2018 21st International Conference on Intelligent Transportation Systems (ITSC)
Maui, Hawaii, USA, November 4-7, 2018

Estimating the Scene-wise Reliability of LiDAR Pedestrian Detectors

Haruya Kyutoku[1], Yasutomo Kawanishi[1], Daisuke Deguchi[2], Ichiro Ide[1], Kazuki Kato[3] and Hiroshi Murase[1]

***Abstract*— Nowadays, development of driving support systems and autonomous driving systems have become active. Pedestrian detection from in-vehicle sensors is one of the most important technologies for these systems. However, outputs of pedestrian detectors can not be fully trusted in real environments. Therefore, we propose an estimation system of pedestrian detector's reliabilities for a given scene. This paper proposes a scene-wise reliability calculation method for LiDAR-based detectors, and a construction method for their estimators. Here, the problem is how we can define the reliability. The proposed method defines the reliability considering oversights as the strictest threshold without oversights. Meanwhile, it defines the reliability considering false detections as the loosest threshold without false detections. Experimental results showed that the proposed method could properly represent the reliability of a given scene, and estimate their reliability.**

I. Introduction

In this paper, we address the problem of estimating how much the output of a pedestrian detector can be trusted for a given scene.

Nowadays, development of driving support systems and autonomous driving systems have become an active research topic. The main focus of these systems is to reduce traffic accidents. Especially, overlooking pedestrians directly links to deadly accidents, so pedestrian detection from in-vehicle sensors is one of the most important technologies for these systems.

Accordingly, various attempts to detect pedestrians from in-vehicle cameras have been made [1], [2]. In addition, pedestrian detection using LiDAR has been actively researched with the growing interest in automatic driving systems [3]–[8].

However, outputs of pedestrian detectors can not be fully trusted in real environments. For example, in the case of using a camera, the detector could not detect pedestrians in difficult scenes as shown in Fig. 1; Figure 1(a) is an example of a scene with blown out highlights due to lens flare. There are two pedestrians at the left front, but the pedestrian detector could not detect them. Figure 1(b) is an example of clipped shadows. There are two pedestrians at the right front, but the pedestrian detector will not be able to correctly detect them. As described above, the output of a detector can not be completely trusted. Thus, in order to make proper use of the pedestrian detection technology, we need to consider the reliability of a detector for a given scene, in addition to its output.

(a) Blown out highlights. (b) Clipped shadows.

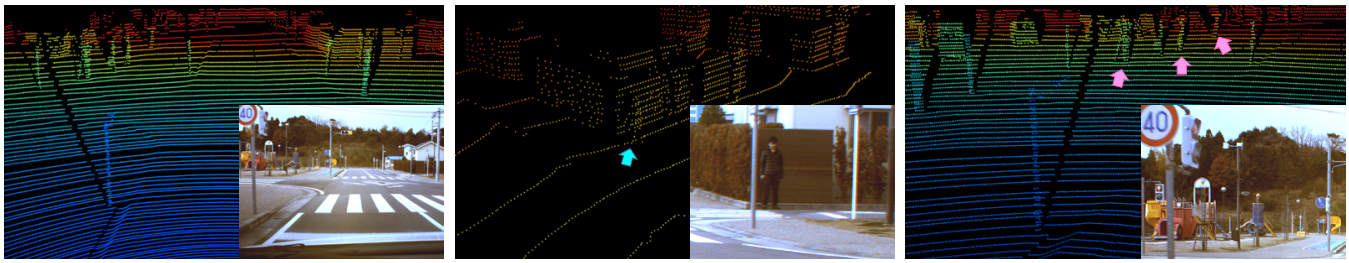
Fig. 1. Examples of difficult scenes for detection with a camera.

Likewise, there are difficult situations in the case of pedestrian detection using a LiDAR. Here, there are two kinds of difficult situations for a detector; oversights and false detections as shown in Fig. 2. Figure 2(a) is the full view of a scene, and Fig. 2(b) and Fig. 2(c) are examples showing each of the difficult situations. Details of each situation are described below.

- **Difficult situations due to oversights**
  Figure 2(b) is a magnified view of the right front part of Fig. 2(a). It is an example where a pedestrian exists close to a wall. The arrow indicates the pedestrian. Such a pedestrian will most likely be integrated with the wall in the detection process, and will be overlooked. However, such situations exist often in practice, so if a vehicle travels trusting the detector's output, it could lead to a serious accident. On the other hand, it may be possible to take actions such as running carefully if the information that oversight is likely to be occurring is provided to the driver.
- **Difficult situations due to false detections**
  Figure 2(c) is a magnified view of the left front part of Fig. 2(a). It is an example where there are many structures similar to pedestrians. The arrows indicate such structures. In this case, there are many rod-shaped structures. A detector will most likely detect such pedestrian-like structures as pedestrians. Such false detections will obstruct smooth traveling.

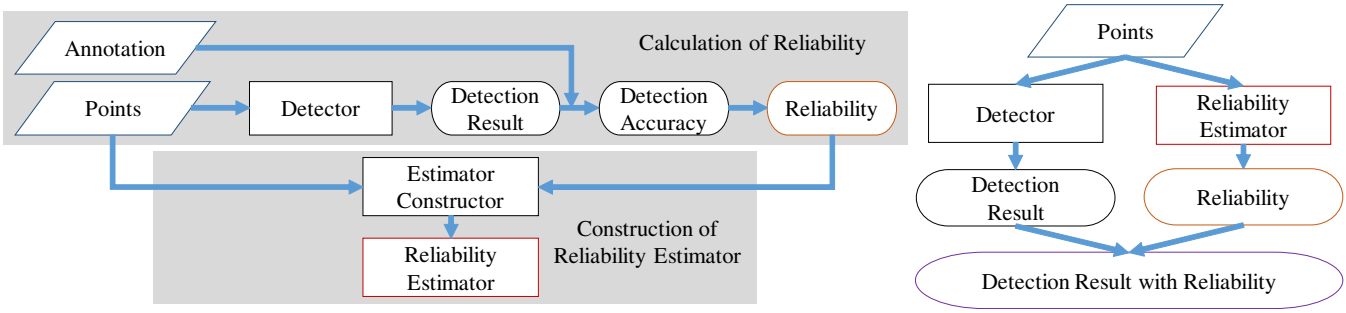
As discussed above, information on the reliability of a detector is very important. If we can estimate the reliability of a detector directly from the input independent of its output, safer autonomous driving and driver assistance can be realized. For example, the system can slow down when the reliability considering oversights is low, because there may be oversights. Moreover, the system can lower the urgency of the detection result when the reliability considering false detections is low, because unnecessary stops will increase the risk of accidents. As a matter of fact, a number of

[1]Haruya Kyutoku, Yasutomo Kawanishi, Ichiro Ide, and Hiroshi Murase are with Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan kyutoku@murase.is.i.nagoya-u.ac.jp
[2]Daisuke Deguchi is with Information Strategy Office, Information & Communications, Nagoya University, Nagoya, Aichi, Japan
[3]Kazuki Kato is with DENSO CORPORATION, Kariya, Aichi, Japan



(a) Full view. (b) Close to large structures such as a wall. (c) Existing pedestrian-like structures such as poles.

Fig. 2. Examples of difficult scenes for detection with a LiDAR. (Bottom right images are corresponding camera images for reference.)



(a) Learning step (b) Estimation step

Fig. 3. Proposed framework.

techniques to estimate the reliability have been developed for sensor fusion [9]–[11]. However, they are techniques for the reliability of the sensor itself or inputs. Therefore, we propose a framework for estimating the reliability of a LiDAR-based pedestrian detector for a given scene. In the following sections, we present the definitions of scene-wise reliabilities (probability of oversight and misdetection) of a detector. The proposed framework is shown in Fig. 3. It is composed of two steps; a learning step (Fig. 3(a)) and an estimation step (Fig. 3(b)). The learning step calculates the reliabilities and constructs reliability estimators. In the estimation step, the constructed estimator estimates the reliability of the detector directly from a given scene, and the system outputs it with the detection results.

This paper is organized as follows: Section II introduces the definition of reliabilities in detail. Section III explains the details of constructing the reliability estimator. Experimental results and discussions are presented in Section IV. Finally, Section V summarizes the paper.

II. Definition of Scene-Wise Reliabilities of a Detector

In this section, we introduce the definition of scene-wise reliabilities of a detector.

As errors in pedestrian detection, there are oversights and misdetections of non-pedestrian objects in the output of a pedestrian detector. Thus, when the number of these errors are small, we can say that the reliability of the detector for a given scene is high. The metric of reliability considering oversights indicates the degree of reliability in correctly detecting pedestrians, namely, the extent to which the oversights do not exist. On the other hand, the metric of reliability considering false detections indicates the degree of reliability considering misdetections, namely, the extent to which false detections do not exist. Each kind of reliability will be described in detail below.

A. Reliability Considering Oversights

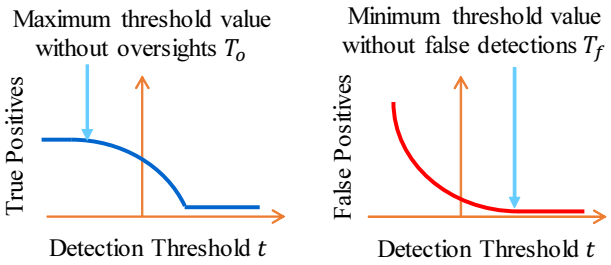
Firstly, the detection rate (recall) can be considered as the metric representing the degree of reliability considering oversights. A scene with lower recall will have more undetected pedestrians, so the scene will have more oversights. Thus, a scene with high recall can be considered as highly reliable. So recall can be used as the metric for the reliability considering oversights. However, since recall is difficult to calculate from a single scene, we propose a metric approximating it as follows.

A detector's threshold where the number of oversights is minimized is expressed as follows.

$$T_o = \max\{\arg\min_t N_{\mathrm{FN}}(t)\} \tag{1}$$

Here, t is a detector's threshold normalized in the range of [0.00, 1.00], and $N_{\mathrm{FN}}(t)$ is the number of false negatives (oversights). An example of the change in the number of true positives with changing the threshold t is shown in Fig. 4(a). As shown here, in general, the number of positive detections increase with lower (looser) threshold values, and decrease with higher (tighter) threshold values. The output from a lower threshold will include more false detections, so we can say that T_o (the maximum threshold value with minimal oversights) is better if its value is higher. Therefore, T_o is used as the reliability considering oversights as follows.

$$R_o = T_o \tag{2}$$



(a) Oversights. (b) False detections.

Fig. 4. Reliability metrics.

B. Reliability Considering False Detections

Meanwhile, precision can be considered as the metric representing the degree of the reliability considering false detections. A scene with lower precision will have more false detections, so the scene will have more fake detections. Thus, a scene with high precision can be considered as highly reliabile. So precision can be used as the metric for the reliability considering false detections. However, since precision is difficult to calculate from a single scene, we propose a metric approximating it as follows.

Next, a detector's threshold where a false detection occurs is expressed as follows.

$$T_f = \min\{\arg\min_t N_{\mathrm{FP}}(t)\} \tag{3}$$

Here, t is a detector's threshold normalized in the range of [0.00, 1.00], and $N_{\mathrm{FP}}(t)$ is the number of false positives (false detections). An example of the change in the number of false positives with changing the threshold t is shown in Fig. 4(b). As explained in Section II-A, in general, results with higher (tighter) threshold values will include more oversights. Conversely, oversights will be avoided with lower (looser) threshold values, so we can say that T_f (the minimum threshold value with minimal false detections) is better if its value is lower. Therefore, inverted T_f is used as the reliability considering false detections as follows.

$$R_f = 1 - T_f \tag{4}$$

III. Reliability Estimation

In this section, we describe the proposed scene-wise reliability estimation method.

The proposed scene-wise reliability estimator extracts several features from the point cloud of a given scene measured by a LiDAR and outputs the estimated reliability. Here, following the methology of general detection tasks, we first extract sets of points by clustering as pre-processing. We defined two types of features representing a scene, namely global features and pedestrian-like features, both described below. Both features are extracted from each cluster and in the end concatenated into a 271 dimensions vector. The reliability regressors of the estimators for (2) and (4) are constructed using ε-SVR (Epsilon Support Vector Regression) with RBF (Radial Basis Function) kernel. They are trained with these features as input and reliabilities as output.

Note that the coordinate axes are the X-axis in the horizontal direction, the Y-axis in the depth direction, and the Z-axis in the height direction.

A. Global Features

The following four features are extracted as global features from each cluster, and in the end concatenated into a 118 dimensions vector.

- **3D covariance matrix**
  Calculated from the three-dimensional covariance matrix of each cluster's centroid. Since this is a 3×3 diagonal matrix, six elements eliminating duplicates are used as features.
- **3D moment**
  Calculated from the three-dimensional moment of each cluster's centroid. Six elements eliminating duplicates are used as features.
- **Distance to candidate**
  Calculated from the set of distances to all of the clusters' centroid. To represent the statistics of this set, the median, mean, and standard deviation are used. In addition, a histogram of distances to all points projected onto the X-Y plane is used as feature. The histogram is composed of 100 bins with 1 m interval.
- **Number of points**
  Calculated from the number of points in each cluster. The mean, median, and standard deviation of these values are used as feature.

B. Pedestrian-like Features

The following three features are extracted as pedestrian-like features from each cluster, and in the end concatenated into a 153 dimensions vector.

- **Statistics for tall objects**
  Generally, pedestrians have a tall shape. Therefore, the ratio of clusters whose height is larger than the width and the depth when the cluster is represented by a rectangular solid is used as the feature. In addition, the areas of each cluster projected onto the X-Y plane are used. The mean, median, and standard deviation of these metrics are used as features.
- **Statistics for heights**
  Calculated from the maximum, minimum, centroid, and height (maximum $-$ minimum) of the Z coordinates of points in each cluster. The mean, median, standard deviation, and histogram of these values are used as features. The histograms of the position (maximum, minimum, centroid) are composed of nine bins with 0.5 m interval. Only the histogram of the height is composed of ten bins with 0.5 m interval.
- **Statistics for reflection intensity**
  Calculated from the maximum, minimum, mean, and median of the reflection intensity of points in each cluster. Their histograms are composed of 25 bins and used as features.

Fig. 5. Experimental vehicle

IV. Experiments

To confirm the effectiveness of the proposed method, experiments on scene-wise reliability estimation were conducted by using actual point clouds measured using a LiDAR as input.

A. Experimental Conditions

To acquire point clouds, we prepared a vehicle that mounted Velodyne LiDAR HDL-64e on the roof as shown in Fig. 5. The vehicle traveled through urban areas during the day time, and obtained 17 sequences composed of 3,871 frames. Each experiment was carried out with a leave-one-out cross-validation scheme where one sequence was used for evaluation and the others for learning.

In this experiment, pedestrians with no occlusion within approximately 40 m from the sensor were targets for the detection, which were annotated manually.

B. Pedestrian Detector

As pre-processing, road plane removal and clustering using Euclidean distance were performed for candidates extraction. Next, these candidates were classified whether they were pedestrian or not by a classifier. The pedestrian classifier used in this experiment was constructed with Real AdaBoost [12]. The following 225 dimensions feature was used for this.

- **Features proposed by Kidono et al. [3]**
  213-dimensional features.
- **Features proposed by Liu et al. [13]**
  Seven-dimensional features based on eigenvalues.
- **Additional features**
  Five-dimensional features. (The height of the cluster, the maximum and minimum of the Z coordinate of each point, the area in the X-Y plane, the median of the intensity of each point.)

C. Validity of the Proposed Reliabilities

Here we present an experiment to confirm the validity of the proposed reliabilities described in Section II.

The proposed method defined "threshold where oversights or false detections are minimized" as reliabilities. For evaluation of these criteria, the thresholds in each frame were calculated with the pedestrian detector constructed in Section IV-B. Note that the pedestrian detector applied to each sequence was trained with other sequences. In addition, clusters extracted from the pre-processing were used for evaluating oversights and false detections.

As an example, the result from sequence #14 is shown in Fig. 6. The threshold of the detector is in the range of [0.00, 1.00]. Figure 6(a) shows the maximum thresholds without oversights defined in (1) for each frame. As mentioned in Section II-A, we can say that outputs from scenes with higher threshold values can be trusted. On the other hand, Figure 6(b) shows the minimum thresholds without false detections defined in (3) for each frame. As mentioned in Section II-B, we can say that outputs from scenes with lower threshold values can be trusted. Note that blanks in Fig. 6(a) are frames with no pedestrian to be detected.

From Fig. 6, we can see that the threshold changes across frames in accordance with the change of scene following the travelling of the vehicle. A frame indicated by a solid circle in Fig. 6(a) is shown in Fig. 7(a) as an example of a low reliability scene considering oversights. A pedestrian existing near a wall, so it was difficult to detect. Similarly, a frame indicated by a dotted circle in Fig. 6(a) is shown in Fig. 7(b) as an example of a high reliability scene considering oversights. A pedestrian existing in an open intersection, so it was easy to detect.

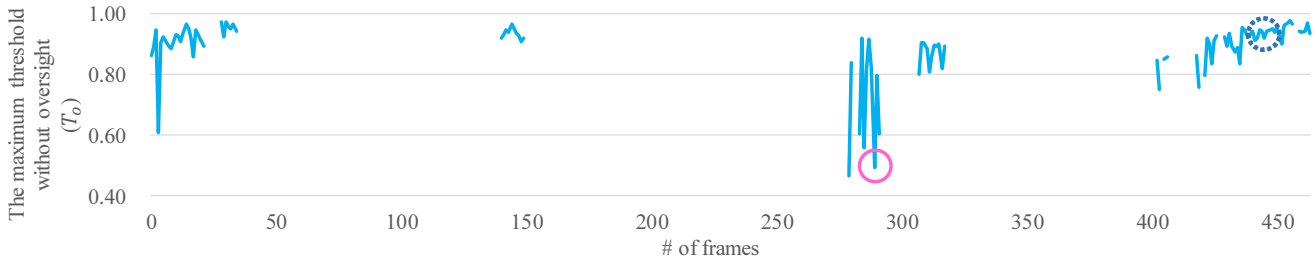
Meanwhile, a frame indicated by a solid circle in Fig. 6(b) is shown in Fig. 7(c) as an example of a low reliability scene considering false detections. Since there were many structures similar to pedestrians such as poles, trees, and so on, false detections were likely to occur. Similarly, a frame indicated by a dotted circle in Fig. 6(b) is shown in Fig. 7(d) as an example of a high reliability scene considering false detections. Although there are trees along the street, they were easy to be separated from pedestrians because they were few and thin.

As described above, we confirmed the validity of the proposed reliabilities. In the remaining part of the experiments, the reliabilities calculated here are used as ground truths.

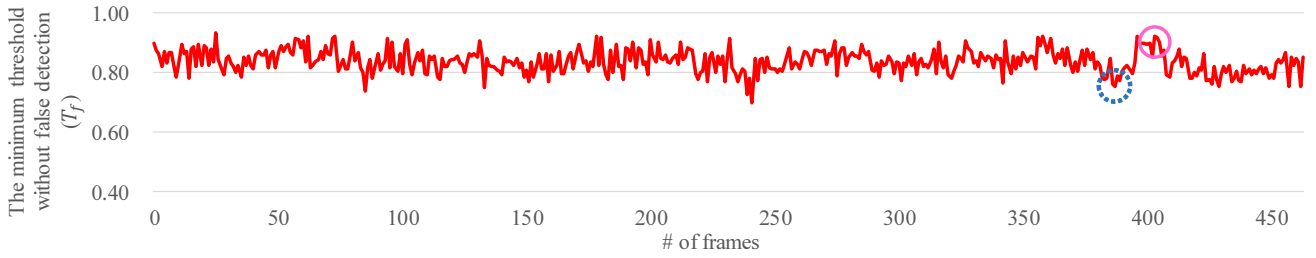
D. Effectiveness of the Proposed Reliability Estimation

Here we present an experiment to confirm the effectiveness of the proposed reliability estimator described in Section III. The estimators were constructed with the reliabilities calculated in Section IV-C as ground truths. Note that the estimators for each sequence were learned with the reliabilities of other sequences.

As a result, the proposed estimator could estimate the reliability considering oversights with 0.055 mean absolute error, and that considering false detections with 0.033 mean absolute error. As an example, the estimation result for sequence #14 is shown in Fig. 8. From Fig. 8(b), we can see that the reliability estimator considering false detections was able to estimate quite well. Meanwhile, from Fig. 8(a), we can see that the reliability estimator considering oversights failed to estimate well when the reliability decreased drastically. The following two reasons could be the cause for this.



(a) Considering oversights.



(b) Considering false detections.

Fig. 6. Thresholds where oversights or false detections are minimized.

- **Lack of features**
  Features leading to oversights such as integration of clusters were lacking.
- **Insufficient amount of training data**
  The reliability considering oversights is calculated from only frames including pedestrians, so the size of the training data was insufficient.

V. Summary

In this paper, we focused on the problem that outputs from pedestrian detectors can not be fully trusted in real environments. Considering this, we defined reliabilities of detectors for a given scene, and proposed a method to construct a scene-wise reliability estimator. The proposed method defined the "threshold where oversights or false detections are minimized" as scene-wise reliabilities.

To demonstrate the effectiveness of the proposed method, experiments were conducted by applying the proposed method to actual point clouds measured from a LiDAR. The experimental results showed that the definitions of reliabilities were reasonable, and that the proposed estimator can estimate the reliabilities properly.

For future work, we need to extend features of the reliability estimator for more accurate estimation. For example, rather than outputting a global reliability for a scene, we may calculate estimation features and the reliability of local regions. Furthermore, we should also design a method that considers the final detection results considering the reliability.

ACKNOWLEDGMENT

Parts of this research were supported by the Center of Innovation Program from Japan Science and Technology Agency, JST.

References

[1] H. Yoshida, D. Suzuo, D. Deguchi, I. Ide, H. Murase, T. Machida, and Y. Kojima, "Pedestrian detection based on deep convolutional neural network with ensemble inference network," in *Proc. 2013 IEEE Intelligent Vehicles Symposium*, Jun. 2013, pp. 654–659.

[2] H. Fukui, T. Yamashita, Y. Yamauchi, H. Fujiyoshi, and H. Murase, "Pedestrian detection based on deep convolutional neural network with ensemble inference network," in *Proc. 2015 IEEE Intelligent Vehicles Symposium*, Jul. 2015, pp. 223–228.

[3] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, "Pedestrian recognition using high-definition LIDAR," in *Proc. 2011 IEEE Intelligent Vehicles Symposium*, Jun. 2011, pp. 405–410.

[4] T. Ogawa, H. Sakai, Y. Suzuki, K. Takagi, and K. Morikawa, "Pedestrian detection and tracking using in-vehicle LIDAR for automotive application," in *Proc. 2011 IEEE Intelligent Vehicles Symposium*, Jun. 2011, pp. 734–739.

[5] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks," in *Proc. 2017 Int. Conf. on Robotics and Automation*, May 2017, pp. 1355–1361.

[6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in Neural Information Processing Systems 30*, pp. 5099–5108, 2017.

[7] T. Yamamoto, F. Shinmura, D. Deguchi, Y. Kawanishi, I. Ide, and H. Murase, "Efficient pedestrian scanning by active scan LIDAR," in *Proc. 2018 Int. Workshop on Advanced Image Technology*, no. C4-2, Jan. 2018, pp. 1–4.

[8] Y. Tatebe, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase, and U. Sakai, "Pedestrian detection from sparse point-cloud using 3DCNN," in *Proc. 2018 Int. Workshop on Advanced Image Technology*, no. C4-4, Jan. 2018, pp. 1–4.

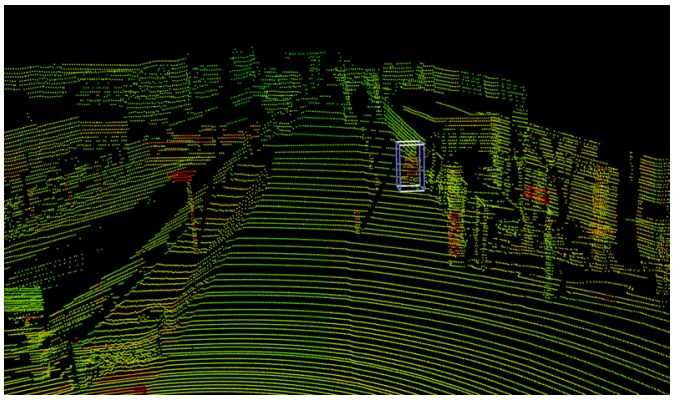
[9] A.-L. Jousselme, A.-C. Boury-Brisset, B. Debaque, and D. Prevost, "Characterization of hard and soft sources of information: A practical illustration," in *Proc. 2014 Int. Conf. on Information Fusion*, Oct. 2014, pp. 1–8.

[10] N.-E. E. Faouzi and L. A. Klein, "Data fusion for ITS: Techniques and research needs," *Transportation Research Procedia*, vol. 15, pp. 495–512, Jun. 2016.

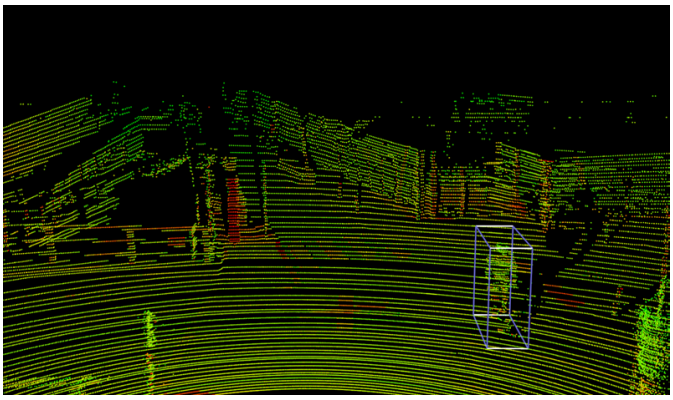
[11] A. R. Hilal, "Context-aware source reliability estimation for multi-sensor management," in *Proc. 2017 Annual Int. Systems Conf.*, Apr. 2017, pp. 1–4.

[12] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, Dec. 1999.

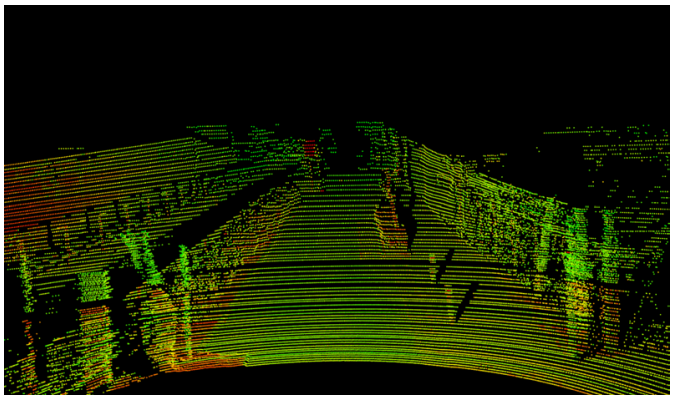
[13] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. on Image Processing*, vol. 24, no. 8, pp. 2502–2514, Apr. 2015.



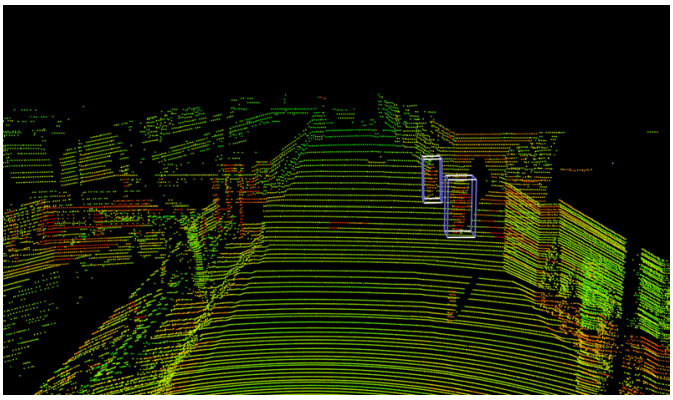
(a) Low reliability scene considering oversights due to large structures.



(b) High reliability scene considering oversights due to open space.

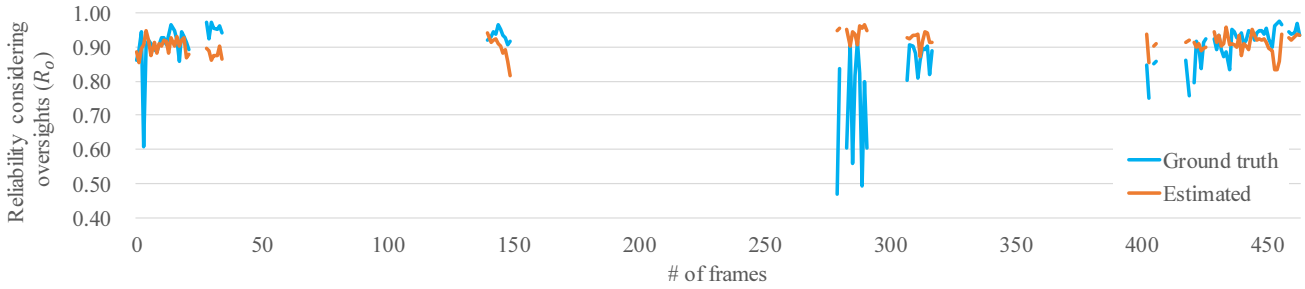


(c) Low reliability scene considering misdetections due to existence of similar objects.

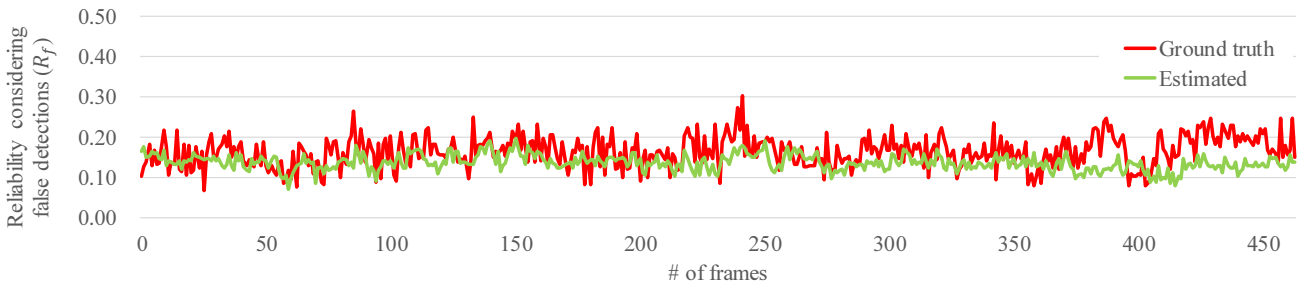


(d) High reliability scene considering misdetections due to inexistence of similar objects.

Fig. 7. Examples of scenes with low and high reliabilities.



(a) Considering oversights.



(b) Considering false detections.

Fig. 8. Estimated reliabilities.