Article

Deep Learning Method on Target Echo Signal Recognition for Obscurant Penetrating Lidar Detection in Degraded Visual Environments

Xujia Liang, Zhonghua Huang, Liping Lu, Zhigang Tao, Bing Yang and Yinlin Li *

School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China; 3120180213@bit.edu.cn (X.L.); huangzh@bit.edu.cn (Z.H.); 3120190216@bit.edu.cn (L.L.); 3120180244@bit.edu.cn (Z.T.); 3120190249@bit.edu.cn (B.Y.)

* Correspondence: liyinlin@bit.edu.cn; Tel.: +86-010-6891-1081

Received: 24 May 2020; Accepted: 16 June 2020; Published: 17 June 2020

Abstract: With the rapid development of autonomous vehicles and mobile robotics, the desire to advance robust light detection and ranging (Lidar) detection methods for real world applications is increasing. However, this task still suffers in degraded visual environments (DVE), including smoke, dust, fog, and rain, as the aerosols lead to false alarm and dysfunction. Therefore, a novel Lidar target echo signal recognition method, based on a multi-distance measurement and deep learning algorithm is presented in this paper; neither the backscatter suppression nor the denoise functions are required. The 2-D spectrogram images are constructed by using the frequency-distance relation derived from the 1-D echo signals of the Lidar sensor individual cell in the course of approaching target. The characteristics of the target echo signal and noise in the spectrogram images are analyzed and determined; thus, the target recognition criterion is established accordingly. A customized deep learning algorithm is subsequently developed to perform the recognition. The simulation and experimental results demonstrate that the proposed method can significantly improve the Lidar detection performance in DVE.

Keywords: obscurant penetrating; Lidar; deep learning; 2-D spectrogram image; visual degraded environment (DVE)

1. Introduction

Light detection and ranging (Lidar) sensors are widely used in detection applications, such as autonomous driving [1–3] and unmanned ground vehicles (UGV) [4], due to the characteristics of high resolution [5] and precision [6]. However, when they are applied in degraded visual environments (DVE), such as smoke, fog, dust, and rain, the laser signals of Lidar sensors are attenuated and absorbed by scattering effect and extinction effect of aerosols, thus, it is difficult to extract effective beat signals and detect the targets further [7]. Therefore, using Lidar for target detection in DVE remains a huge challenge [8].

Nowadays, researches on obscurant penetrating mainly focus on fixed distance measurement. Several denoise and backscatter suppression methods have been developed and implemented. Donoho et al. [9] proposed a denoise method, based on the wavelet transformation method. After Huang proposed the Empirical Mode Decomposition (EMD) method, the researchers proposed several EMD-based denoise methods to process the disturbed signal, such as the EMD direct wavelet thresholding method, the EMD interval thresholding method and the EMD-based partial reconstruction [10–13]. Shamsudin et al. [14] developed a denoise method using SVM and KNN to distinguish the fog and target echo signal in the point

cloud data. Heinzler et al. [2] presented a Convolutional Neural Network based (CNN-based) method to distinguish and filter out scatter points caused by adverse weather in point cloud data. Charron et al. [15] proposed a method to remove snow noise by processing point cloud data using a 3-D outlier detection algorithm, which considers the change of point cloud density with the increase of distance from the sensor. In terms of underwater Lidar detection, Illig et al. [16] applied blind signal separation (BSS) to measure and suppress backscattering signal. O'Connor et al. [17] developed several techniques using RF modulation to suppress background clutter and enhance the behavior of target detection, which is quite effective in both forward and backscatter suppression. Djuricic et al. [18] proposed a multiple-pulse laser scanning method to distinguish the target and soft obstacle. However, denoise and backscatter suppression methods are widely analyzed in DVE, yet limited by their computational cost.

Other methods focus on upgrading the optical design of sensors. The OPAL Lidar developed by Neptec company utilize high power emitter with a sensitive receiver to suppress the backscatter from aerosols [19]. Starr et al. [20] fused infrared camera and Lidar for range finding in fire environments. However, the expensive and complex shortcomings limit their application in obscurant penetrating Lidar.

In real application scenarios, Lidar is often installed on moveable platforms, such as unmanned ground vehicles and robots. For practical applications, it is feasible to analyze and apply multiple echo signals when the platform moves. To overcome the above-mentioned deficiencies, this paper presents a novel obscurant penetrating detection method for Lidar, which does not require complex denoising functions. The major contributions of this paper are as follows.

- (1) We propose a novel multi-distance measurement method that acquires multiple echo signals in the process of approaching the target. Unlike the fixed-distance measurement, this method acquires multiple echo signals during the movement of the Lidar sensor.
- (2) We propose a novel target echo signal recognition method based on 2-D spectrogram images. The 2-D spectrogram images are constructed by using the frequency-distance relation derived from the 1-D echo signals of the Lidar sensor individual cell in the course of approaching target.
- (3) We propose a customized deep learning algorithm based on Faster Region Convolutional Neural Network (R-CNN) to recognize the target echo signal and to predict the target distance.

The rest of this paper is organized as follows: Section 2 describes target echo signal recognition method based on 2-D spectrogram images and the deep learning algorithm in detail. To evaluate the effectiveness of the proposed method, the simulation and experimental results are given in Section 3. Section 4 draws the summary and directions for future work of this paper.

2. The Obscurant Penetrating Lidar Detection Method

2.1. Method Overview

This paper presents a novel obscurant penetrating detection Lidar method. Among all the obscurants, smoke interference is the most typical interference factor that affects the recognition accuracy of Lidar [21,22]. Thus, this paper is mainly focusing on the smoke environment. The workflow of the method is shown in Figure 1. First, the multi-distance measurement acquires multiple 1-D echo signals during Lidar movement. Next, the 2-D spectrogram images are constructed by using the frequency-distance relation derived from the 1-D echo signals. Then, the deep learning algorithm based on Faster R-CNN extracts the target echo signal based on the joint relationship of the 1-D echo signal sequence. Meanwhile, the target distance will be further predicted. Finally, the Lidar detects the target during the movement.

Figure 1. The framework of the obscure penetrating detection method in this paper.

2.2. The Multi-Distance Measurement

First, it is necessary to define the multi-distance measurement. The purpose of explaining the method in detail is to distinguish it from the fixed position measurement method. As Figure 2 illustrates, the Lidar sensor is installed on a moveable platform. In the smoke environment, the Lidar regularly emits signals and receives echoes as the platform moves. R is the distance that the platform moves between two probes, n is the total number of measurements. Measurement distance is the distance the platform moves between the first detection and the last detection, and the target distance indicates the distance from the target during the last time of measurement. Moreover, the distance between each measurement interval is the same, which can be achieved by controlling the moving speed of the platform.

Figure 2. Multi-distance measurement model and application process. The picture shows the multiple measurements made by light detection and ranging (Lidar) during the movement.

2.3. The 2-D Spectrogram Images

Before proposing the target echo signal recognition criterion based on 2-D spectrogram image, we first analyzed the 1-D echo signals received in different conditions.

Figure 3 demonstrates the 1-D echo signals in different visibility conditions. As Figure 3 illustrates, echo signals in Figure 3a are obtained in clear air environments, which can be considered a case of infinite visibility. In the smokeless condition, the highest peak of echo signal demonstrates the location of the target echo signal [7]. It can be found that the frequency of echo signal peak increases with the target distance. Therefore, we visualize the distribution of the frequency of beat signal peak in different visibility conditions as Figure 4a. It can be noticed that, in a smokeless environment, the frequency of target echo signal linearly increases over target distance. Therefore, the relationship between the target distance and the frequency of the target echo signal in smokeless condition can be described as:

$$f = \rho \cdot d \quad (1)$$

where ρ is a constant determined by the frequency of beat signal, the period and frequency offset of the FM signal [23–25]. However, in the harsh visibility conditions shown in Figure 3, the noise leads to multiple peaks with the rise of smoke concentration. Due to the presence of multiple peaks, it is difficult to extract target echo signal peak in harsh visibility conditions. Moreover, it is not feasible to use Equation (1) to calculate the target distance in a smoke environment.

Figure 3. The 1-D echo signals in different conditions. (a) In smokeless conditions; (b) in conditions of 15 m visibility; (c) in conditions of 10 m visibility; (d) in conditions of 5 m visibility. The label below each spectrum ('5 m', '8 m' and '10 m') indicates the measurement distance from the target.

Figure 4. The distribution of beat signal peaks and target echo peaks in different visibility conditions: (a) The distribution of beat signal peaks in degraded visual environments (DVE); (b) the distribution of target echo signal peaks in DVE.

In addition, it should be noted that, as Figure 3 shows, the noise is mainly distributed in the low frequency part, while the frequency of target echo signal increases over target distance. Thus, it is obvious

that the target echo signal is separated from the noise. In order to validate the characteristic, 1-D echo signals in different visibility conditions are analyzed. At first, we design a filter to filter out the peaks of the low frequency part of the echo signals. The filter can be described as:

$$f_r = \max(0, f - f_{th}) \quad (2)$$

where f_r is the remaining frequency peaks, f_{th} is the threshold value of the frequency. Next, the distribution statistics of peaks in the remaining part are recorded. As Figure 4b shows, the peaks in the remaining part have the same distribution with the peaks in the smokeless environment. It proves that filtering the peak of the low frequency part is helpful to identify the target echo signal. More importantly, the noise is separated from the target echo signal.

Therefore, in order to clearly separate the target echo signal and noise, we began to look for a representation that can fuse signals from multiple distances. Consequently, we proposed the 2-D spectrogram images, which are constructed by using the frequency-distance relation derived from the 1-D echo signals acquired from the multi-distance measurement. The construction process of 2-D spectrogram image is shown in Figure 5. The multiple 1-D signals are combined, as Equation (3) shows. In this process, the multiple 1-D echo signals are normalized in 2-D spectrogram image, which contains the characteristics of the target echo signal and noise. In addition, the size of the 2-D spectrogram image can also be adjusted as needed.

$$A = \frac{1}{M} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} \quad (3)$$

where A represents the matrix of the 2-D spectrogram, P_i represents the sequence of the i -th 1-D echo signal, M is the maximum value of all 1-D echo signals.

Figure 5. The construction process of 2-D spectrogram image. In 2-D spectrogram image, the yellow part represents the high amplitude value, and the blue part represents the low amplitude value.

The most significant advantage of 2-D spectrogram image is that it demonstrates the amplitude–frequency characteristics of target echo signal during the process of approaching the target, as the noise is separated from the target echo signal. Therefore, the target echo signal can be demonstrated as Figure 6, as the correlation of adjacent data in the sequence is very obvious. Moreover, in order to further measure the target distance, the positions of target echo signal peaks are labeled as well. When the target echo signal is recognized in the spectrogram image, it indicates that the target is detected.

Figure 6. The 2-D spectrogram image in smokeless environment. The target echo signal and target echo signal peaks are annotated in the figure.

2.4. The Target Echo Signal Recognition Algorithm

The target echo signal can be annotated in the 2-D spectrogram image, while the target frequency linearly increases over target distance. Therefore, a novel signal recognition algorithm based on the visual signature extracted from the echo signal's time-frequency representation is presented for target detection. As Figure 7 illustrates, two Faster R-CNN networks are utilized in the algorithm, one for signal recognition, and the other is used for target ranging further.

Figure 7. The structure of the proposed algorithm.

Faster R-CNN contains two modules: the region proposal network (RPN) module, and the Fast R-CNN detection module [26]. The RPN module is utilized for extracting candidate regions. Then, the Fast R-CNN module classifies the extracted candidate regions into target and background. As Figure 8 shows, the Faster R-CNN consists of two full-connection layers: a regression layer and a classification layer. The former refines and corrects the proposal regions, so the predicted bounding box is closer to the ground truth. The latter classifies the proposal regions.

Figure 8. The Faster R-CNN workflow diagram.

In RPN, the proposal regions are predicted based on the preset anchors' scales. Two types of anchors are applied for the detection network and the ranging network respectively. The inputs' scales are clustered to optimize the parameters. As is shown in Figure 9, due to the huge difference between the two types of anchors, two RPN modules are adopted, to predict the proposal regions respectively.

Figure 9. Illustration of the distribution of the two anchors according to the box area and the aspect ratio.

We now describe the loss function applied for training the algorithm. As analyzed from above, the loss function is divided into detection part and ranging part. The first output $L_{reg_det}(t_i, t_i^g)$ determines the target's coordinate, where t_i is a vector and contains (t_x, t_y, t_w, t_h) , representing the predicted offset of the anchor's location, and t_i^g represents the true offset of the anchor relative to the ground truth. The second output is a cross-entropy loss function of binary classification, denoted as $L_{class_det}(p_i, p_i^g)$, where p_i estimate the presence of target in the foreground, and p_i^g represents the ground truth [26]. The last two of the four loss functions, $L_{reg_rag}(t_j, t_j^g)$ and $L_{class_rag}(p_i, p_i^g)$ evaluate the ranging part of the algorithm. The descriptions of these losses are shown as:

$$\begin{cases} L_{reg_det} = \lambda \frac{1}{N_{reg_det}} \sum_i p_i^* R(t_i - t_i^*) \\ L_{class_det} = \frac{1}{N_{class_det}} \sum_i -\log[p_i p_i^* + (1 - p_i)(1 - p_i^*)] \\ L_{reg_rag} = \lambda \frac{1}{M_{reg_rag}} \sum_i p_i^* R(t_i - t_i^*) \\ L_{class_rag} = \frac{1}{M_{class_rag}} \sum_i -\log[p_i p_i^* + (1 - p_i)(1 - p_i^*)] \\ L = L_{reg_det} + L_{class_det} + L_{reg_rag} + L_{class_rag} \end{cases} \quad (4)$$

where p_i and t_i are the prediction probability and predicted coordinate, p_i^* and t_i^* are the label and coordinate of ground truth, λ is weight balance parameter, we set it to 10, R is the Smooth L1 function, N_{reg_det} and M_{reg_rag} are the number of anchors in the two networks respectively, N_{class_det} and M_{class_rag} are the number

of mini batch sizes in the two networks, respectively [26]. The sum of the above four losses is minimized until convergence, indicating that the network has finished training.

So far, the recognition algorithm has been proposed, the target distance can be calculated by the location of target echo signal peak as:

$$D = \frac{(n-1)R}{\sum_{i=2}^n (f_i - f_{i-1})} f_n \quad (5)$$

where D is the predicted distance of the Lidar from the target, f_i is the frequency of target echo signal peak, n is the total number of measurements, R is the distance between each measurement.

Average precision (AP) and average ranging error (ARE) are applied as evaluation indicators. AP evaluates the performance of the detection part and ARE evaluates ranging part. AP evaluates the detection performances under different threshold conditions. The higher the AP value, the better the detector's detection performance. Meanwhile, the ARE is given as:

$$ARE = \frac{1}{N} \sum_0^N |D - R| \quad (6)$$

where D is predicted distance calculated by Equation (5), R is the actual distance between the detector and the target. The lower the ARE, the better the detector's ranging performance.

3. Simulation and Experimental Results

In order to more realistically analyze the detection performance of the proposed method, the recognition algorithm was evaluated by simulation and experimental echo signals, respectively. Subsequently, the evaluate results were then compared to the other algorithms under the same conditions. In the end, the evaluate results in different visibility conditions were analyzed.

3.1. The Simulation Results

3.1.1. The Simulation and Training Settings

The entire simulation process is shown in Figure 10. The Monte Carlo model was adopted to simulate the echo signals of Lidar in DVE [27–29]. The delay time and amplitude of the received signal can be calculated by calculating the time when the photon enters the detector and the energy of all the photons received at that time. The beat signal is simulated after mixing with the local oscillator signal and low-pass filtering. In order to simulate the real scene, the simulation can change the scattering coefficient and extinction coefficient of smoke aerosol, according to the different smoke concentrations. The relationship of visibility and smoke concentration can be expressed as:

$$v = \sqrt[y]{\frac{c_0}{M_c}} \quad (7)$$

where the M_c is the smoke concentration, v is the visibility, c_0 and y are constants, depending on the aerosol type, and in a smoke environment, the values of these two constants are 37.3 and 1.07 respectively [30].

A total of 135 2-D spectrogram images were simulated in different visibility conditions, ranging from 1 m to 15 m, and the training set and test set were separated as 7:3. Figure 11 shows the 2-D spectrogram images constructed from the simulated signals. The construction detail of the 2-D spectrogram image is shown in Table 1.

Figure 10. The simulation setup process of echo signal in obscurant environments.

Figure 11. The 2-D spectrogram images of different visibility conditions: (a) spectrogram image in smokeless condition; (b) spectrogram image in a condition of 15 m visibility; (c) spectrogram image in a condition of 10 m visibility; (d) spectrogram image in a condition of 8 m visibility.

Table 1. Construction details about 2-D spectrogram images.

Visibility/m	Measurement Distance	Measurement Times	Annotation Situation
1–3	13 m to 3 m	11	No
4–6			Yes
7–15	16 m to 6 m		

Afterwards, we annotated the target echo signals. As Figure 3 shows, the amplitude of target echo signal becomes smaller and smaller as the smoke density increases. In conditions with 1 m, 2 m, 3 m visibility, the target echo signals were too weak to annotate. In order to further evaluate the extreme performance of the proposed algorithm, we shortened the measurement distance in harsh visibility conditions. Measurement times refers to the number of measurements during the multi-distance measurement.

Then, the stochastic gradient method (SGD) was applied in the optimization process, of which the number of iterations was 80,000, the momentum and the weight decay were 0.9 and 0.0005, respectively. The initial learning rate of the first 60,000 mini-batch was 0.0001, and the last 20,000 mini-batch was 0.00001. Since ResNet50 has a better detection performance than VGG16, we applied ResNet50 as the pre-trained module in the algorithm. In total, the algorithm was trained for 100 epochs from scratch. We applied NVIDIA's GTX 960M graphics card for accelerating computing.

3.1.2. Detection Results of the Simulated Signals

(1) Comparison with Other Deep Learning Algorithms

The deep learning detection methods are divided into region proposal based and non-region proposal based algorithms. The representative of the former is R-CNN, Fast R-CNN, and Faster R-CNN, the representative of the latter is You Only Look Once (YOLO). The region proposal based algorithm has a higher recognition accuracy, while the non-region proposal based method, such as YOLO, has better real-time performance with lower detection accuracy. Moreover, YOLO cannot detect the small target in the image [31]. We investigated the performances of different algorithms (YOLO, R-CNN, Fast R-CNN, and Faster R-CNN) and different resolutions (227×227 and 500×500).

As can be seen from Table 2, YOLO cannot effectively identify the target echo signal in the 2-D spectrogram image, but, compared with other algorithms, it takes the fastest time, which is consistent with the previous view that the non-region proposal based method cannot identify small targets. Therefore, we focus on comparing the recognition results, based on the region proposal based algorithms. The performances of models with 500×500 inputs are better than those with 227×227 , as larger inputs contain more detail. At the same time, the larger the input image, the more time it consumes. By using 500×500 instead of 227×227 , the AP of Faster R-CNN has been improved by about 6.4%, and the ARE has been reduced by 1.67 m. In 500×500 conditions, the AP of R-CNN and Fast R-CNN are 75.1% and 86.3% respectively, lower than that of Faster R-CNN. While R-CNN and Fast R-CNN adopt selective search, which is based on random size cropping to predict the ROI, Faster R-CNN employs RPN [32]. RPN is much faster than the selective search on extracting candidate regions, because RPN reduces the number of candidates [31]. The detection results by selective search and RPN can be shown in Figure 11. As Figure 12 shows, the RPN generates much fewer proposal regions, with more accurate locations. Moreover, the ARE of Faster R-CNN, 0.24 m, is lower than the other, as the detection function contributes to the ranging performance. Thus, the model applying Faster R-CNN with 500×500 input has the best performance in the simulation experiment.

Table 2. The results of the different algorithms.

Methods	Resolution	AP (%)	ARE (m)	Time(s)
YOLO	227 × 227	0	∞	0.02
	500 × 500	0	∞	0.08
R-CNN	227 × 227	57.0	2.75	5.63
	500 × 500	75.1	1.32	53.34
Fast R-CNN	227 × 227	75.9	2.53	1.24
	500 × 500	86.3	1.34	4.42
Faster R-CNN	227 × 227	85.0	1.91	0.49
	500 × 500	91.4	0.24	0.86

Figure 12. The detection results by selective search and the region proposal network (RPN): (a) results of selective search; (b) target echo signal location by RPN.

(2) Detection Performances in Different Visibility Conditions

The proposed method is evaluated by simulated signals of different visibility conditions ranging from 1 m to 15 m. According to the results, in conditions with visibility of 1 m, 2 m, and 3 m, the algorithm did not recognize the presence of target echo signal, which was in line with the fact that we have not annotated the ground truth in those spectrogram images. The remaining results are shown in Figure 13.

Figure 13. The performance in different visibility conditions.

As can be seen in Figure 13, in conditions that visibility above 7 m, the AP decreases gradually, since the impact of noise increase with smoke concentration. It is worth noting that the AP of 6 m visibility is significantly improved compared with that of 7 m visibility. The possible reason is that the 2-D spectrogram images of 6 m visibility are measured from 13 m to 3 m, which are much closer, compared with the 7 m

visibility chart, which measured from 16 m to 6 m. In addition, ARE increases slightly as visibility decreases, which indicates that the detection performance has some positive impact on the ranging function.

Figure 14 shows detection results by utilizing the trained model on the test samples. It is clear that when the visibility is relatively high, the target echo signal has a strong intensity and is relatively significant in the 2-D spectrogram images, making it easy for recognition. As the visibility drops, the recognition becomes difficult, because the characteristics of target echo signal are not dramatic.

Figure 14. The detection performances in different visibility conditions: (a) results in a 4 m visibility condition; (b) results in a 14 m visibility condition.

3.2. Experimental Results

3.2.1. Experimental and Training Settings

In order to verify the method, a FMCW Lidar was applied to collect echo signal in DVE. Figure 15 demonstrates the experimental process. This experiment was conducted in a $7\text{ m} \times 1.5\text{ m} \times 2\text{ m}$ site, which was constructed using black flannel with a reflection coefficient of 0.05. The trolley carrying the Lidar and spectrum analyzer is placed 6 m away from the target board. At first, the smoke generator sprays smoke into the test site, the visibility test module measures the smoke concentration when the visibility is stable. Next, the trolley is pushed towards the target board at a constant speed. In this process, the echo signal is recorded every fixed distance. After that, the echo signals are computed by FFT to form 2-D spectrogram images. Eventually, the 2-D spectrogram images are sent into the proposed algorithm.

The construction details of the 2-D spectrogram images are shown in Table 3. Due to the limitation of the experiment site, the measurement distance of the experimental process is shorter than that of the simulation process. After the spectrogram images are constructed, the algorithm was trained with the same parameter settings as before.

Figure 15. The experimental process in DVE.

Table 3. Construction details of the spectrogram images.

Visibility/m	Measurement Distance	Measurement Times	Annotation Situation
1–2	6 m to 2 m	5	No
3–15			Yes

3.2.2. Experimental Results

(1) Comparison with Non-Deep Learning Methods

We compared this method with other non-deep learning methods, including the wavelet transformation method (Wavelet), the EMD direct wavelet thresholding method (EMD-DT), the EMD interval thresholding method (EMD-IT), and the correlation-based EMD partial reconstruction method (EMD-correlation). Since the denoise methods of non-deep learning are all for 1-D signals, we only use ARE and time consuming as the evaluation standard in this section. In order to ensure the fairness of the experiment, the 1-D echo signal closest to the target collected in the multi-distance measurement process is used as the input signal of the non-deep learning method.

It can be seen from Table 4 that, among the non-deep learning methods, compared with the EMD based methods, the wavelet transformation has a faster calculation speed and higher accuracy. The possible reason is that all of these EMD based methods rely on thresholds to eliminate noise, but the threshold settings for these methods are very rough and do not have a good denoising effect on noise. The proposed deep learning recognition method has the highest accuracy, as we do not need to consider setting a threshold for noise, we recognize the target echo signal based on the frequency-distance relation derived from the course of approaching the target.

Table 4. Comparison with non-deep learning methods.

Methods	Wavelet	EMD-DT	EMD-IT	EMD-Correlation	Our Method
ARE(m)	0.32	0.48	0.42	0.45	0.14
Time(s)	0.82	5.27	4.86	4.42	0.75

(2) Comparison with Other Deep Learning Algorithms

We evaluated our method alongside other deep learning algorithms, and the results from different algorithms are shown in Table 5.

Table 5. The results of the different algorithms.

Algorithms	Resolution	AP (%)	ARE (m)	Time(s)
YOLO	227 × 227	0	∞	0.04
	500 × 500	0	∞	0.11
R-CNN	227 × 227	62.9	1.46	6.72
	500 × 500	80.6	0.81	46.65
Fast R-CNN	227 × 227	78.3	1.08	0.95
	500 × 500	89.1	0.34	4.48
Faster R-CNN	227 × 227	82.5	0.84	0.34
	500 × 500	92.3	0.14	0.75

As Table 5 shows, similar to the simulation results, YOLO cannot effectively identify the target echo signal. The performances of 500×500 inputs are better than those of 227×227 inputs. The AP of R-CNN and Fast R-CNN with 500×500 are, respectively, 80.6% and 89.1%, while the Faster R-CNN with 500×500 is 92.3%. The ARE of Faster R-CNN with 500×500 , 0.14 m, is lower than that of R-CNN and Faster R-CNN. It indicates that the Faster R-CNN with 500×500 inputs have the best detection performance, which is consistent with previous simulation conclusion. It should be noted that the experimental detection performance is better than that of simulation result. It is probably because the experiment process has a closer measurement distance compared with the simulation process.

(3) Detection performances in different visibility conditions

We evaluated the detection performances of Faster R-CNN with 500×500 input in different visibility conditions ranging from 1 m to 15 m. The results are shown in Figure 16. Nevertheless, in conditions with visibility of 1 m and 2 m, the algorithm did not recognize the presence of the target echo signal, which was in line with the fact that we have not annotated the ground truth in those images. In this section, the algorithm can recognize the target echo signal at a visibility of 3 m, but in the simulation result it cannot be recognized, the possible reason being that the measurement distance (from 6 m to 2 m) in the experimental process is closer than the simulation process (from 13 m to 3 m).

Figure 16. Detection results in different visibility.

As presented in Figure 16, as the measurement distance is settled, the AP decreases gradually with increasing smoke concentration, while the ARE slightly increases. The results of the experiment prove the conclusions of the previous section. In addition, the detection performances of the test samples are shown in Figure 17.

Figure 17. The detection performances in different visibility conditions: (a) results in a 4 m visibility condition; (b) results in a 14 m visibility condition.

However, as Figures 14 and 17 demonstrate, there are some differences between the simulation and experiment signals that need to be explained. As the target echo signals appear as a continuous ridge shape in the simulation images, while they look like independent mountain peaks in the experimental images. The reason is that the frequency resolution of the experimental signal is higher than the simulated signal. Therefore, the adjacent spectral peaks in the measured data are separated, while they are continuous in the simulated data. These occasions lead to wide main lobe width of the simulation spectrums while the main lobe widths of the experimental spectrums are narrow. In conclusion, the characteristics of target echo signals in the 2-D spectrogram images obtained from the experiment still obey the recognition criterion.

4. Conclusions

In this paper, we have proposed a novel obscurant penetrating detection method for Lidar in DVE. As automatic driving is rapidly developing, our proposed method is applicable to mobile application scenarios. The proposed multi-distance measurement method can collect multiple 1-D echo signals during movement. An effective target echo signal recognition method is also presented, as the 2-D spectrogram images contain dynamic characteristics of 1-D echo signal sequence in the process of approaching the target. Our method requires no denoising and suppression functions to process backscattering signal. Furthermore, we have proposed a recognition algorithm based on deep learning to recognize the target echo signal. The simulation and experimental results have verified that the proposed method can be well trained and utilized for Lidar detection in DVE. In the future, we will transplant the algorithm to Field Programmable Gate Array (FPGA) and consider practical application problems.

Author Contributions: X.L. designed and performed the simulation experiments and writing of the corresponding paragraphs. Z.H. provided many constructive comments. L.L. proofread the manuscript and improved the style of the writing and grammar. Z.T. and B.Y. worked on the experiment. Y.L. conceived of the idea of the whole study,

provided guidance for the experimental testing, and contributed to the drafting and implementation of the manuscript. All authors read and approved the final manuscript.

Funding: This research received no external funding.

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

References

1. Zhang, Y.; Wang, J.; Wang, X.; Dolan, J.M. Road-segmentation-based curb detection method for self-driving via a 3D-LiDAR sensor. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3981–3991. [\[CrossRef\]](#)
2. Heinzler, R.; Piewak, F.; Schindler, P.; Stork, W. CNN-Based Lidar Point Cloud De-Noising in Adverse Weather. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2514–2521. [\[CrossRef\]](#)
3. Hecht, J. Lidar for self-driving cars. *Opt. Photo. News* **2018**, *29*, 26–33. [\[CrossRef\]](#)
4. Wallace, L.; Lucieer, A.; Watson, C.S. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7619–7628. [\[CrossRef\]](#)
5. Starr, J.W.; Lattimer, B.Y. A comparison of IR stereo vision and LIDAR for use in fire environments. *Sensors* **2012**, *8*, 1–4.
6. Ma, R. DEM generation and building detection from lidar data. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 847–854. [\[CrossRef\]](#)
7. Jiang, Z.; Song, C. The Simulation of FMCW Laser Transmission in Smoke Environment. In Proceedings of the 2019 4th International Conference on Control and Robotics Engineering (ICCCE), Nanjing, China, 20–23 April 2019; pp. 175–178.
8. Fritsche, P.; Zeise, B.; Hemme, P.; Wagner, B. Fusion of radar, LiDAR and thermal information for hazard detection in low visibility environments. In Proceedings of the 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR), Shanghai, China, 11–13 October 2017; pp. 96–101.
9. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [\[CrossRef\]](#)
10. Huang, N.E.; Long, S.R.; Zheng, S. The Mechanism for Frequency Downshift in Nonlinear Wave Evolution. *Adv. Appl. Mech.* **1996**, *32*, 59–117.
11. Kopsinis, Y.; McLaughlin, S. Improved EMD using doubly-iterative sifting and high order spline interpolation. *EURASIP J. Adv. Signal Process.* **2008**, *1*, 128293. [\[CrossRef\]](#)
12. Kopsinis, Y.; McLaughlin, S. Empirical mode decomposition based soft-thresholding. In Proceedings of the 2008 16th European Signal Processing Conference, Lausanne, Switzerland, 6 April 2008; pp. 1–5.
13. Kopsinis, Y.; McLaughlin, S. Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding. *IEEE T. Signal Proces.* **2009**, *57*, 1351–1362. [\[CrossRef\]](#)
14. Shamsudin, A.U.; Ohno, K.; Westfechtel, T.; Takahiro, S.; Okada, Y.; Tadokoro, S. Fog removal using laser beam penetration, laser intensity, and geometrical features for 3D measurements in fog-filled room. *Adv. Robot.* **2016**, *30*, 729–743. [\[CrossRef\]](#)
15. Charron, N.; Phillips, S.; Waslander, S.L. De-noising of Lidar point clouds corrupted by snowfall. In Proceedings of the 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 9–11 May 2018; pp. 254–261.
16. Illig, D.W.; Rumbaugh, L.K.; Banavar, M.K.; Bollt, E.M.; Jemison, W.D. Backscatter suppression via blind signal separation for a 532 nm underwater chaotic lidar rangefinder. In Proceedings of the 2015 SPIE, Washington, DC, USA, 19–22 October 2015; pp. 1–6.
17. O’connor, S.; Lee, R.; Mullen, L.; Cochenour, B. In Waveform design considerations for modulated pulse lidar. In Proceedings of the Ocean Sensing and Monitoring VI, Maryland, MD, USA, 5 May 2014; p. 91110P.
18. Djuricic, A.; Jutzi, B. Supporting Uavs in Low Visibility Conditions by Multiple-Pulse Laser Scanning Devices. *Isprs Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-1/W1*, 93–98. [\[CrossRef\]](#)
19. Trickey, E.; Church, P.; Cao, X. Characterization of the OPAL obscurant penetrating LiDAR in various degraded visual environments. In Proceedings of the Degraded Visual Environments: Enhanced, Synthetic, and External Vision Solutions, International Society for Optics and Photonics, Maryland, MD, USA, 29 April 2013; p. 87370E.

20. Starr, J.W.; Lattimer, B.Y. Evidential Sensor Fusion of Long-Wavelength Infrared Stereo Vision and 3D-LIDAR for Rangefinding in Fire Environments. *Fire Technol.* **2017**, 1961–1983. [\[CrossRef\]](#)
21. Bosch, T.; Lescure, M.; Myllyla, R.; Rioux, M.; Amann, M.C. Laser ranging: A critical review of usual techniques for distance measurement. *Opt. Eng.* **2001**, 40, 10–12. [\[CrossRef\]](#)
22. Journet, B.; Bazin, G. A low-cost laser range finder based on an FMCW-like method. *IEEE Trans. Instrum. Meas.* **2000**, 49, 840–843. [\[CrossRef\]](#)
23. Liu, B.; Song, C.; Li, Y.; Duan, Y. A frequency-modulated-continuous-wave laser detection system based on the four-quadrant photodetector. *Appl. Phys. B* **2018**, 124, 186. [\[CrossRef\]](#)
24. Batet, O.; Dios, F.; Comeron, A.; Agishev, R. Intensity-modulated linear-frequency-modulated continuous-wave lidar for distributed media: Fundamentals of technique. *Appl. Opt.* **2010**, 49, 3369–3379. [\[CrossRef\]](#)
25. Stann, B.L.; Ruff, W.C.; Sztankay, Z.G. Intensity-modulated diode laser radar using frequency-modulation/continuous-wave ranging techniques. *Opt. Eng.* **1996**, 35, 3270–3278. [\[CrossRef\]](#)
26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, 39, 1137–1149. [\[CrossRef\]](#)
27. Brusciagioni, P.; Ismaelli, A.; Zaccanti, G. Monte-Carlo calculations of LIDAR returns: Procedure and results. *Appl. Phys. B* **1995**, 60, 325–329. [\[CrossRef\]](#)
28. Liu, B.; Song, C.; Duan, Y. The characteristics simulation of FMCW laser backscattering signals. *Opt. Rev.* **2018**, 25, 197–204. [\[CrossRef\]](#)
29. Lee, H.; Wang, Q.; Shan, B.; Li, X.; Gong, Y.; Zhao, J.; Peng, Z. A New Monte Carlo Simulation Model for Laser Transmission in Smokescreen. In *MATLAB*; SPIE: Washington, DC, USA, 2016.
30. Ryde, J.; Hillier, N. Performance of laser and radar ranging devices in adverse environmental conditions. *J. Field Robot.* **2009**, 8, 712–727. [\[CrossRef\]](#)
31. Cao, X.; Peng, W.; Cai, M.; Bai, X.; Gong, G.; Liu, M.; Qi, J. Region Based CNN for Foreign Object Debris Detection on Airfield Pavement. *Sensors* **2018**, 18, 737. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Uijlings, J.R.R.; Sande, K.E.A.V.D. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, 104, 154–171. [\[CrossRef\]](#)

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).