Received May 8, 2019, accepted June 24, 2019, date of publication July 5, 2019, date of current version July 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927036

ROI-Based LiDAR Sampling Algorithm in on-Road Environment for Autonomous Driving

XUAN TRUONG NGUYEN[1], KHAC-THAI NGUYEN[1], HYUK-JAE LEE[1], (Member, IEEE), AND HYUN KIM[2], (Member, IEEE)

[1]Inter-University Semiconductor Research Center, Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea
[2]Department of Electrical and Information Engineering and the Research Center for Electrical and Information Technology, Seoul National University of Science and Technology, Seoul 01811, South Korea

Corresponding author: Hyun Kim (hyunkim@seoultech.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Korean Government (MSIT) under Grant 2017-0-00721-001 (Development of Intelligent Semiconductor Technology for Vision Recognition Signal Processing for Vehicle Based on Multi-Sensor Fusion) and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A6A1A03032119.

ABSTRACT As the acquisition of laser range measurements such as those from light detection and ranging (LiDAR) sensors requires a considerable amount of time, to design an effective sampling algorithm is a critical task in numerous laser range applications. The state-of-the-art adaptive methods such as two-step sampling are highly effective at handling less complex scenes such as indoor environments with a moderately low sampling rate. However, their performance is relatively low in complex on-road environments, particularly when the sampling rate of the measuring equipment is low. To address this problem, this paper proposes a region-of-interest (ROI)-based sampling algorithm in on-road environments for autonomous driving. With the aid of fast and accurate road and object detection algorithms, particularly those based on convolutional neural networks, the proposed sampling algorithm utilizes the semantic information and effectively distributes samples in the road, object, and background areas. The experimental results demonstrate that the proposed algorithm significantly reduces the mean-absolute-error in the object area by at most 52.8% compared to two-step sampling; moreover, it achieves robust reconstruction quality even at a very low sampling rate of 1%.

INDEX TERMS Autonomous driving, LiDAR sampling, on-road environment, ROI-based sampling, two-stage sampling.

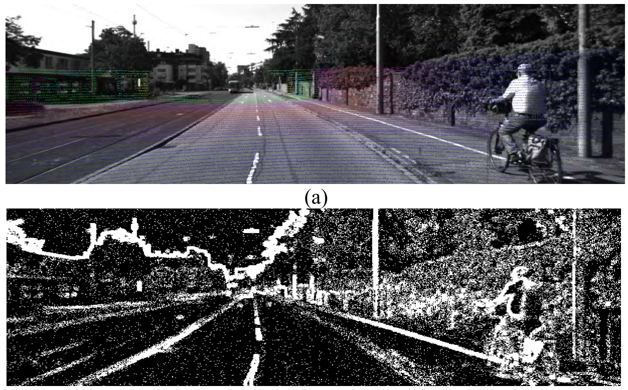
I. INTRODUCTION

In recent years, autonomous driving has become an emerging trend. From the sensing aspect, 3-D cameras such as RGB-D cameras and light detection and ranging (LiDAR) sensors are becoming more affordable, enabling both academic studies and industrial (commercial) applications, such as self-driving cars employing video analytics on LiDAR captured data for path planning as well as obstacle detection [1]. To mimic the complex natural sensing system of humans, a vehicle is installed with different types of sensors such as grayscale/color cameras, inertial and GPS navigation sensors, radio detection and ranging (RADAR), and LiDAR sensors [2], [3]. One of the most critical and challenging tasks in autonomous driving is the generation of a local map of objects (i.e., road, vehicles, and pedestrians) surrounding a car. This task directly relies on the depth sensing technologies. As depth information is crucial for understanding the visual world, many studies have been explored ways to acquire accurate depth information efficiently in both hardware and software systems. In software-based solutions, disparity estimation algorithms using single or stereo cameras have been studied to estimate accurate depth cues in shorter processing time [4]–[6]. In hardware-based solutions, depth sensors such as Microsoft Kinect and LiDAR sensors have been developed to capture better quality depth information with portability and low cost [7]–[9]. Whereas classical stereo vision techniques are only capable of estimating distances of close objects, a LiDAR sensor can produce rich information of a wide and broad field of view (FOV) with a range

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqi Wang.

reaching to 120 meters or more [10]–[12]. However, there are two mainly challenging problems for designing LiDAR sensors and utilizing them to a commercial application. Firstly, despite the progress in active depth sensing, the quality of acquired depth maps is still low when compared against their color image counterparts. Therefore it is highly required to develop efficient and effective depth recovery techniques (especially for improving the spatial resolution of depth cameras), i.e. compressive sensing, image super-resolution or depth completion methods [13]–[17]. Secondly, although a LiDAR is capable of constructing a high-definition map of objects, this requires considerable resources to process and store a large-scale point cloud, requiring resampling or encoding methods [18]–[20]. From the sensor aspect, to solve these two problems, fast and accurate sampling is required to reduce the spatial resolution of a LiDAR sensor, especially for autonomous driving in an on-road environment, where the spatial and temporal resolution of a LiDAR sensor is significantly sparser than that of an image sensor (see Fig. 5(c) and Section IV-A for further details).

This study addresses the problem of finding sampling locations for LiDAR scanners to minimize the reconstruction error in an entire scene or a specific region-of-interest (ROI) for a given sampling budget. This problem is directly related to the compressive sensing theory, which has intensively studied in many decades, and several approaches to find a sampling matrix have been presented [21]–[26]. Motivated by the property of the wavelet transform that the relevant coefficients coincide with discontinuities, Hawe et al. [21] recommend that a data acquisition system should pick samples at the discontinuities or along gradients. However, this approach is not practical for two reasons. First, the gradient of the disparity map is not available prior to sampling. Therefore, all the gradient information needs to be inferred from the color image. Second, the gradient of a color image could be significantly different from that of the disparity map. Thus, inferring the disparity gradient from the color image is challenging. Liu et al. [22] suggest to use outlier elimination prior to edge disparity estimation. Meanwhile, Schwartz et al. [23], [24] propose a saliency-guided sampling approach to perform sampling in a two-stage manner. First, approximately 10% of the samples are sampled randomly, and an approximate depth map is derived from those sampled data. Subsequently, object information or saliency is extracted from the estimated depth to select better locations with the remaining sample budget. Following this approach, L. K. Liu et al. propose a similar two-step sampling in [25]. Particularly, at the pilot stage, half of the sample budget is sampled randomly or along the gradients of a color image. In the second stage, called the refinement stage, sampled points are used to estimate a round disparity map and then to compute locations for the remaining sample budgets. However, these approaches [22]–[25] involve time-consuming rough disparity estimation. In [26], L. K. Liu proposed a sampling framework for acquiring a depth video. Considering the merit of spatial information, this method estimates the motions between two color frames and uses them to compute a depth map from a previous estimated depth frame. A rough estimated map is used as a guide for a gradient-based sampling; so that the sampling procedure is completed in one stage rather than two. However, this method still involves time-consuming disparity estimation. In addition, the method is verified only with relatively simple synthetic datasets, so it becomes relatively convenient to achieve highly accurate motion estimation and sampling. However, in out-door environments; it is challenging to estimate motion accurately, owing to illumination, noise, and other camera factors. Furthermore, the previous sampling schemes [21]–[26] are inappropriate for autonomous driving in on-road environments for the following two reasons. Firstly, as shown in Fig. 1, a scene in an outdoor environment generally consists of a complex background; this caused previous methods to allocate excessively high number of samples into non-interested areas (i.e., trees in Fig. 1). Secondly, it is challenging to obtain a reliable gradient image of a scene in an outdoor scenario because generally, its RGB image is complicated and its raw depth image is too sparse to estimate a reliable gradient map (i.e., $1 \sim 2\%$ sparser than an RGB image).



(a)

(b)

FIGURE 1. Example of a sampling mask in on-road scenario. a) RGB image and LiDAR sampling points. b) Sampling mask for a sampling rate of 25%, sampled points are marked in white, and unsampled ones are marked in black.

To address these two problems, this study proposes a sampling method to distribute samples in road and object regions based on the ratio between their areas. Given a fixed number of samples in road and ROI areas, the method efficiently distributes samples from road to object regions and thus significantly enhances the reconstruction quality of objects. As a single type of sensor is likely to be inadequate for mimicking the sensing system of humans during driving, the main objective of this study is to present a framework exploiting the semantical information from camera sensors to enhance the data acquisition of a LiDAR. The proposed framework addresses a general LiDAR sampling case, where the budget of an entire frame is fixed. In addition, the proposed sampling method utilizes the state-of-the-art Oracle random sampling method in [25]; hence, it is highly efficient and applicable for various scenarios, i.e., on-road environments. To this end, the contributions of this work are as follows:

1. A systematical framework of depth acquisition in an on-road environment is presented. Unlike previous approaches, the proposed scheme detects the objects in a road and segments a scene into background, ROI, and object areas. From the segmentation results, the approach distributes samples across the segmented areas.
2. An ROI-based sampling problem is proposed to optimize a depth sensing system in a LiDAR for an on-road environment. The optimization problem has an optimal solution, which effectively addresses the two problems of the LiDAR discussed above. Experimental results demonstrate that the proposed approach significantly reduces the mean-absolute-error (MAE) in the object area by at most 52.8%. Moreover, it achieves robust reconstruction quality at a very low sampling rate of 1%. In addition, the proposed sampling is remarkably fast (i.e., within a few milliseconds), rendering it applicable to a real-time acquisition system.

The rest of this paper is organized as follows. Section II describes the sampling model and then briefly reviews the related results and their limitations for on-road environments. Section III introduces the proposed sampling approach. The experimental results are presented in Section IV and conclusions and future work are presented in Section V.

II. BACKGROUND

This section briefly describes the definition of a sampling problem and introduces previous studies on gradient-based sampling and two-step sampling algorithms [25], which are the most relevant sampling approaches to this study.

A. SAMPLING MODEL

Let $x \in R^N$ be an $N \times 1$ vector representing a depth map of an entire scene in an FOV of a capturing device such as a LiDAR. For straightforwardness, x is normalized such that $0 \leq x_i \leq 1$ for $i = 1, \ldots, N$. In general, a LiDAR sensor cannot acquire data for all the locations in the FOV so that it reconstructs the depth map of the entire FOV from the sampled data. Let M denote the number of samples from which a capturing device is capable of acquiring data. The *sampling problem* is an optimization problem of selecting samples in the FOV to minimize the reconstruction error with the constraint that the number of the samples satisfies the target budget M. For mathematical formulation, let $\{1, \ldots, N\}$ represent the set of the indexes that correspond to the locations of the entire FOV and $\{i_1, \ldots, i_{\mathrm{M}}\}$ represent the set of the indexes that correspond to the sampling locations among $\{1, \ldots, N\}$.

Problem P1 (Sampling Problem): The sampling problem is to derive $\{i_1, \ldots, i_{\mathrm{M}}\}$ that minimizes the following objective function

$$\min_{i_1,\ldots i_M} \frac{1}{N} \sum_{j=1}^{N} \left\| x_j - \widetilde{x}_j \right\| \tag{1}$$

where $x_1, \ldots, x_N$ are real values and $\widetilde{x_1}, \ldots, \widetilde{x_N}$ are values estimated from the M measurements $x_{i_1}, \ldots, x_{i_M}$.

Generally, it is not feasible to determine a solution in a brute-force search manner. Hence, a heuristic method is likely to be used. The next subsection presents a heuristic algorithm that is called Oracle random sampling or gradient-based sampling derived in [25].

B. GRADIENT-BASED SAMPLING

In [25], a probabilistic model is used to represent the sampling problem. For N locations in an FOV, a diagonal matrix $S \in R^{N \times N}$ is used to represent the sampling operation with the $(i, i)^{\text{th}}$ entry of S being as follows:

$$S_{ii} = \begin{cases} 1, & \text{with probability } p_i, \\ 0, & \text{with probability } 1 - p_i, \end{cases} \tag{2}$$

where $0 \leq p_i \leq 1$ is a pre-defined probability of sampling the i-th location, for $i = 1, \ldots, N$. Given S, the sampled data $b \in R^{N \times 1}$ is defined as follows:

$$b = Sx. \tag{3}$$

where the i^{th} entry b_i is zero if $S_{ii} = 0$. The target budget is defined by the target sampling ratio ξ with $0 < \xi < 1$, which represents the average sampling frequency. Then, the following constraint is obtained:

$$\frac{1}{N} \sum_{i=1}^{N} p_i = \xi. \tag{4}$$

For a large N, the standard concentration inequality guarantees that the average number of ones in S is approximately ξN (i.e., $\xi N = M$) [25].

Let $a = [a_1, \ldots, a_N]^T$ be a vector representing the magnitude of the gradient of the depth map. It can be calculated as follows:

$$a = \nabla x = \sqrt{(D_x x)^2 + \left(D_y x\right)^2}. \tag{5}$$

The intuition underlying the gradient-based sampling method is that the average gradient computed by all the N samples is similar to the average gradient computed from a subset of ξN samples [25]. Let $\left\{p_j\right\}_{j=1}^{N}$ be the optimal sampling probability to define the sampling map S. For a specified sampling ratio ξ and the gradient map, the derivation of the optimal sampling probability $\left\{p_j\right\}_{j=1}^{M}$ is formulated as the following optimization problem:

$$\min_{p_1,\ldots,p_N} \frac{1}{N} \sum_{j=1}^{N} \frac{a_j^2}{p_j} \tag{6}$$

subject to $\frac{1}{N} \sum_j^N p_j = \xi$ and $0 \leq p_j \leq 1$.

The solution is formulated as follows:

$$p_j = \min\left(\tau a_j, 1\right). \tag{7}$$

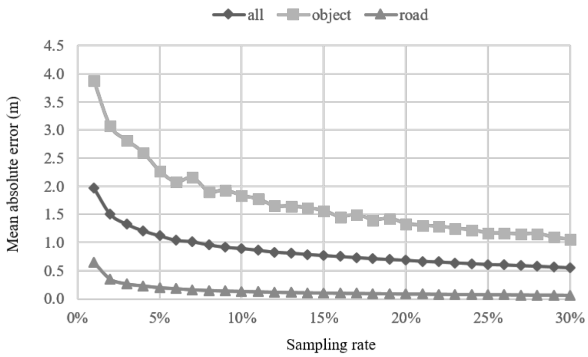


FIGURE 2. Average mean absolute error (m) over sampling rates.

where τ is the solution of $g(\tau) = 0$. Here,

$$g(\tau) = \sum_{j}^{N} \min(\tau a_j, 1) - \xi N. \quad (8)$$

Note that $g(\tau)$ is a piecewise linear and monotonically increasing function, with $g(+\infty) = N(1-\xi)$ and $g(0) \leq 0$. Therefore, τ can be uniquely determined as the root of $g(\tau)$. Moreover, an efficient solution for deriving τ is available.

In practice, the gradients of a depth image are not available prior to sampling. Consequently, a practical sampling method is performed in two steps or utilizes a gradient inferred from an RGB image as described in Section I. However, a vector $a = [a_1, \ldots, a_N]^T$ becomes excessively noisy in a practical outdoor scenario.

III. PROPOSED SAMPLING ALGORITHM

This section presents the main concept and mathematical derivation of the proposed sampling.

A. MOTIVATING EXAMPLES

Before addressing the sampling problem, particularly for an on-road scenario, it is necessary to determine what kind of a sampling pattern should be used in this case. This subsection demonstrates the characteristics of road, object, and overall regions and the utilization of these characteristics to generate a sampling pattern. Fig. 2 shows an average mean-absolute-error (MAE) of test images for specified sampling rates ranging from 1% to 30%. These errors are measured in three regions: an overall scene, road, and object areas. The figure demonstrates that those three errors gradually decrease when the sampling rate increases. In particular, the errors of the overall scene decreases from 1.97 m to 0.55 m when the sampling rate increases from 1% to 30%. In addition, the errors of the road areas is approximately two times that on the overall scene. For example, at the sampling rate of 1%, the MAE error on the object area is approximately 3.88 m, which is approximately two times of that on the overall area (1.97 m). Meanwhile, the MAE error of the road area is relatively small. When the sampling rate increases from 1% to 30%, the error decreases from 0.65 m to 0.06 m. The profiling results in Fig. 2 suggest a strong message: For a given sampling budget, 1) it is likely to give more samples on an object area, especially because this area plays a critical role for self-driving tasks such as obstacle detection and path planning; 2) a sampling budget on a road area is likely to be decreased without a significant degradation on MAE error.

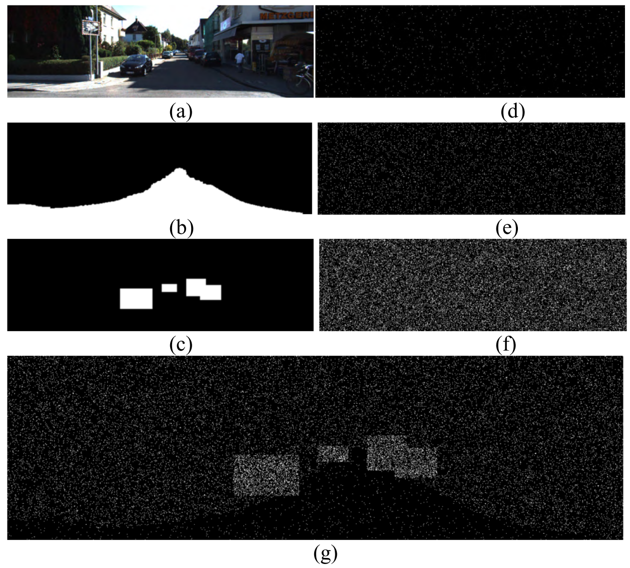


FIGURE 3. Motivational example for ROI-based sampling. Sampled locations are marked with "white", and unsampled ones are marked with "black". (a) RGB image, (b) Road mask, (c) Object mask, (d), (e), (f) Random sampling masks at sampling rates of 1%, 5%, and 20%, respectively, and (g) Expected output sampling pattern.

Fig. 3 demonstrates a sampling pattern constructed under the above characteristics. First, thanks to fast and accurate road and object detection algorithms [27]–[33], it is assumed that a single scene is segmented into road, object, and background regions as shown in Figs. 3(a), (b), and (c), respectively. Meanwhile, three randomly sampling patterns at the sampling rates of 1%, 5%, and 20% are given in Figs. 3(d), (e), and (f), respectively. The expected output sampling pattern shown in Fig. 3(g) is derived as follows. The sampling locations on the road region are obtained by applying locations on the road mask in Fig. 3(b) and those in the random sampling pattern in Fig. 3(d). Next, the locations on the object region are derived from the pattern at the rate of 20% (Fig. 3(f)) with the object mask in Fig. 3(c). Finally, the remaining locations in a background region are selected from the pattern at a sampling rate of 5% (i.e., Fig. 3(e)). The final sampling map is shown in Fig. 3(g); it has a sampling rate of 5.1%, which is close to that in Fig. 3(e). Especially, the MAE error on the object area with the sampling pattern in Fig. 3(g) is approximately 1.39 m, which is significantly smaller than that of the error with the uniformly random sampling pattern in Fig. 3(e) (i.e., 2.49 m). In this particular case, the overall MAE by the desired sampling pattern in Fig. 3(g) is approximately 0.82 m, which is smaller than that given by a uniformly random sampling pattern (0.89 m).

B. AN ROI-BASED SAMPLING ALGORITHM

To address the sampling problem in an on-road environment, a scene is assumed to be segmented into three regions:

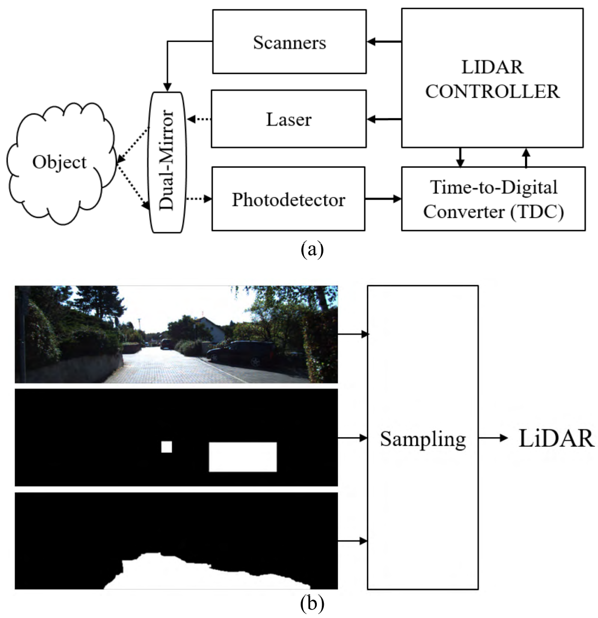


FIGURE 4. (a) Block diagram of a typical LiDAR and (b) System configuration for LiDAR sampling in an on-road environment. It is assumed that object and road masks are specified prior to sampling. The output sampling is used by a LiDAR sensor.

road, object, and background ones. Aided by convolutional neural networks (CNNs), various road and object detection algorithms have been intensively studied in recent years; and consequently, their accuracy and speed have improved significantly [27]–[33]. Numerous road detection approaches submitted to the KITTI road detection benchmark have precisions of over 96%, whereas their runtime is approximately 0.06 s [30]. Considering the effectiveness of available state-of-the-art road and object detection algorithms, it is reasonable to assume that both the road and object masks yielded by an RGB image of a scene are specified prior to the sampling operation in a LiDAR.

A LiDAR usually operates by performing multiple point-wise measurements in a FOV. A block diagram of a LiDAR is illustrated in Fig. 4(a). A typical measuring procedure of the LiDAR system is described as follows. A controller in the LiDAR system starts by computing a target location in the FOV. In the next step, the target position is transmitted to a mechanical scanner that controls motors and mirrors to direct the emitted light. This step requires the communication and motor control time. After the mirror is aimed at the target, the laser diode in the LiDAR system emits a laser beam. Next, the LiDAR waits until the laser reaches an object and its reflected signal arrives at a photodetector. The time interval between the emitted and detected signals is generally referred to as time of flight (TOF) and is denoted as t^{TOF}. Finally, the measurement of t^{TOF} is converted to an electric signal and transmitted to the optical device controller that calculates the TOF from the signal. In the last step, the result is transmitted to the main controller. The sampling problem in LiDAR aims to determine sampling locations for its controller. Fig. 4(b) illustrates a system configuration to generate a sampling pattern for LiDAR in an on-road environment. To compute a sampling map, three input images are an RGB image, an object mask, and a road mask. The sampling procedure operates as follows: First, an RGB image is captured, requiring approximately 16.7 ms for a high-speed 60-fps camera sensor. Next, both the object and road masks are derived from the specified RGB image, requiring less than 20 ms because the object and road detections can be performed at the frame rate of 50 fps [28], [29], [32]. Finally, a sampling pattern is derived in 1 ms. Based on the computed pattern, sampling locations are given to the LIDAR controller which moves scanners to the desired positions, and measures distances.

With this configuration, the sampling problem in a LiDAR is derived as follows. Let S_B, S_R, and S_O denote the index sets of the points in background, road, and object areas, respectively. Thus, the union of three sets is equal to $\{1, \ldots, N\}$, and an intersection of any two sets is an empty set.

$$S_B \cup S_R \cup R_O = \{1, \ldots, N\}. \tag{9a}$$

$$S_B \cap S_R = S_B \cap S_O = S_O \cap S_R = \emptyset. \tag{9b}$$

The points in the background, road, and object areas have different characteristics so that they are likely to be sampled with different priorities. Let λ_B, λ_R, and λ_O be weighted-parameters representing sampling priorities for the background, road and object areas, respectively. Recall that a road area is generally flat so that the MAE does not change significantly as the sampling rate changes. Therefore, the parameter λ_R for a road area is likely to be smaller than the one for a background area. Meanwhile, points in an object area are more likely to be sampled rather than those in a background area. Considering these, λ_B, λ_O and λ_R must satisfy the following equations:

$$\lambda_R = \alpha \times \lambda_B \tag{10}$$

$$\lambda_O = \beta \times \lambda_B \tag{11}$$

where α and β are constant and $\alpha \leq 1$, $\beta \geq 1$. Therefore, the sampling problem (6) is modified as follows:

Problem P2 (Sampling Problem for a LiDAR): Given α and β, the sampling problem is to derive $\{i_1, \ldots, i_\mathrm{M}\}$ that minimizes the following objective function:

$$\min_{i_1,\ldots i_M} \frac{1}{N}\left(\sum_{i\in S_B} ||x_i - \widetilde{x}_i|| + \alpha \sum_{j\in S_R} ||x_j - \widetilde{x}_j|| + \beta \sum_{k\in S_o} ||x_k - \widetilde{x}_k||\right) \tag{12}$$

where $x_1, \ldots, x_N$ are real values and $\widetilde{x_1}, \ldots, \widetilde{x_N}$ are the values estimated from the M measurements $x_{i_1}, \ldots, x_{i_M}$.

Compared to Problem P1 in Section II.A, Problem P2 has additional weight parameters α and β Apparently, when α and β are equal to one, the objective function in (10) becomes that in (1). That is, the background, road, and object areas have identical sampling priority. As mentioned above, it is not feasible to determine a solution in a brute-force search manner; therefore, a heuristic method is necessary. Similar to the derivation in Section II.A, Problem 2 is transformed into a gradient-based sampling in the following section.

C. MODIFIED GRADIENT-BASED SAMPLING

For consistency, let $a = [a_1, \ldots, a_N]^T$ be a vector indicating the prior information of the depth map. It should be noted that the vector a may be defined as the gradient of the depth map as shown in (5) in Section II.B. However, it represents any prior information. Using the probabilistic model in [15], the gradient-based sampling is modified as follows. For a given sampling ratio ξ, the prior map, and parameters α and β, the derivation of the optimal sampling probability $\{p_j\}_{j=1}^{M}$ is formulated as the following optimization problem:

$$\min_{p_1,\ldots p_N} \frac{1}{N}\left(\sum_{i\in S_B}\frac{a_i^2}{p_i} + \alpha\sum_{j\in S_R}\frac{a_j^2}{p_j} + \beta\sum_{k\in S_O}\frac{a_k^2}{p_k}\right) \tag{13}$$

subject to $\frac{1}{N}\sum_j^N p_j = \xi$ and $0 \le p_j \le 1$; here, S_B, S_R, and S_O are defined as (9a) and (9b).

In practice, a gradient generally becomes excessively noisy in an on-road environment so that it is challenging to generate a ''desired'' sampling pattern from the problem in (6). Therefore, this study considers an extreme and widely used case where the gradient information is unspecified or is under the following condition:

$$a_1 = a_2 = \cdots = a_N = 1 \tag{14}$$

D. α, β–DISTORTION OPTIMIZATION PROBLEM

Obviously, given weight parameters α and β, solving (13) is similar to solving (6) with an exaction solution described in (7) and (8). This subsection presents an optimization problem to derive these parameters.

Problem P3 (α, β – Distortion Problem): Parameters α, β are derived by solving the following optimization problem:

$$\min_{\alpha,\beta} MAE_{obj} + \lambda MAE_{all} \tag{15}$$

where MAE_{obj} and MAE_{all} are reconstruction errors on road and overall regions, respectively; and λ is a weight factor.

Problem P3 clearly demonstrates that for α, β selection, it is necessary to consider both a quality enhancement on an ROI and a degradation on the other. However, similar to the sampling problem, it is a chicken-and-egg problem so that it is not feasible to determine a solution by using brute-force search. In practice, a numerical solution is obtained as shown in the following subsection IV.B.

IV. EXPERIMENTAL RESULTS

This section describes the experimental environments and then demonstrates the results of the proposed sampling algorithm. Firstly, Sections V.A and B describe the datasets and experimental results of the proposed ROI-based sampling method with the parameters α and β. Subsequently, a detailed comparison with the previous results is presented. Sampling rate ξ is set to either 5%, 10%, 15% or 20%, and 18 datasets in the KITTI dataset are used in the experiments.

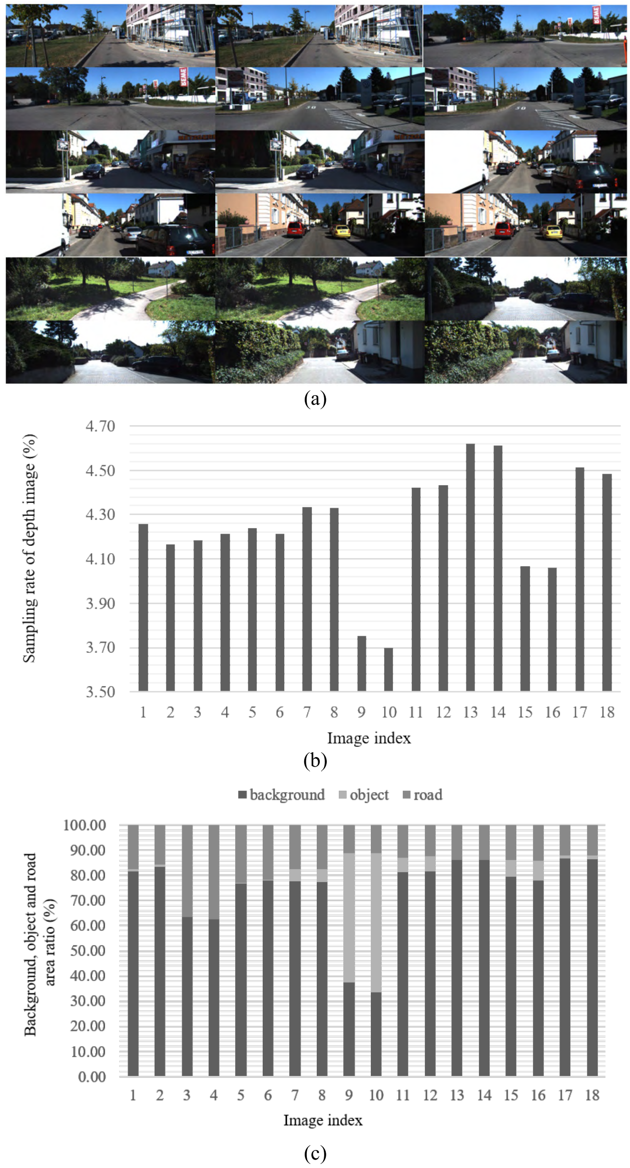


FIGURE 5. Eighteen images acquired from KITTI datasets [2], [3]. (a) Color images, (b) Sampling rates of depth images, and (c) Background, object, and road area ratio.

A. DATASETS

A set of 18 test images is acquired from the well-known KITTI dataset [2], [3] to cover typical on-road scenarios. The set profiles are included in Fig. 5. Firstly, Fig. 5(a) shows the captured color images of these datasets; these images display various typical on-road scenarios such as a clear scene without any object and a scene with close and distant vehicles. Fig. 5(b) shows the sampling rates of the groundtruth depth images. A depth image is obtained by projecting a point cloud data provided by a LiDAR sensor into a color image domain [2], [3]. However, the number of measurements of the LiDAR

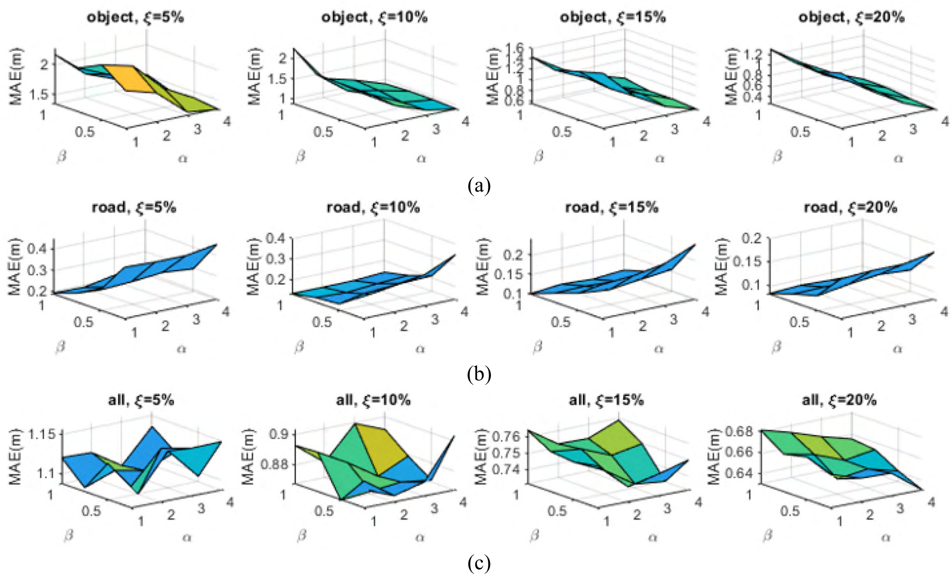


FIGURE 6. MAEs at the sampling rate of 5%, 10%, 15%, and 20% on different regions: (a) Object, (b) Road, and (c) Overall.

(i.e., a 64-channel Velodyne LiDAR) is significantly sparser than that of an RGB image. For these datasets, the sampling rate is approximately 4.26% on an average and ranges from 3.70% to 4.62%, as shown in Fig. 5(b). It is noteworthy that the sampling method in [25] is based on a wavelet, and the contour-based reconstruction method performs ineffectively when the sample budget is small (i.e., 1% or 2%) [14]. That is, it is challenging to obtain a reliable gradient image of a scene in an outdoor scenario because generally, its RGB image is complicated and its raw depth image excessively sparse to estimate a reliable gradient map (i.e., 1% $\sim$ 2% sparse compared to an RGB image).

In addition, Fig. 5(c) demonstrates the profiling area ratios of the background, road, and object regions in the datasets. In particular, the ratio of the object areas is approximately 8.24% on average and ranges from 0% to 55.7%. Meanwhile, the ratio of the road areas is approximately 17.24% on average and ranges from 10.94% to 37.5%. The remaining background occupies an average area ratio of 74.33% and ranges from 33.35% to 86.81%.

B. FINDING PARAMATERS α AND β

This subsection presents the experimental results with various values of the parameters α and β. As the sampling rate is set to 5%, 10%, 15%, or 20%, the minimum value of α is set as 0.25 to maintain a considerable sampling rate in a road area (approximately 1.25%, 2.5%, 3.75%, or 5%, respectively). Meanwhile, the maximum value of β is set as four as the maximum sampling rate in the object regions is approximately 80% (= 20% $\times$ 4). In particular, α is set to 0.25, 0.5, 0.75, or one, whereas β is set to one, two, three, or four. Fig. 6(a), Fig. 6(b), and Fig. 6(c) show the average MAEs of the object, road, and entire areas, respectively. The first, second, and third rows display the MAEs of the object, road, and overall areas, respectively. In each row, the first, second, third, and fourth columns display the MAEs at the sampling rates of 5%, 10%, 15%, and 20%, respectively. It should be noted that the MAEs on the object areas significantly decrease when the value of the parameter β increases. For example, considering the baseline sampling $\alpha = \beta = 1$ and the sampling $\alpha = 0.25$, $\beta = 4$. Note that with $\alpha = \beta = 1$, the sampling method becomes a uniform random sampling. The MAEs on the object area with this setting are shown in Fig. 6(a); here, the MAEs decrease by 0.83 m, 0.97 m, 0.87 m, and 0.48 m at the sample rsates of 5%, 10%, 15%, and 20%, respectively. Meanwhile, the overall MAEs shown in Fig. 6(c) do not change significantly when the MAE differences between the various parameters shown in the third row of Fig. 6 are likely less than 0.1 m in all the cases. In addition, the average MAEs on the road area shown in Fig. 6(b) increase to an average of approximately 0.18 m and ranges from 0.1 m to 0.26 m. To reflect the importance of the object regions, $\beta = 4$ is used for comparison. Meanwhile, the lower and upper values of α are experimentally selected as 0.25 and one, respectively. These experimental results demonstrate that the proposed sampling method with $\alpha = 0.25$, $\beta = 4$ significantly improves the reconstruction quality on the object area. Similarly, the proposed sampling with $\alpha = 1$, $\beta = 4$ also achieves a large improvement on the road area compared to the baseline. In this case, the average MAE on the road areas decrease to an average of approximately 0.84 m and ranges from 0.67 m to 0.98 m for various sample rates. It should be noted that 0.84 m is highly critical when compared to the sizes of the cars on the road.

Fig. 7 demonstrates the subjective comparison between the proposed sampling with $\alpha = 0.25$, $\beta = 4$ and the uniformly

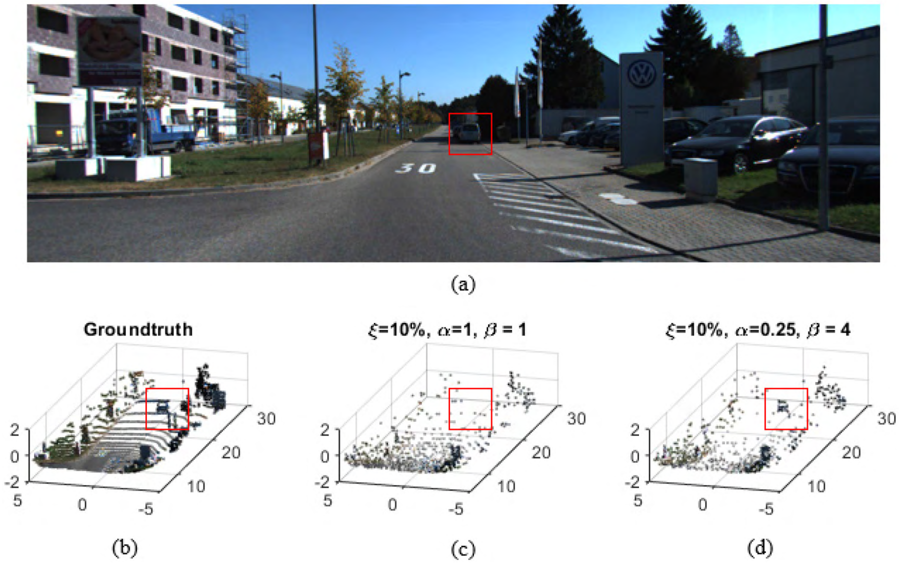


FIGURE 7. Subjective comparison between the proposed sampling ($\alpha = 0.25$, $\beta = 4$) with the baseline ($\alpha = 1$, $\beta = 1$) or uniformly random sampling. (a) RGB image, (b) Ground truth point cloud, (c) Samples obtained by random sampling at the sampling rate 10%, and (d) Samples obtained by the proposed sampling.

random sampling at the sampling rate of 10%. Fig. 7(a) shows the color view of a scene, because a car is considered as an object of interest. The 3D point cloud view of a scene is displayed in Fig. 7(b). Meanwhile, the sampling points of the random sampling and the proposed method are shown in Fig. 7(c) and (d), respectively. It is apparent that the number of samples in Fig. 7(c) and (d) are substantially less than that in Fig. 7(b); this results in a substantial reduction in memory storage and computational power in an actual case. However, visually, the samples in Fig. 7(d) appear significantly better than those in Fig. 7(b) because the samples surrounding the car are effectively captured in this sampling. It should be noted that only very few points of the car are captured in Fig. 7(c); this is likely to cause ambiguity in obstacle avoidance or path planning in practical autonomous driving. This subjective comparison clearly demonstrates the advantage of the proposed ROI-based sampling over the uniformly random sampling.

C. OBJECTIVE AND QUATITATIVE EVALUATIONS

This subsection compares the proposed sampling method with three previous sampling approaches: random sampling, color-image-guided sampling [21], and two-stage sampling [25]. To reflect the importance of the object regions, $\beta = 4$ is used for comparison. The lower and upper values of α are selected as 0.25 and one, respectively. In particular, for the proposed method, two settings are $\alpha = 1$, $\beta = 4$ and $\alpha = 0.25$, $\beta = 4$. It should be noted that the sampling methods in [11] and [25] are used with synthetic disparity datasets rather than on-road KITTI datasets [2], [3]. Therefore, for a fair comparison, they are modified to be used for KITTI datasets. In particular, the color-image-guided sampling [21] uses a half of the sample budget used for random sampling.

TABLE 1. Comparisons of MAEs (M) among various sampling algorithms on object area at sampling rates of 5%, 10%, 15%, and 20%.

	5%	10%	15%	20%
Random	2.176	1.871	1.433	1.315
Color image-guided [21]	2.306	1.702	1.350	1.136
Two-stage [25]	1.981	1.230	0.819	0.581
Proposed $\boldsymbol{\alpha = 1, \beta = 4}$	1.506	0.892	0.667	0.365
Proposed $\boldsymbol{\alpha = 0.25, \beta = 4}$	**1.350**	**0.900**	**0.561**	**0.274**

For the remaining budget, it computes the gradient of a gray image and computes the remaining locations based on the gradient-based sampling in Section II.B. Moreover, the two-stage sampling [25] is modified by using the interpolation method in Matlab for estimating a depth map because the reconstruction quality of the method in [25] is excessively low when the sparsity of the depth map is approximately 1–4%.

The comparison of the MAEs on the object, road, and overall areas are reported in Tables 1, 2, and 3, respectively. On each table, the first, second, and third rows display the experimental results with uniformly random sampling, color-image-guided sampling [21], and two-stage sampling [25], respectively; the fourth and fifth rows display the results with two variations in the proposed methods. In each row, the second, third, fourth, and fifth columns report the results with the sampling rates of 5%, 10%, 15%, and 20%, respectively. Table 1 shows that the proposed sample achieves the highest performance among all the methods on the object area, which is critical in autonomous driving. Compared to [21], the variation in the proposed method in the fifth row reduces the MAE

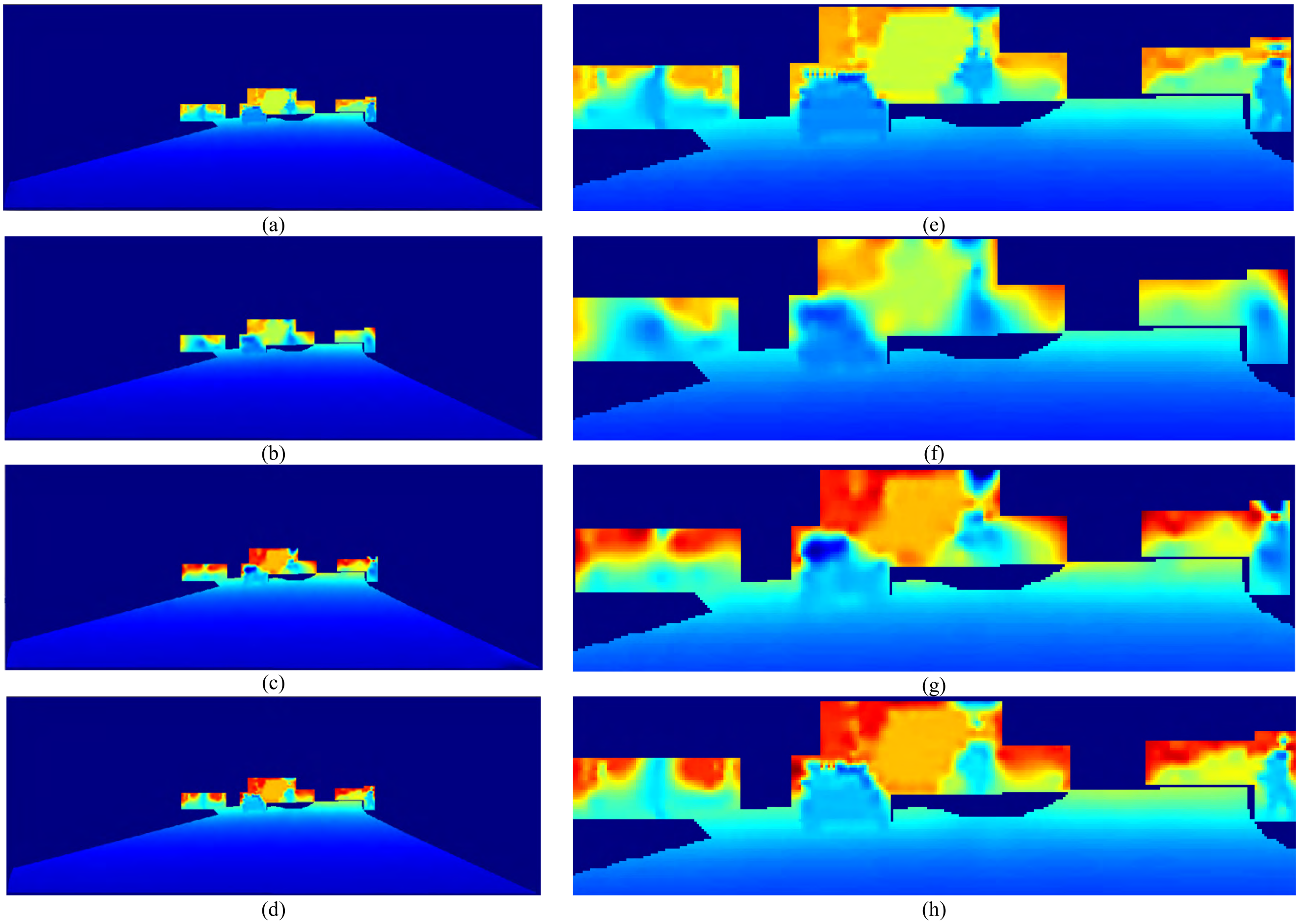


FIGURE 8. Example of reconstruction results at road and object areas by various sampling methods. (a) raw depth; (b) random sampling, (c) two-step sampling [25]; (d) this work and (e), (f), (g), and (h) the zoom-out results of the object areas from of (a), (b), (c), and (d), respectively.

TABLE 2. Comparisons of MAEs (m) among various sampling algorithms on road area at sampling rates of 5%, 10%, 15%, and 20%.

	5%	10%	15%	20%
Random	0.193	0.135	0.100	0.083
Color image-guided [21]	0.211	0.137	0.103	0.084
Two-stage [25]	**0.189**	**0.116**	**0.086**	**0.067**
Proposed $\boldsymbol{\alpha = 1, \beta = 4}$	0.211	0.144	0.110	0.083
Proposed $\boldsymbol{\alpha = 0.25, \beta = 4}$	0.454	0.350	0.243	0.183

by 0.852 m on average, ranging from 0.79 m to 0.955 m. A reconstruction error or an MAE is commonly used to verify the effectiveness of a sampling pattern. Objects refer to cars, trucks, or pedestrians, whose size is relatively small. Thus, a reduction in the MAE by approximately 1 m is significant.

This paper also presents a visual example in Fig. 7. Apparently, the absence of measurements in the object areas in Fig. 7(c) causes significant degradation of an MAE. When compared to the two-stage sampling [25], the variation of the proposed method reduces the MAE by 0.382 m on an average (ranging from 0.258 m to 0.631 m). That is, the proposed method achieves 35.75% reduction on an average (ranging from 26.84% to 52.8%) in the MAE on the object area when compared to [25]. Tables 2 and 3 present the MAEs on the road and overall areas of less importance. The results indicate that the two-stage sampling [25] achieves the highest performance with these criteria. However, compared to [25], the proposed method decreases the MAE on the road and entire areas by at most 0.265 m and 0.194 m, respectively. It should be noted that the MAE degradation mainly occurs on the background region; it consists of trees or buildings, which are less important for obstacle detection and localization. Hence, the proposed sampling method provides an effective trade-off between errors on the object and those on the remaining areas. In addition, the proposed sampling algorithm has significantly low complexity so that it can be performed in 1 ms.

The final reconstructed images are shown in Fig. 8. Fig. 8(a) shows the reconstruction result from a raw depth map. Note that a raw depth is also sparse. As demonstrated in Fig. 8(b), (c), and (d), the road areas are effectively reconstructed for all the three methods, i.e., random sampling, modified two-stage sampling [25], and this work, respectively. In addition, the proposed algorithm generates the

TABLE 3. Comparisons of MAEs (m) among various sampling algorithms on overall area at sampling rates of 5%, 10%, 15%, and 20%.

	5%	10%	15%	20%
Random	1.121	0.893	0.764	0.682
Color image-guided [21]	1.123	0.877	0.735	0.627
Two-stage [25]	**1.041**	**0.748**	**0.573**	**0.437**
Proposed $\boldsymbol{\alpha = 1, \beta = 4}$	1.137	0.891	0.759	0.655
Proposed $\boldsymbol{\alpha = 0.25, \beta = 4}$	1.149	0.903	0.750	0.631

closest output compared to the upper-bound performance, whereas the random sampling yields the lowest performance. Figs. 8(e), (f), (g), and (h) demonstrate the reconstruction results and zoomed-out object regions from a raw depth image, random sampling, modified two-stage sampling [25], and this work, respectively. Evidently, the proposed sampling method yields the best performance on an object area. Meanwhile, the reconstruction errors on the road areas are visually similar among all methods.

V. CONCLUSION

In this paper, a concept of using results from object and road detection in an advanced-driver-assistance-systems (ADAS) is introduced to more effectively guide the sampling process in a LiDAR system, specifically increasing the sampling rate on the ROI regions. The main concept is to distribute the sampling budget of a LiDAR across the road and object areas through ROI detection. As a result, the proposed algorithm is highly optimized for typical ADAS, and consequently, it offers better results in both objective and quantitative evaluation. Therefore, the proposed system is expected to contribute greatly to the generalization of the autonomous driving systems.

REFERENCES

[1] B. Prabhakaran, Y.-G. Jiang, H. Kalva, and S.-F. Chang, "Editorial IEEE transactions on multimedia special section on video analytics: Challenges, algorithms, and applications," *IEEE Trans. Multimedia*, vol. 20, no. 5, p. 1037, May 2018.

[2] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[4] W.-S. Jang and Y.-S. Ho, "Efficient disparity map estimation using occlusion handling for various 3D multimedia applications," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1937–1943, Nov. 2011.

[5] Z. Lee and T. Q. Nguyen, "Multi-array camera disparity enhancement," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2168–2177, Dec. 2014.

[6] R. Ranftl, V. Vineet, Q. Chen, and V. Koltun, "Dense monocular depth estimation in complex dynamic scenes," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4058–4066.

[7] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[8] B. Schwarz, "LIDAR: Mapping the world in 3D," *Nature Photon.*, vol. 4, no. 7, pp. 429–430, 2010.

[9] D. S. Hall, "High definition LiDAR system," U.S. Patents 7 969 558, B2, Jun. 28, 2011.

[10] RIEGL USA Inc. (2017). *VUX-1UAV Survey-Grade Unmanned Laser Scanners for UAS/UAV-Based Laser Scanning*. (Accessed: Mar. 31, 2019). [Online] Available:http://products.rieglusa.com/product/all-categories-unmanned-scanners/vux-1uav-survey-grade-unmanned-laser-scanners

[11] Velodyne LiDAR Inc. (2017). *Velodyne LiDAR HDL-64E S3: High Definition Real-Time 3D LiDAR Sensor*. (Accessed: Mar. 31, 2019). [Online]. Available: http://velodynelidar.com/hdl-64e.html

[12] X. T. Nguyen, Van L. Dinh, H.-J. Lee, and H. Kim, "A high-definition LIDAR system based on two-mirror deflection scanners," *IEEE Sensors J.*, vol. 18, no. 2, pp. 559–568, Jan. 2018.

[13] J. Xie, R. Feris, S.-S. Yu, and M.-T. Sun, "Joint super resolution and denoising from a single depth image," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1525–1537, Sep. 2015.

[14] F. Ma, L. Carlone, U. Ayaz, and S. Karaman, "Sparse sensing for resource-constrained depth reconstruction," in *Proc. IEEE Int. Conf. Intell. Robot. Syst. (IROS)*, Oct. 2016, pp. 96–103.

[15] J.-I. Jung and Y.-S. Ho, "Depth image interpolation using confidence-based Markov random field," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1399–1402, Nov. 2012.

[16] J. Uhrig, N. Schneider, L. Schneider, T. Brox, U. Franke, and A. Geiger, "Sparsity invariant CNNs," in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2017, pp. 11–20.

[17] Y. Song and Y.-S. Ho, "High-resolution depth map generator for 3D video applications using time-of-flight cameras," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 386–391, Nov. 2017.

[18] Y. Chen, X. Zhao, L. Zhang, and J.-W. Kang, "Multiview and 3D video compression using neighboring block based disparity vectors," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 576–589, Apr. 2016.

[19] S. Chen, D. Tian, C. Feng, A. Vetro, and J. Kovačević, "Fast resampling of three-dimensional point clouds via graphs," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 666–681, Feb. 2018.

[20] K. Bahirat, S. Raghuraman, and B. Prabhakaran, "Real-time, curvature-sensitive surface simplification using depth images," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1489–1498, Jun. 2018.

[21] S. Hawe, M. Kleinsteuber, and K. Diepold, "Dense disparity maps from sparse disparity measurements," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2126–2133.

[22] L.-K. Liu, Z. Lee, and T. Nguyen, "Sharp disparity reconstruction using sparse disparity measurement and color information," *IVMSP Workshop*, Jul. 2013, pp. 1–4.

[23] S. Schwartz, A. Wong, and D. A. Clausi, "Multi-scale saliency-guided compressive sensing approach to efficient robotic laser range measurements," in *Proc. 9th Conf. Comput. Robot Vis.*, May 2012, pp. 1–8.

[24] S. Schwartz, A. Wong, and D. A. Clausi, "Saliency-guided compressive sensing approach to efficient laser range measurement," *J. Vis. Commun. Image Represent.*, vol. 24, no. 2, pp. 160–170, Feb. 2013.

[25] L.-K. Liu, S. H. Chan, and T. Q. Nguyen, "Depth reconstruction from sparse samples: Representation, algorithm, and sampling," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1983–1996, Jun. 2015.

[26] L.-K. Liu and T. Q. Nguyen, "A framework for depth video reconstruction from a subset of samples and its applications," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4873–4887, Oct. 2016.

[27] F. Yang, W. Choi, and Y. Lin, "Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2129–2137.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

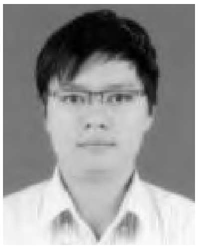
[29] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.

[30] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[31] G. L. Oliveira, W. Burgard, and T. Brox, "Efficient deep models for monocular road segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4885–4891.

[32] Z. Chen and Z. Chen, "RBNet: A deep neural network for unified road and road boundary detection," in *Proc. Int. Conf. Neural Inf. Process.*, Nov. 2017, pp. 677–687.

[33] *KITTI Road Benchmark*. Accessed: Mar.31, 2019. [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_road.php



XUAN TRUONG NGUYEN received the B.S. degree in electrical engineering from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2011, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2015 and 2019, respectively, where he is currently a Postdoctoral Fellow with the BK21+ of Electrical and Computer Engineering Department. His research interests include the areas of algorithm, and SoC design for low-complexity computer vision and multimedia applications.

KHAC-THAI NGUYEN received the B.S. degree in electronics and telecommunications from the Hanoi University of Science and Technology, Vietnam, in 2014, and the M.S. degree in electrical engineering and computer science from Seoul National University, South Korea, in 2018. He is currently a System Software Engineer with Google.

HYUK-JAE LEE received the B.S. and M.S. degrees in electronics engineering from Seoul National University, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 1996. From 1998 to 2001, he was with the Server and Workstation Chipset Division, Intel Corporation, Hillsboro, OR, USA, as a Senior Component Design Engineer. From 1996 to 1998, he was on the faculty of the Department of Computer Science, Louisiana Tech University, Ruston, LA, USA. In 2001, he joined the School of Electrical Engineering and Computer Science, Seoul National University, South Korea, where he is currently a Professor. He is the Founder of Mamurian Design, Inc., a fabless SoC design house for multimedia applications. His research interests include the areas of computer architecture and SoC design for multimedia applications.

HYUN KIM received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2009, 2011, and 2015, respectively. From 2015 to 2018, he was with the BK21 Creative Research Engineer Development for IT, Seoul National University, Seoul, as a BK Assistant Professor. In 2018, he joined the Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul, where he is currently an Assistant Professor. His research interests include the areas of algorithm, computer architecture, memory, and SoC design for low-complexity multimedia applications and deep neural networks.