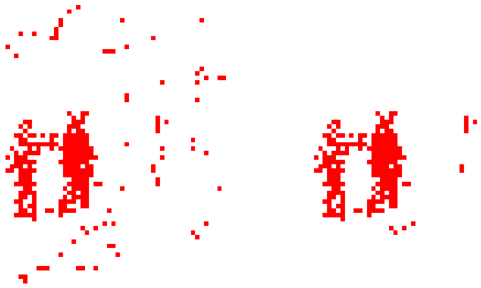
Although the MoG technique is regarded as a highly robust approach for optical video processing, as demonstrated in Fig.



(a) Detected foreground (by red) in a singel time frame of the NRCS Lidar image sequence

(b) Detected foreground region (by red) displayed over the generated high-resolution background range image

Figure 4. Foreground detection results (by red) in the *City Center* scene, displayed in 3D point cloud representation.



(a) Foreground detection without the spatial filtering adjustment in the central area of Fig. 4a

(b) Foreground detection result with the proposed method in the central area of Fig. 4a

Figure 5. Foreground detection results (by red) in the *Courtyard* scene, displayed in range image representation.

4(b), the above described Fg-Bg classification process is notably noisy for NRCS Lidar-based range image sequences, especially in scenarios recorded in large outdoor environments. Various sources of noise are present, including oscillations and small movements in the background (tree leaves, branches), whose regions are often classified falsely as foreground. Although by fine-tuning the parameters of the algorithm, the negative effects of oscillations can be decreased, usually these artefacts cannot be eliminated in acceptable quality. As a consequence, to reliably eliminate the oscillation artefacts, further noise filtering steps are needed, as described in the next subsection.

As for the speed of adaption, the initialization period of the method in a new scene needs about 50-100 time frames, to obtain an efficient initial background range value for each pixel of the HR background map. Additional 100-300 frames are required to let the background model's MoG distributions parameters converge, exploiting the repetitive sensor measurements from the observed background scene.

### 2.3 Foreground noise filtering

In this section, we propose filtering steps applied to the Mixture of Gaussians-based segmentation output, to obtain a smoothly uniform and observation consistent segmentation of the point cloud sequence recorded by the NRCS Lidar.

Let us observe that vibrations of objects (e.g. tree leaves, branches) in the background area are usually composed of relatively small, but frequent movements. The vibrating objects' edge points often oscillate between neighboring pixels of the range image lattice, causing challenges for the original MoG approach. As the background oscillations are often quasi-periodic, we can frequently observe for the pixels of these areas two high-weight Gaussian components, thus based on the thresholding rule of eq. (1) these regions receive in majority of background labels. However, there are regions in the observed area where real foreground objects (persons, cars, etc.) often pass. This frequent occurrence of an object in a typical distance also means, that the method will store this distance in the model, in the second component as well, while the first component will contain the real background distance value with the highest weight. In order not to be misguided for these real foreground points, we apply an additional filtering condition: if the deviation of the highest weight Gaussian component is saliently small (which indicates a compact background surface), we do not allow to include further Gaussian components in the local background model.

Since the above described MoG-based method works independently on each pixel of the range image, noise may result in many standalone false foreground pixels surrounded by background regions, which can be removed by morphological filtering operations.

As a result, the number of false-positive foreground pixels can be significantly decreased (see Fig. 5), and we can obtain compact and largely connected object shapes as shown in Fig. 6.

## 3. DATASET COLLECTION

For the development and the evaluation of the proposed method, two measurement sequences were recorded by a tripod-mounted Livox AVIA sensor in two different, outdoor locations.

In the *Courtyard* scene, five people were walking in a narrow inner courtyard surrounded by large building facades, while canopies of trees and bushes are waving in the background due to the wind. The observed courtyard is 15m wide, its width is parallel to the NRCS Lidar's front plane, while the length of the observed area is 40m. This measurement setup was suitable for the $70^\circ$ horizontal field of view of the Livox sensor (see Fig. 3.) The sensor was placed horizontally, looking towards the horizon. 5-7 walking pedestrians formed the foreground regions of the scene, while the background consisted of parking cars, walls, trees, ground areas, etc. This setup utilized the benefits of the NRCS Livox sensor, as the foreground regions appeared close to the center of the sensor's field of view, resulting in better spatial resolution than in the peripheral FoV regions. Also, the distance regions of the observed are were suitable for the sensor's angular resolution.

The *City Center* sequence was recorded in a busy scene in downtown Budapest, Hungary, containing several moving vehicles and pedestrians. The selected square and junction were observed from a higher location, where the sensor was placed looking towards the ground. The foreground regions of this scene include various types of moving objects, including pedestrians, cars, trams, cyclists, etc. In this experiment, the observed area was in an open space, thus the observed distances were also limited by the sensor's reflection detection capabilities, not only by the static field objects such as buildings/vegetation. As the observed area was farther from the sensor than in the *Courtyard* scene, the *City Center* sequence has sparser data. As a consequence of the sparser measurements, we observed here a slightly longer initialization period of the high-resolution background model.

## 4. RESULTS AND DISCUSSION

The method was tested and evaluated using the *Courtyard* and *City Center* Livox Lidar measurements (see Sec. 3).

A demonstrating example for foreground classification on a sparse sample frame from the *Courtyard* sequence and the generated dense background model are displayed in Fig. 4 in the range image representation.

A sample result from the *City Center* dataset is displayed in Fig. 6 in point cloud representation. Here both the foreground and background objects were at larger distances, resulting in even sparser Lidar point cloud frames.

### 4.1 Quantitative Results

Numerical evaluation of the algorithm's performance was conducted via comparing the detection results to ground truth segmentation, which was manually generated for selected keyframes of both the *Courtyard* and the *City Center* Lidar measurement sequences. More specifically, we considered 25s long measurement segments in both scenes, and manually annotated every 5th point cloud (i.e. the annotation frame rate was 2fps) via a 3D annotation tool, separating the foreground and background regions.

| | Courtyard | City Center |
|---|---|---|
| Precision | 0.72 | 0.62 |
| Recall | 0.82 | 0.77 |
| F1 Score | 0.76 | 0.67 |
| IoU | 0.62 | 0.52 |

Table 1. Result of the quantitative evaluation of the method on the annotated Courtyard and City Center datasets

The quantitative performance analysis was performed by the comparison of each point's label after the assignment of the 3D corresponding points of the ground truth and the output clouds. To measure the similarity between the binary annotation of the ground truth point cloud, and the binary classification of each point in the result point cloud, the mean F1-score, Intersection over Union (IoU) were calculated alongside precision and recall. The used metrics' definition follows the standard binary classification metrics (Metz, 1978).

The results of the quantitative evaluation are listed in Table 1. The mean point-level F1-score of the method was 0.76 on the *Courtyard*, and 0.67 on the *City Center* sequences. These initial results are satisfying considering our low-level classification approach, which observation is can also be confirmed by qualitative experiments. In practical use, the existing classification errors can be eliminated by considering various higher-level object- or scene features, e.g. results of object detection using PointPillars deep neural network (Lang et al., 2019). The lower F1-score result of the City Center dataset is explained by the greater distance between objects and the sensor, which yielded a lower spatial resolution of the measurement.

The average running speed of the method was 80ms for each point cloud on a PC with an i7-7500U K CPU @2.7 GHz x4, 16 GB RAM.

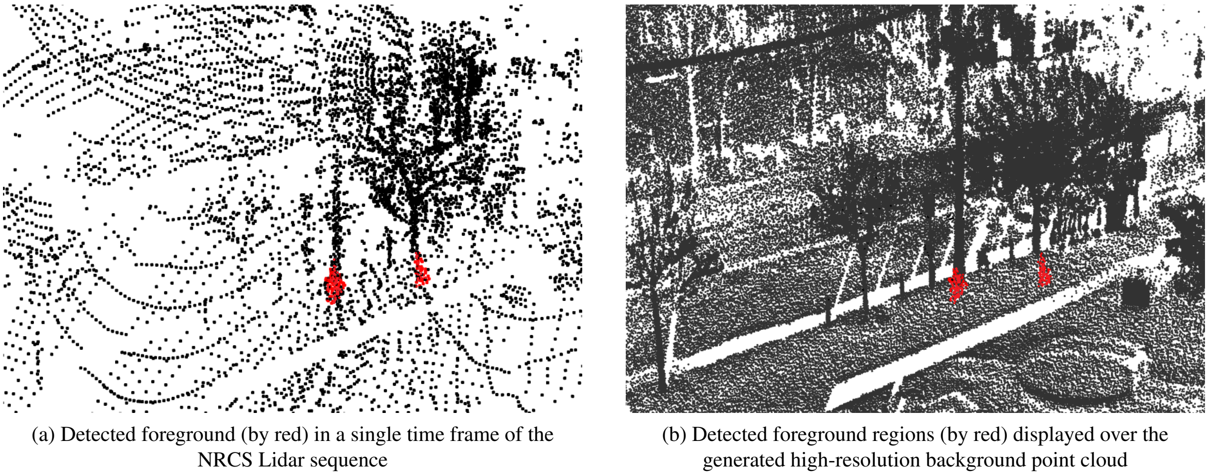
### 4.2 Qualitative Results

For qualitative analysis, we constructed first a dense 3D point cloud from the 2D high-resolution background model.

Then the moving objects detected in the consecutive Lidar frames (Fig. 6a) can be displayed with the background's dense point cloud in the same coordinate system, which can provide a useful visualization effect for the operators of a surveillance system (Fig. 6b).

We demonstrate the development phases of the dense background model by the adopted MoG approach in Fig. 7. As the time elapses, the sensor's non-repetitive scanning pattern covers more and more regions of its field of view, resulting in a step-by-step evolution of the background point cloud. By the end of the initialization process, all undefined regions disappear, and all pixels in the FoV receive a valid range value. Once the high-resolution background model is built, it is updated continuously during the surveillance process.

During the experiments, we also tested the adaptivity of the background model, by investigating the transition of different scene regions from foreground to background classes and vice versa. For example, Fig. 8 displays consecutive point cloud frames, where a walking pedestrian stopped for a certain time period, and its point cloud was built into the background model. We should also mention when the pedestrian started to walk



(a) Detected foreground (by red) in a single time frame of the NRCS Lidar sequence

(b) Detected foreground regions (by red) displayed over the generated high-resolution background point cloud

Figure 6. Foreground detection results (by red) in the *City Center* scene, displayed in 3D point cloud representation.

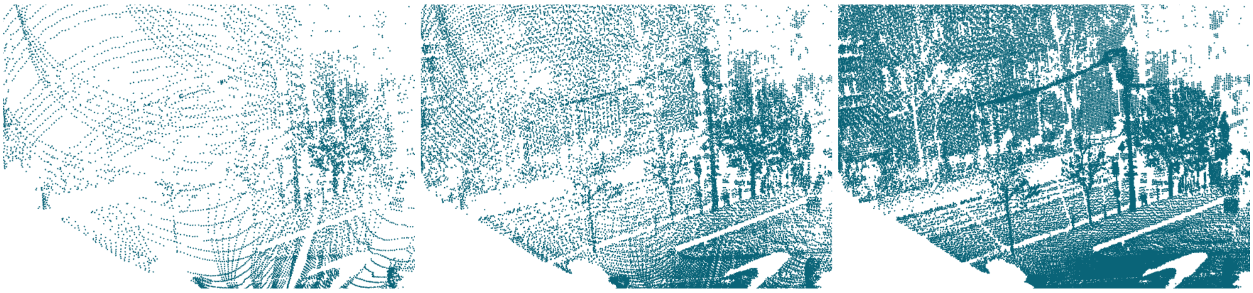


Figure 7. Evolution of the high-resolution background model in the City Center dataset

again later, we could see a quick "revival" of the hidden background region, whose range values were temporarily stored in the second strongest Gaussian components of the concerning pixels.

## 5. CONCLUSIONS

In this paper, a novel, robust and quick foreground-background segmentation method was presented, which works efficiently on point clouds recorded by Non-repetitive Circular Scanning Lidar sensors. The method can be extended in various ways, for example by taking into account object-level features, or the temporal dynamics of the observed scene via object tracking.

## ACKNOWLEDGEMENTS

This work was supported in part by the National Research Development and Innovation (NRDI) Office within the frameworks of the Autonomous Systems National Laboratory and the Artificial Intelligence National Laboratory programs, and in part by the PPKE-NVA-27 grant (L.K.) and by the SZTAKI-NVA-01 grant (M.K.,C.B.) financed under the TKP2021 funding scheme.

## REFERENCES

Alkhalili, Y., Luthra, M., Rizk, A., Koldehofe, B., 2019. 3-d urban objects detection and classification from point clouds. *13th ACM International Conference on Distributed and Event-Based Systems*, DEBS '19, New York, NY, USA, 209–213.

Avia User Manual, https://www.livoxtech.com, 2021.

Benedek, C., Gálai, B., Nagy, B., Jankó, Z., 2018. Lidar-Based Gait Analysis and Activity Recognition in a 4D Surveillance System. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(1), 101-113.

Benedek, C., Molnár, D., Szirányi, T., 2013. A dynamic MRF model for foreground detection on range data sequences of rotating multi-beam lidar. *Advances in Depth Image Analysis and Applications*, Springer Berlin Heidelberg, 87–96.

Börcs, A., Nagy, B., Benedek, C., 2017. Instant Object Detection in Lidar Point Clouds. *IEEE Geoscience and Remote Sensing Letters*, 14(7), 992-996.

Glennie, C. L., Hartzell, P. J., 2020. Accuracy assessment and calibration of low-cost autonomous sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 371–376.

Kaestner, R., Engelhard, N., Triebel, R., R.Siegwart, 2010. A Bayesian approach to learning 3D representations of dynamic environments. *Proc. International Symposium on Experimental Robotics (ISER)*, Springer Press, Berlin.

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O., 2019. Pointpillars: Fast encoders for object detection from point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12697–12705.

Li, K., Li, M., Hanebeck, U. D., 2021. Towards High-Performance Solid-State-LiDAR-Inertial Odometry and Mapping. *IEEE Robotics and Automation Letters*, 6(3), 5167-5174.

Lin, J., Zhang, F., 2020. Loam livox: A fast, robust, high-precision lidar odometry and mapping package for lidars of

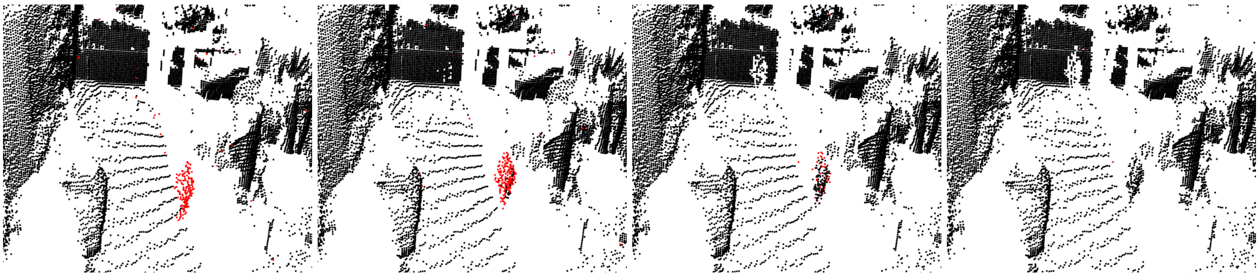


Figure 8. Transition of a region from the foreground (red) to the background (black), while a pedestrian stopped and stood in place for a longer time

small fov. *IEEE International Conference on Robotics and Automation (ICRA)*, 3126–3131.

Metz, C. E., 1978. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4), 283-298.

Schiller, I., Koch, R., 2011. Improved video segmentation by adaptive combination of depth keying and Mixture-of-Gaussians. *Proc. Scandinavian Conference on Image Analysis, Ystad, Sweden*, LNCS, 6688, Springer-Verlag, Berlin, Heidelberg, 59–68.

Stauffer, C., Grimson, W. E. L., 2000. Learning Patterns of Activity Using Real-Time Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 747–757.

Wang, Y., Lou, Y., Zhang, Y., Song, W., Huang, F., Tu, Z., 2021. A Robust Framework for Simultaneous Localization and Mapping with Multiple Non-Repetitive Scanning Lidars. *Remote Sensing*, 13(10). https://www.mdpi.com/2072-4292/13/10/2015.