

Master of Science Thesis

**ROI-based LiDAR sampling in
on-road environment for autonomous
driving**

August 2018

**Graduate School of Engineering
Seoul National University
Electrical and Computer Engineering Major
Nguyen Khac Thai**

ROI-based LiDAR sampling in on-road environment for autonomous driving

Supervisor: Prof. Hyuk-Jae Lee

This work is submitted as a Master of Science thesis

August 2018

**Graduate School of Engineering
Seoul National University
Electrical and Computer Engineering Major**

Nguyen Khac Thai

**Confirming the master's thesis written by
Nguyen Khac Thai
August 2018**

Chair Prof. Choi, Kiyoun_____

Vice Chair Prof. Lee, Hyuk-Jae_____

Examiner Prof. Cho, Nam-Ik_____

Abstract

The acquisition of laser range measurements (i.e. Light Detection and Ranging sensor (LiDAR)) can be a time-consuming process if a high spatial resolution is required. Hence, designing an effective sampling algorithm is essential for many laser range applications. Previous approaches, such as two-step sampling, can be useful in general situations involving indoor and less complex scenes. However, they show a deficiency in the outdoor complex environment, especially in the condition of a very low sampling rate. To address this problem, this paper proposes an ROI-based sampling algorithm in the road environment, the typical environment for ADAS (advanced driver-assistance systems). Taking the merit of existing road and object detection algorithms, i.e. YOLO, the proposed method utilizes the semantic information and effectively distribute the sample budget to maximize the reconstruction quality, especially in the objects area. Experimental results show that the proposed method significantly reduces the mean-absolute-error (MAE) in the objects area and in the overall ROI by 44.9% and 15.1% compared to the two-step sampling. In addition, it achieves robust reconstruction quality with a very low sampling rate (i.e. 1% in experiments).

Keyword: ADAS, advanced driver-assistance system, Light Detection and Ranging sensor, LiDAR, sampling algorithm, sampling strategy

Student Number: 2016-24658

Table of Contents

Abstract.....	I
List of Figures.....	V
List of Table.....	VII
CHAPTER 1 - INTRODUCTION.....	1
1.1. Overview	1
1.2. LiDAR and LiDAR sampling.....	2
1.3. Thesis outline.....	3
CHAPTER 2 - BACKGROUND AND RELATED WORKS	4
2.1. Background.....	4
2.1.1 Sensor systems in ADAS.....	4
2.1.2 Calibration between LiDAR and camera.....	6
2.1.3 Object detection algorithms	9
2.1.4 Iterative Closest Point algorithm	9
2.2. Related works	11
2.2.1 Depth-gradient-guided sampling	11
2.2.2 Color-image-guided sampling	13

CHAPTER 3 - SAMPLING ALGORITHM	15
3.1 System Workflow.....	15
3.2 System configuration.....	16
3.3 Distribution of budget inside ROI.....	17
3.3.1 Reconstruction characteristics of each area in ROI.....	17
3.3.2 Distribution budget between road and objects area.....	23
3.4 Distribution budget between ROI and background area	24
CHAPTER 4 - EXPERIMENTAL RESULTS.....	27
5.1. Dataset	27
5.2. Evaluation inside ROI	30
5.2.1 Objective evaluation	30
5.2.2 Quantitative evaluation.....	36
5.3. Evaluation in the entire frame	37
CHAPTER 5 - CONCLUSION.....	42
REFERENCES.....	43

List of Figures

Figure 1. A test system with a LiDAR and a camera in my lab	5
Figure 2. Calibration result. Color indicates the distance. Occlusion is observed due to a different viewpoint.	7
Figure 3. Mean projection error according to the number of samples used for optimization	8
Figure 4. Workflow of two-step sampling.....	12
Figure 5. Workflow of the proposed system	15
Figure 6. Road and objects mask generated from KITTI dataset	16
Figure 7. Reconstruction characteristics of the road area.....	18
Figure 8. MAE from 10 random scans	19
Figure 9. Reconstruction characteristics of objects area	19
Figure 10. Error images after reconstruction. Blue indicates smaller error. More red means higher error.	20
Figure 11. Fitting results of MAE of road and objects area	22
Figure 12. Localization results on dependence of β	25
Figure 13. Raw scan and ground truth depth from KITTI.....	27

Figure 14. Raw depth and ground truth in moving object area	28
Figure 15. From top to bottom: masks of raw depth and sampling mask of random sampling, two-step sampling and proposed algorithm	31
Figure 16. Depth ground truth and the reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm	33
Figure 17. Zoomed out of objects area in depth ground truth and the reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm	34
Figure 18. Error images from reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm	35
Figure 19. Sampling maps comparison of different methods.....	39

List of Table

Table 1. Sampling rate inside ROI and ratio between road and objects area	29
Table 2. Number of samples inside road and objects area	30
Table 3. Mean Absolute Error comparison between methods.....	36
Table 4. Odometry results of various setups	38
Table 5. Reconstruction results of various setups	40

CHAPTER 1 - INTRODUCTION

1.1. Overview

The development of advanced driver-assistance systems, or ADAS, is considered the next logical step of the car industry. It is seen as the new revolution that changes the way people live and transport. Not only granting more personal freedom, allow people to free their hands and mind while driving, self-driving cars are also expected to vastly reduce road accidents, congestion, pollution; and eliminate the huge cost of owning personal vehicles when integrating with blooming share-driving services [1]. There are five self-driving levels, starting with level 0 of no automation to level 5 of complete automation where drivers simply just enter the destination and let the systems do the rest of work. While the ultimate goal is a safe and fully autonomous driving system, current commercial systems now can constantly sense the surrounding environments and immediately alert the drivers of danger, traffic situations, and road conditions. Being considered a key to the future, ADAS have been actively developed in recent year [2] [3] not just in academia but many companies are competing to release the first commercial system.

In the stream of those developments, object detection algorithms have been experiencing rapid developments. Their accuracy and speed have been

improved significantly over time. Among them, YOLO is a real-time model which is fast, efficient, optimized end-to-end and thus provides better efficiency with respect to the computational cost [4]. YOLOv3, which could have the frame rate up to 40-50fps, is much faster than the typical frame rate of LiDAR systems. Besides, the road detection, which could be considered as an easier problem has also been solved effectively. Many road detection approaches submitted to the KITTI road detection benchmark have the precision of more than 96% while the runtime is just about 0.06s [5].

1.2. LiDAR and LiDAR sampling

In ADAS, the sensor system is the critical part which let the car monitor the surrounding environment. For the task of detecting obstacles on the road, cameras and light detection and ranging sensors (LiDAR), such as Velodyne LiDAR, have been proved to be valuable in these systems. LiDAR is the range sensing sensor that uses laser pulses to estimate distances to surrounding objects by measuring the time of flight of the laser beams. The quality of the measurements is affected by the reflection property and the angle of the reflecting surface. Many measurements are missing because laser beams do not return. It is common for a LiDAR system to have more than 1 laser unit, ie. the Velodyne HDL-64E has 64 lasers that can measure up to more than 2.2 million points per second [6]. In our lab, a LiDAR model has also been

developed [7]. Having the same problem with other systems, that LiDAR is limited to 100.000 measurements per second. However, equipped with two-mirror deflection scanners, it has the flexibility and capability to sample any position in the field of view in any given order. This improvement opens an opportunity for developing a sampling algorithm that could exploit the sampling budget in a non-uniform scanning pattern to enhance the reconstruction results.

While being able to produce high-quality depth information, the scanning rate of LiDAR system is relatively slow, typically 10fps [6]. In other words, the sampling budget for a given duration is limited. An efficient sampling technique could help to overcome this limitation. Reducing the requirement for spatial resolution could increase the scanning rate or lower the cost of the system. This thesis focuses on developing an effective LiDAR sampling algorithm for real-world situation using semantic cues.

1.3. Thesis outline

The rest of this thesis is organized as follows. The background and related works are described in Chapter 2. Workflow of the proposed system and the sampling algorithm are presented in Chapter 3. The evaluation method and the experimental results are presented in Chapter 4. Chapter 5 summarizes and concludes the thesis.

CHAPTER 2 - BACKGROUND AND RELATED WORKS

In this thesis, various terms related to my work are presented. This section gives a brief introduction to them along with previous works of this research. In the first part, sensor systems in ADAS are introduced including camera and LiDAR used in the proposed system, then methods to calibrate data from these two sensors together are described. After that, the current status of object and road detection algorithms is summarized. The second part shows previous researches which consist of depth-gradient-guided sampling and color-image-guided sampling approaches.

2.1. Background

2.1.1 Sensor systems in ADAS

A car is a highly complex system with hundreds of processors and a variety of types of sensors. These sensor systems give the car the ability to monitor its internal conditions to detect malfunctions. They are also used to sense the external environment, which is the key component to the success of an ADAS system. Active sensors like light detection and ranging (LiDAR) and radio detection and ranging (radar) or passive sensors like cameras are used in experimental and commercial systems.



Figure 1. A test system with a LiDAR and a camera in my lab

For object detection, camera and LiDAR are well supplementary to each other. The LiDAR has advantages of high-precision range measurement; which is independent of the distance, immunity to the night condition and light intensity change. However, its resolution is very limited. While cameras are cheap, capable of much higher resolution and frame rate but have the drawbacks of the sensitivity to intensity shift and inability to work during night time. Therefore, fusion techniques to combine them together are developed. Calibration is necessary to fuse these data together, which

specifies how to convert measurements from the LiDAR coordinate to the image coordinate. The next part describes in detail about this calibration step.

2.1.2 Calibration between LiDAR and camera

Many calibration algorithms use checkerboards as the target. However, checkerboards are often observed with range discrepancies in the LiDAR [8] and that may affect the final estimated outputs [9]. To tackle this problem, in [9], the author uses a circular target. While it is straightforward to detect the circle's center in the image, finding its 3D correspondence point becomes less reliable in sparse data. In [8], the author uses a monochromatic polygonal board, and estimates its edges and intersects them to find the board's vertices. This method improves the accuracy for a low-resolution LiDAR. However, with the monochromatic board, specifying the correct board segment is more challenging in both 2D and 3D range domain. In [8], this task requires manual interventions.

Therefore, I propose to apply the background subtraction technique to automate that work. First, two samples with and without the board are taken. Then, the background subtraction is performed on both 2D and 3D data while the board area is the largest segment of the subtraction output.

The background image and board image are processed as follows. First, these two are converted to grayscale and then subtracted to generate a binary

image by comparing absolute subtraction value and a threshold. Second, the board area is detected as the largest region in the subtraction image. Third, that area is grown to the full board segment using the region growing technique. Finally, its edges are estimated and their intersections are determined as the board's vertices.

For 3D data, both scans are first converted to a range image. Then, the subtraction and board segment detection are carried out as in normal images. The final set of 3D measurements belong to the board is that consists of points corresponding to the board segment on the range image.

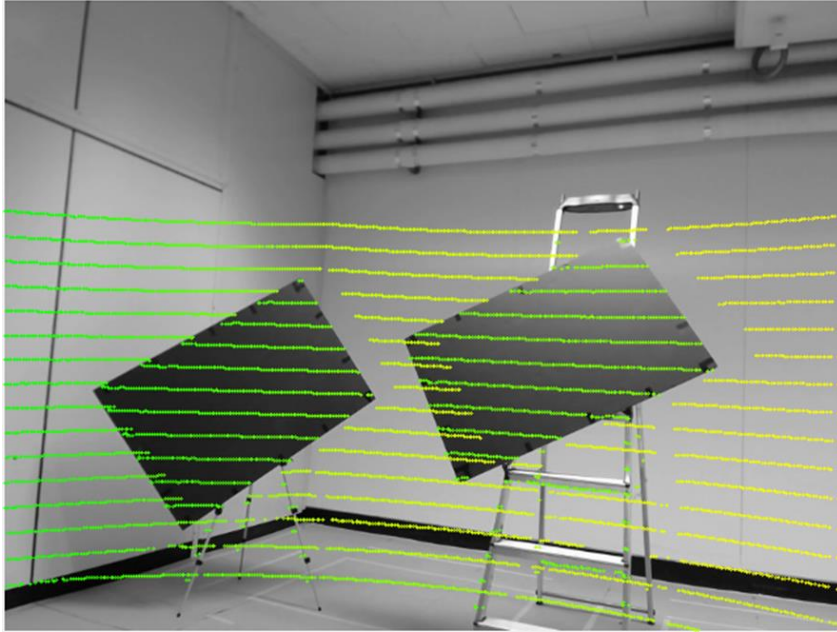


Figure 2. Calibration result. Color indicates the distance. Occlusion is observed due to a different viewpoint.

Finally, after extracting corner points on the 2D image and board data from LiDAR, the calibration process is performed as described in [8]. In my work, the perpendicularity property between two adjacent edges of the rectangle is also enforced. That improves the estimation results when there are little points on edges.

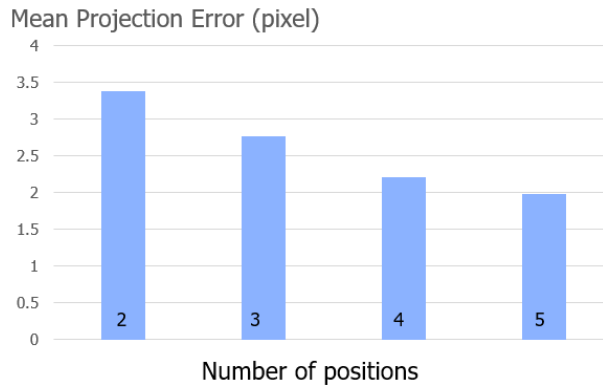


Figure 3. Mean projection error according to the number of samples used for optimization

The experiment is conducted by using a Velodyne VLP16 and a camera. The target is a rectangle slanted to the ground and placed 3-5 meter away from the LiDAR to ensure there are about 10 data lines. The mean projection error in a pixel is provided in Figure 3, which gradually declines when increasing the number of samples used for optimization. Figure 2 depicts the calibration result. Color indicates distance. Occlusion is observed due to a different viewpoint between the camera and LiDAR.

2.1.3 Object detection algorithms

Object detection algorithms have been experiencing rapid developments. In 2012, the first CNN network outperforms human in ImageNet challenge, since then it has become the gold standard in object recognition. Their accuracy and speed have been improved significantly over time. Among them, YOLO is a real-time model which is fast, efficient and optimized end-to-end and thus provides better efficiency with respect to the computational cost [4]. YOLOv3, which could have frame rate up to 40-50fps, is much faster than the typical frame rate of a LiDAR system. Besides, the road detection, which could be considered as an easier problem has also been solved effectively. Many road detection approaches submitted to KITTI road detection benchmarks have the precision of more than 96% while runtime is just about 0.06s [5]. CNN networks are also the gold standard in this area.

2.1.4 Iterative Closest Point algorithm

In this section, the definition of Iterative Closest Point (ICP) algorithm will be given. In the proposed system, ICP is used to register data between two successive LiDAR scans, the result indicates the 6D movement of the LiDAR sensor, which could be converted to the movement and velocity of the car.

Given two sets of corresponding point $X1$ and $X2$, ICP algorithm finds the translation vector t and rotation matrix R that minimize the sum of squared error between these two sets:

$$E(R, t) = \frac{1}{N} \sum_{i=1}^N (X1_i - R * X2_i - t)^2$$

Where $X1_i$ and $X2_i$ are the corresponding points from two sets. This minimization problem is solved in an iterative manner, the gradient of error from the previous iteration is used to set up variables for the next iteration. Above term is called point-to-point error. This term is effective for the cases when the number of samples is dense enough. However, with LiDAR the sample density is very sparse, hence, the point-to-plane error is much more effective and stable.

In more complex systems, visual cues are used to increase the odometry performance and avoiding the ICP to fall to a local minimum. (V-LOAM) [10], which is the top performer in KITTI odometry dataset at the moment of my thesis, is one of those follow this direction. Visual odometry is used for ego-motion estimation then refinement step is carried out using lidar odometry (LOAM) [11], which itself stands at the second position on the KITTI odometry benchmark. This proves that LiDAR measurements are very important for the odometry problem.

2.2. Related works

This section begins with an introduction to the sampling problem. Then, previous approaches and their limitations are subsequently discussed.

Considering the problem of measuring a scene consists of $W \times H$ sampling location with a total budget of N points per scan. A sampling algorithm defines a sampling mask that consists of binary value 0 and 1, where a location with the value of 1 indicates the location would be sampled and a location with the value of 0 would not. Depend on each objective, the above sampling mask is optimized with different criteria. In this work, the goal is to minimize the reconstruction error inside the ROI area. There are some sampling techniques can carry out this task. Uniform or grid sampling is the most straightforward and widely used in real systems. However, they could not exploit additional information to build a better sampling strategy. Several non-uniform sampling strategies [12] [13] [14] [15] have been proposed. Non-uniform sampling techniques for LiDAR can be divided into two following categories.

2.2.1 Depth-gradient-guided sampling

The first category uses properties of the depth images by assuming that the depth images are already available or performing sampling to acquire a

part of it. In [12], the author proposes that sampling along the depth gradient could give the optimal result. The paper is able to recover the depth map in Middlebury dataset with only 5% of the samples. However, this approach is not practical since the depth gradient is not available prior to sampling. Therefore, two-step sampling [13] is designed to apply gradient-based sampling in a more practical manner. At the first step, uniform sampling is carried out with a half of budget. Then the author reconstructs depth from these samples to find the gradient of the reconstructed map. At the second step, the sampling is performed along the gradient of the reconstructed map from the first step. Drawbacks of this method is a delay between two steps for the reconstruction of the depth image, which requires very heavy computation. Second, when the sampling rate becomes too low, there are not enough samples to conduct a good reconstruction from the first step so this affects the overall results.

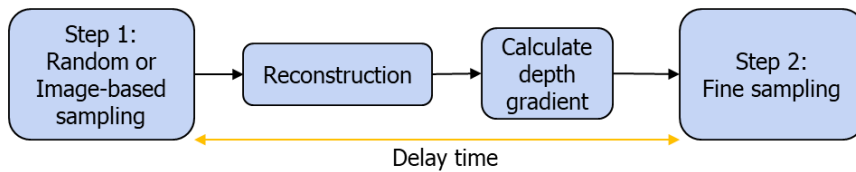


Figure 4. Workflow of two-step sampling

2.2.2 Color-image-guided sampling

The second category uses images to guide the sampling process. In [9], the author uses stereo images to compute a depth disparity map, then samples along the edge of this raw estimation. However, the depth computation from stereo images is complex and time-consuming, and large occlusion is usually found in the estimated depth. A sampling method which uses saliency information from the color image to select sampling position is proposed in [15]. This method has an advantage that the computation of saliency map is cheaper compared to methods mentioned above, however, it is noisier with many outliers remaining in the saliency map.

Despite different types and categories, all previous methods were introduced and tested in the indoor scenarios where the gradient and details of both color image and depth image are much simpler than the outdoor situation. Targeting the real-world environment of ADAS system, the proposed method is not only designed for outdoor scenes but also for a very low sampling rate with a budget about 2000 samples per scan (equivalent to about 1% sampling rate). To achieve this, semantic information detected from cameras is used.

Although this also requires complex computation, I argue that in most existing ADAS systems, object detection is a must-be-done task and some

algorithms such as YOLO is real-time and very efficient. Therefore, the target is to inherit results from this step to improve the LiDAR sampling efficiency based on the observation that the road area is much more simpler and can be constructed with very few samples and the object area is more semantically important. Hence, sampling at the objects area is conducted denser while the sampling rate of road area is kept just enough to a decent reconstruction result. A portion of samples is also moved from the background to ROI area while preserving the localization result acceptable. Practically, to enforce this system, a synchronization method between camera and lidar needs to be developed. However, this task is left for a future work and not included in this thesis.

CHAPTER 3 - SAMPLING ALGORITHM

In this section, the system architecture and configuration are first described then the thesis will analyze each area inside ROI and finally explain the algorithm to distribute sampling budget between areas.

3.1 System Workflow

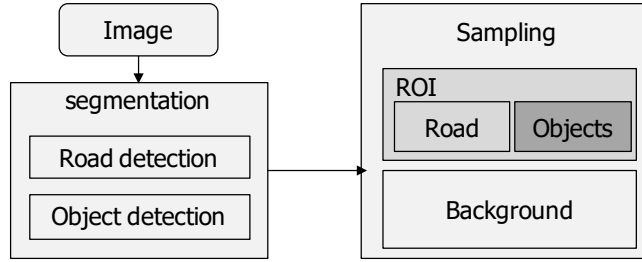


Figure 5. Workflow of the proposed system

The workflow of the proposed system is described as in figure 5. First, an image is captured using a camera. Then, the object and road detection are carried out simultaneously. The result of this step is used to guide the proposed sampling algorithm. At this step, the camera can take the next image immediately instead of waiting for the sampling process to be finished. That way, there will be no delay time in the system. The reconstruction and evaluation step are used for evaluating the sampling algorithm only. They could be performed offline at any time if the samples are stored. The steps require most computing power in this system is road and object detection. They are usually implemented by a neuron network. The state-of-the-art

algorithms can run at about 40-50fps as mentioned at the background part. This frame rate is already much higher than LiDAR frame rate, hence, the possible delay from this part is also not too long.

3.2 System configuration



Figure 6. Road and objects mask generated from KITTI dataset

As described in figure 5, the proposed system includes road and object detection step. However, this thesis only proposes and evaluates the sampling phase so ground truth data is used for those steps from KITTI. The KITTI object detection dataset also contains unrelated objects and cars are not on the road so they must be filtered out manually. On the road, KITTI road dataset only includes the road in the current direction. The road in the opposite

direction and the pavement are not included. In the two-way streets, all lanes are included [16]. This is a very reasonable setup since it is the part of the road where most of the traffic situations related to current car happen. There are many available methods for reconstruction step. Recently, while some conventional image processing methods still have good performance [17], there is a domination of the CNN-based methods in this area. However, in this application, linear interpolation is used for evaluation, in details it is the griddata function in MATLAB. That is because I believe that it is a classical method, not specifically modeled for any sampling method, so it is able to give a fair comparison. The evaluation method used is the KITTI provided evaluation code. In this study, mean absolute error (MAE) is used as the criteria for all comparisons.

$$\mathbf{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}.$$

3.3 Distribution of budget inside ROI

3.3.1 Reconstruction characteristics of each area in ROI

Given the total budget of N samples in ROI area, the problem is to specify the budget for road area N_r and budget for object areas N_o ($N_r + N_o = N$), which minimizes the reconstruction error in ROI area. Inside each area, random sampling is used with a minor modification that the closer area

is sampled sparser than further area. The reconstruction error can only be calculated after sampling. Therefore, the MAE result has to be predicted prior to sampling to be able to optimize final output. Which model has better prediction will result in better performance.

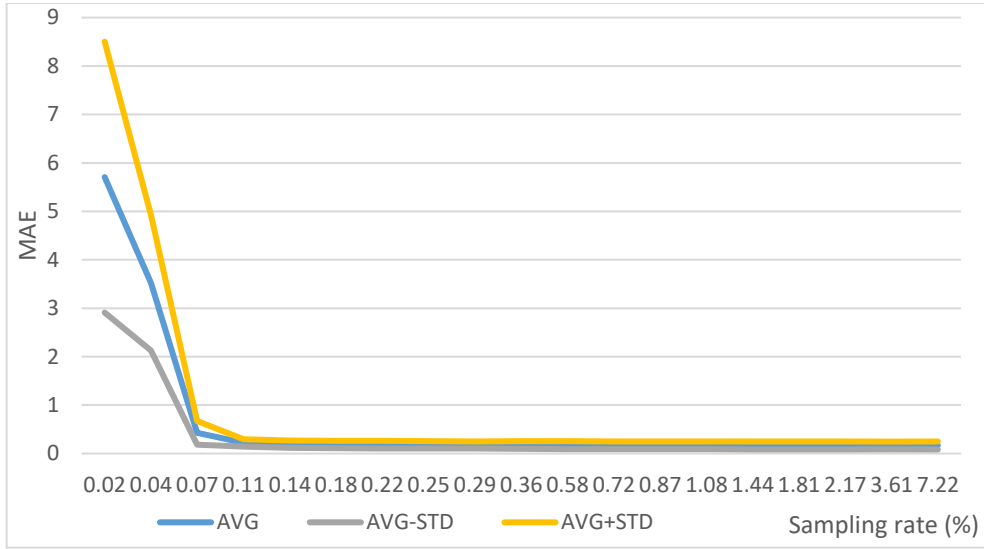


Figure 7. Reconstruction characteristics of the road area

Figure 7 examines the MAE output of road area based on the sampling rate. Three lines are demonstrated: AVG as the mean MAE results of all images, AVG-STD is the mean results minus the standard deviation and AVG+STD is the mean result plus standard deviation. It is clear that the road has much smaller errors and it converges very fast, starting at 0.11%. At this point, the small standard deviation indicates that results among images

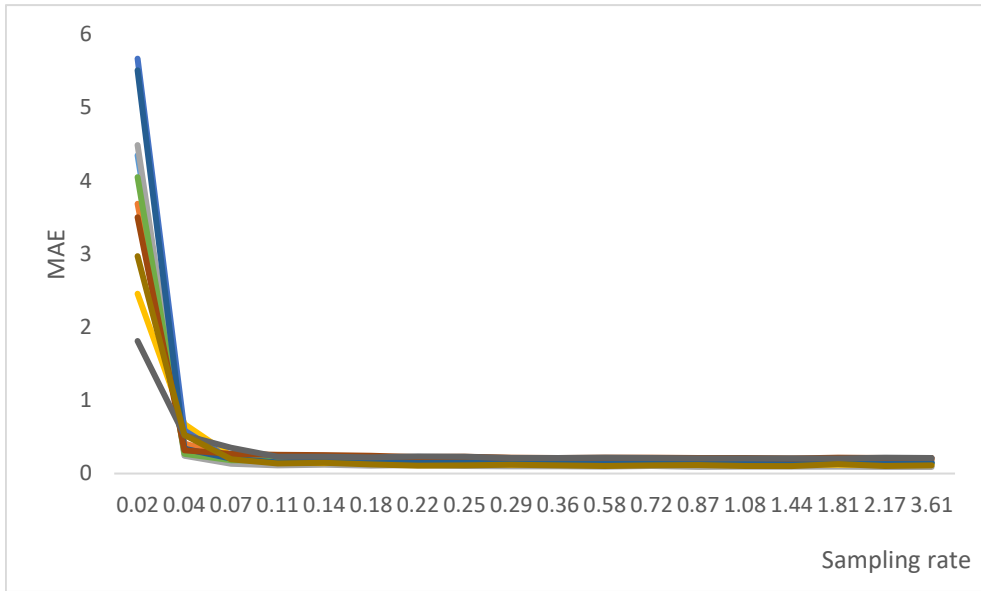


Figure 8. MAE from 10 random scans

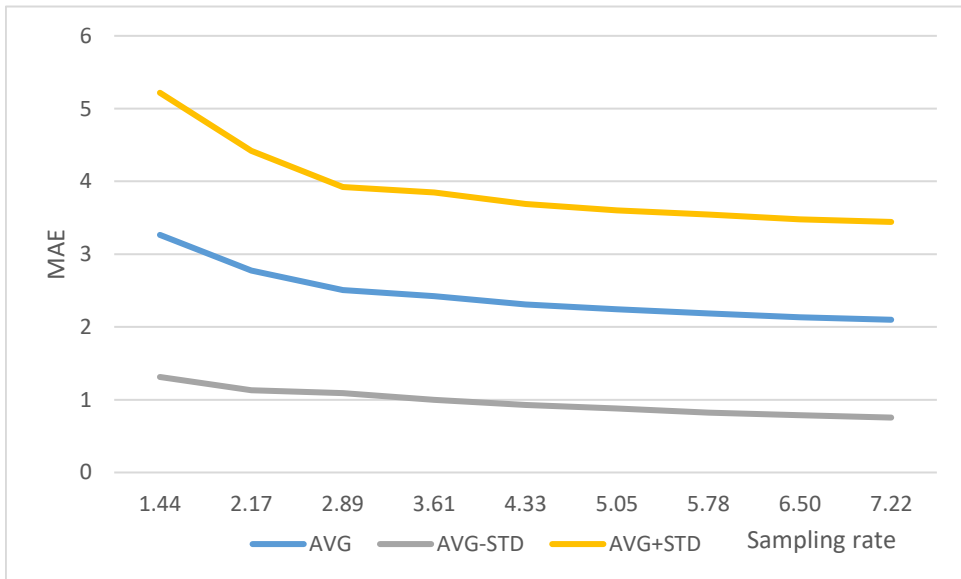


Figure 9. Reconstruction characteristics of objects area

are consistent. This is demonstrated again in figure 8, where MAE of the reconstruction map of 10 random scans is shown. The same result holds and

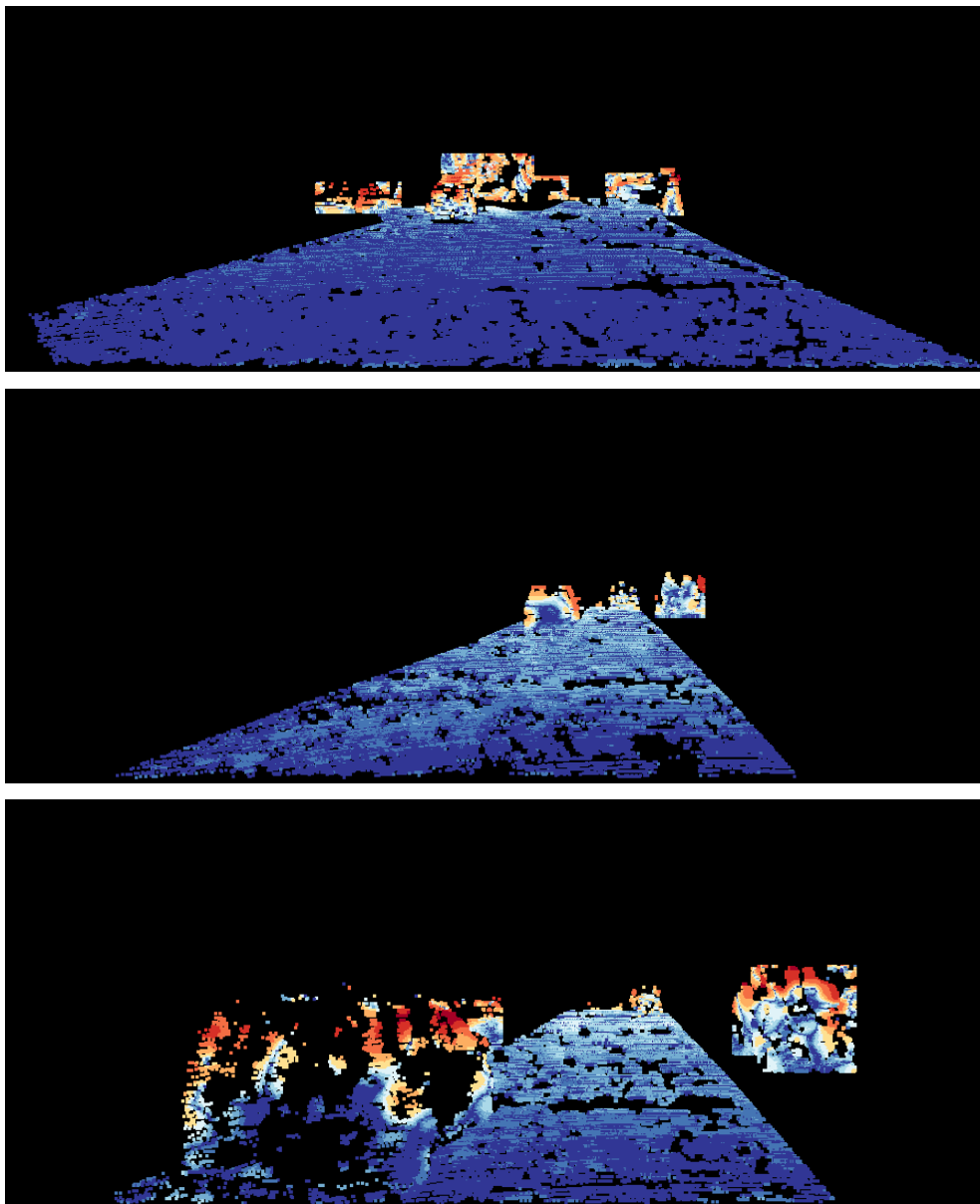


Figure 10. Error images after reconstruction. Blue indicates smaller error. More red means higher error.

they start to converge at a sampling rate of 0.11%.

The reconstruction characteristics of objects area are shown in figure 9. As can be seen, the MAE of objects is worse and no convergence was found. The standard deviation stays high along the sampling rate axis. This means that the sampling rate of the objects area must be kept as high as possible. Fortunately, the decrease in MAE of this area starts to slow down at 4.33% mean that they begin to become more stable at this point. In figure 10, a comparison of these differences is shown. In all three images, the error in objects area is much higher than in road area. The second image also shows that in road area, the error in the further area is higher than closer area. In the third image, although the object is close, its error is still high. In this particular image, object area is larger than it is in the other two images.

As mention before, to distribute sample budget, the MAE result of road and objects area are predicted. Base on above observation, the MAE of road and objects area are modeled into a ration function with three parameters to control its behavior as follows:

$$F(x) = a + \frac{b}{x-c} \quad (1)$$

Where $a = 0.1627$, $b = 0.0074$, $c = 0.0255$ for road and $a = 1.831$, $b = 2.011$, $c = 0.03956$ for objects. The SSE of these fittings are 0.0000057 and 0.002,

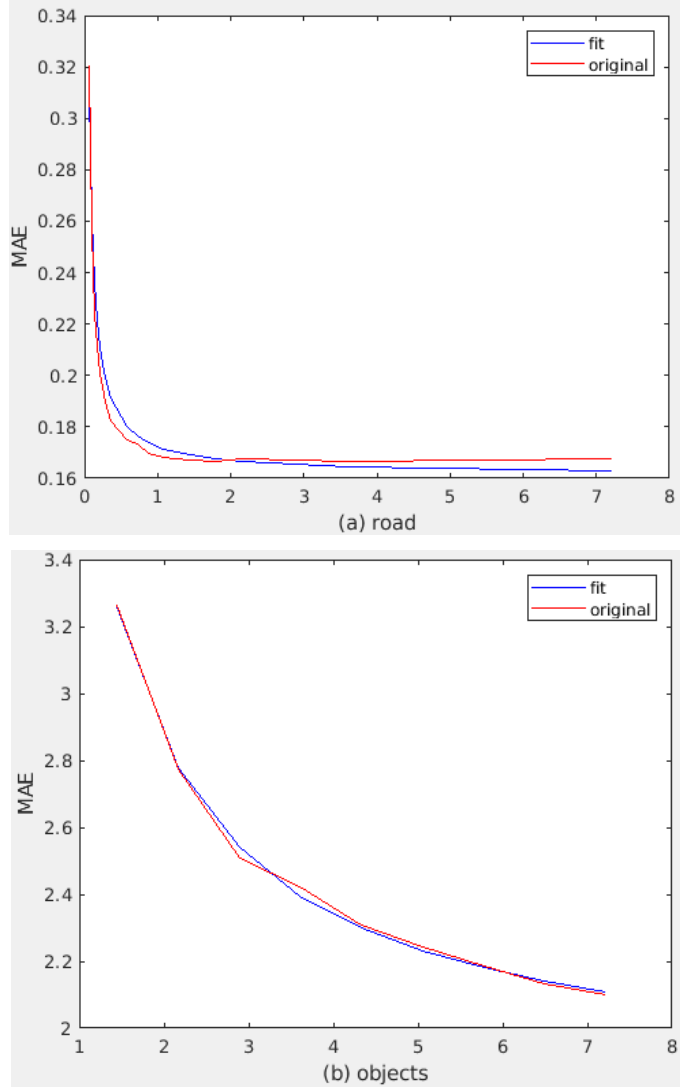


Figure 11. Fitting results of MAE of road and objects area

respectively. The results of these fittings are shown in figure 11. As analyzed as above, the fitting of road area is reliable since MAE results of all images converge. While the MAE results vary across scenes, hence, it is fitted with

mean results of all images, so it may not be good when considering a single case. However, for all test images, this method still gives a good result.

3.3.2 Distribution budget between road and objects area

Based on the model achieved from above section, the proposed algorithm is stated as follow: Given budget N , the area of road S_r and objects area S_o in pixels, the problem is to find the budget for the road N_r and budget for objects area N_o ($N_r + N_o = N$). It is solved by optimizing following criteria:

$$\text{Minimize } \frac{S_r \times F_1(N_r/S_r) + \gamma \times S_o \times F_2(N_o/S_o)}{S_r + \gamma \times S_o} \quad (2)$$

With F_1 and F_2 are the fitted function for road and objects area in (1), respectively. γ is the weighting parameter between the road and the objects area. When γ equals to 1, the criteria become the predicted MAE of the entire ROI area. if γ increases, the budget for objects area also increases. This equation is a compromise between enhancing reconstruction result of entire ROI and improving reconstruction result of objects area only. Sampling budget is finite and an integer, so this minimization could be solved numerically effectively.

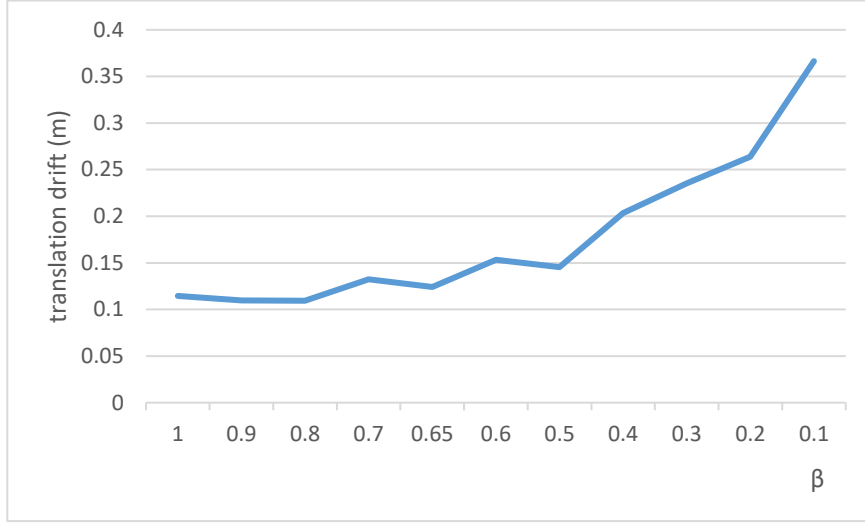
In KITTI dataset, the road area is about 4-5 times larger than object area. So, when sampling budget is set at 1000 points per scan, the criteria usually converges with about 300-400 samples for road area, the rest of the budget

for objects area. When the objects area/road area ratio increases this number decrease to reserve more samples for objects area. There are some cases where there are no valid objects, hence all samples are used for the road.

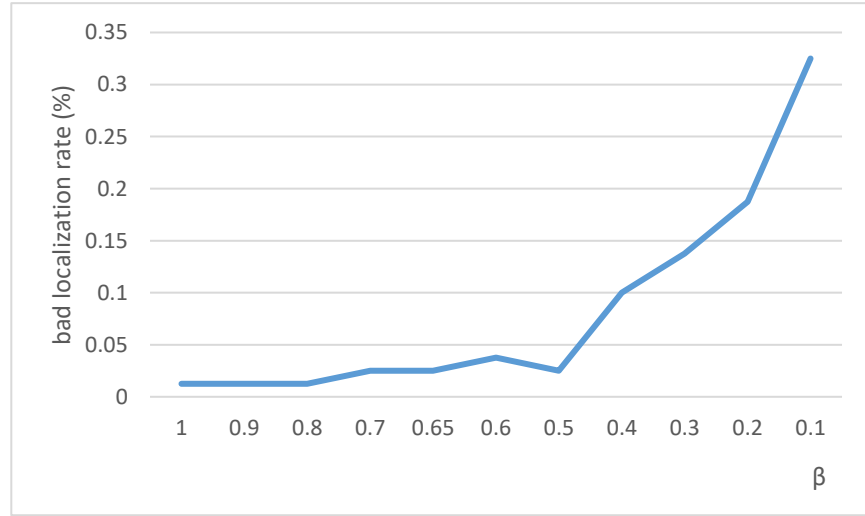
3.4 Distribution budget between ROI and background area

Ideally, when the objective is increasing the reconstruction results as much as possible, then if all budget is distrusted to the ROI area, the reconstruction results inside ROI will be best. In real ADAS system, although it can be argued that only samples insides ROI will give the semantic meaning and be helpful for tasks like obstacle detection, distance measurement from the car to objects. However, samples in non-ROI area (also called background area) are also helpful for tasks like odometry, detecting motion and movement of the car. When testing odometry using LiDAR samples by Iterative Closest Point (ICP) algorithm, a huge decrease in performance is detected when sampling is performed inside ROI region only. The localization results after moving β percent of the budget from the background area to the ROI are presented in figure 12.

Figure 12 demonstrates translation errors and bad localization rate (the ratio of cases when the translation error is higher than 50cm) of random sampling in the entire image with a budget of 2000 results after moving β percent of the budget from the background area to the ROI. As β decreases,



(a) Translation error



(b) Bad localization rate

Figure 12. Localization results on dependence of β

the translation error and bad localization rate increase. When β is smaller than 50%, they start to rise sharply. When using the proposed algorithm inside the ROI area only, the odometry result is very bad. With the same budget of random sampling, the output is 5 time worse compared to the case β equals to

1. This shows that to get better odometry result, it is necessary to have samples outside ROI area. The main objective is getting better reconstruction result; therefore, the sampling density inside ROI area is kept denser than that in non-ROI area. The budget for non-ROI will be set only to make its odometry results keep up with that of random sampling.

Given N total budget for the entire field of view, call N_{ROI} and N_{nROI} are budget for ROI and non-ROI area ($N_{ROI}+N_{nROI}=N$), then N_{ROI} and N_{nROI} are determined as follow:

$$N_{ROI} = \min \left(\frac{\alpha * N * S_{ROI}}{S_{nROI} + \alpha * S_{ROI}}, N - \frac{\beta * N * S_{nROI}}{S_{nROI} + S_{ROI}} \right) \quad (3)$$

With S_{ROI} and S_{nROI} are the area of ROI and background area. α is the weighting parameter, the increase of α indicates that more samples are expected to be moved from the background area to the ROI. β indicates the minimum of sampling budget must be kept in the background area to preserve the localization result. In this experiment, α and β are set to 1.5 and 0.65, respectively. In equation (1), which term is chosen depends on the ratio S_{ROI}/S_{nROI} . If this ratio is small, the first term is chosen. In reverse, if this ratio is large, the second term is chosen to preserve the localization result.

CHAPTER 4 - EXPERIMENTAL RESULTS

This section explains the experimental methodology and results. First, the KITTI dataset is introduced with some adjustments made. Then the objective and quantitative results are shown and explained in detail.

5.1. Dataset



Raw scan



Ground truth

Figure 13. Raw scan and ground truth depth from KITTI

KITTI dataset is used for experiments, the raw depth map as input and ground truth data for evaluation. Figure 13 shows an example of the raw and ground truth depth data in KITTI after mapping to its corresponding color image. The raw depth data is taken using Velodyne HDE64 sensor [18]. The

ground truth is generated by combining 11 raw scans together and discarding measurements that are not consistent among those scans [19]. One drawback of this method is that the measurements on moving objects are usually not consistent and being removed, leading to a sparser ground truth on these important areas, whereas the static object areas have denser sample density. This can be seen in figure 14, the cyclist even has more samples in raw depth than in ground truth while technically the ground truth is combined with 11 raw depth images. In some other areas, the ground truth is very dense, some other areas do not have any ground truth data. Therefore, in this experiment, even the ground truth has much more data than the raw depth image, the raw depth images give better reconstruction result than ground truth after sampling.



Figure 14. Raw depth and ground truth in moving object area

This thesis also applied the ground truth of both object detection and road detection benchmarks in this experiment. Therefore, images that could be used in these tests must exist in three benchmarks: depth completion, road

detection, and object detection. Only a few images, exactly 106 images, match these requirements. For this sampling problem, it is believed this is enough to develop the proposed algorithm and give a comprehensive evaluation. Inside ROI, raw depth has the sampling rate of about 7-8%. In table 1, in the dataset, the minimum sampling rate of raw depth data is 3.65%, the maximum is 8%, while the average number is 7.19%. Between road and objects area, on average the road is 4 times larger than the objects area. However, in some exceptional cases, the objects area is 2-3 times larger than road area. In these cases, the objects are very close to the sensors. About the number of samples inside each area, on average the road areas have 6.75 times more samples than the objects area. These number shows that the current sampling strategy is not efficient. The road need smaller sampling rate than objects area, while in current systems, the opposite number is found.

Table 1. Sampling rate inside ROI and ratio between road and objects area

	Area Object/Road	N. sample Object/Road	Sampling rate inside ROI
Min	0.00344115	0.000524109	0.036568
Max	2.703898226	1.728091529	0.0807
Mean	0.232687667	0.148326604	0.071916

A sampling with the budget of 1000 samples per scan for ROI and 2000 samples for the entire frame is used in the experiment. In table 2, the average number of samples inside ROI is 7202, this means that about 13.89% of available samples are used. The reconstruction with all samples in raw depth is used as upper-bound performance because that is all measurements available. The random sampling is used as base performance and two-step sampling for comparison. Grid sampling is not possible in these experiments, because the raw depth does not have regular sample grid.

Table 2. Number of samples inside road and objects area

	Number of samples		
	Total	Road	Objects
Min	3527	3348	3
Max	12722	10806	7099
Mean	7202.388	6397.369	805.0194

5.2. Evaluation inside ROI

This section shows evaluation results correspond to the method in section 3.3 with the budget of 1000 samples inside the ROI.

5.2.1 Objective evaluation

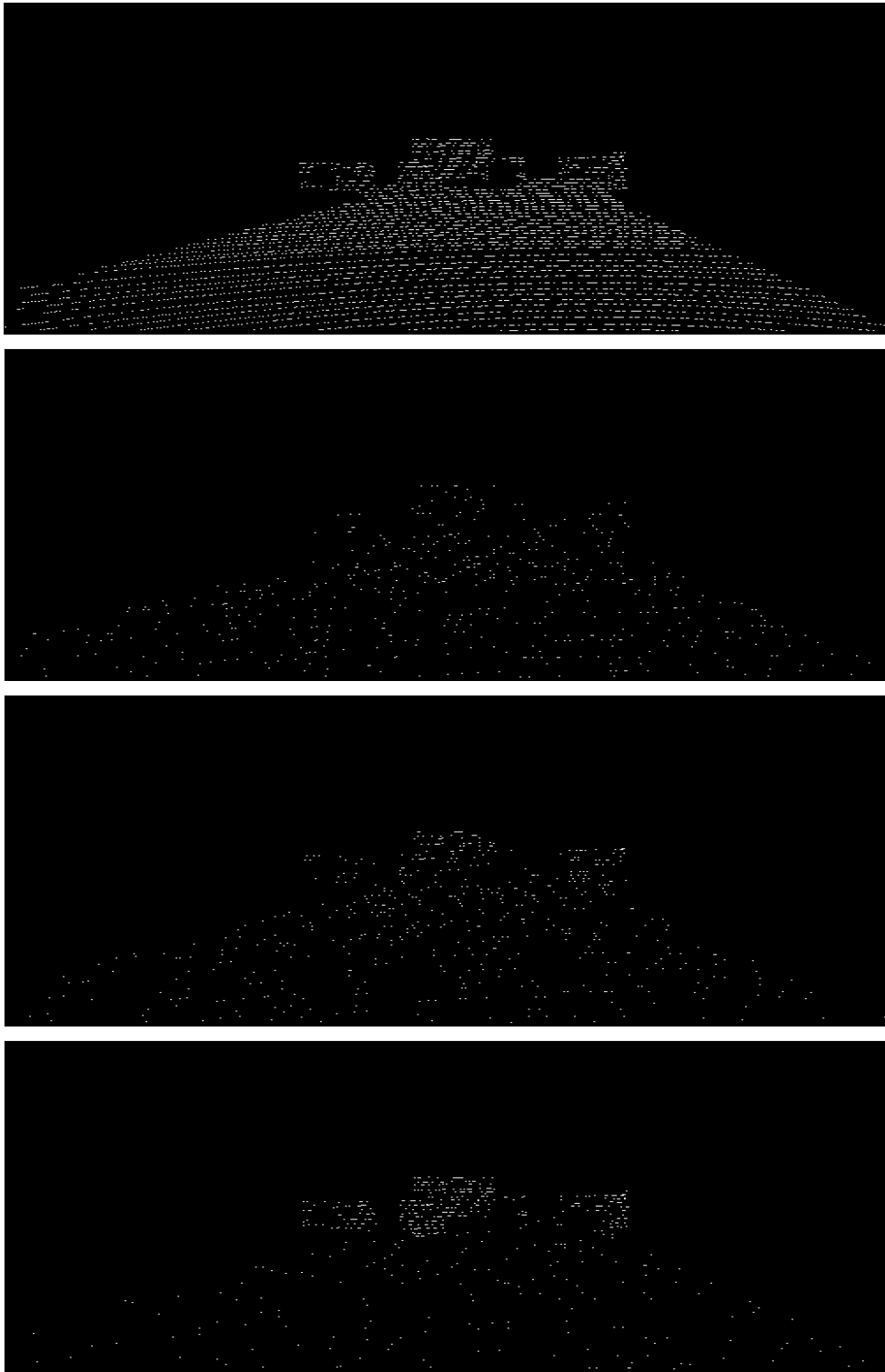
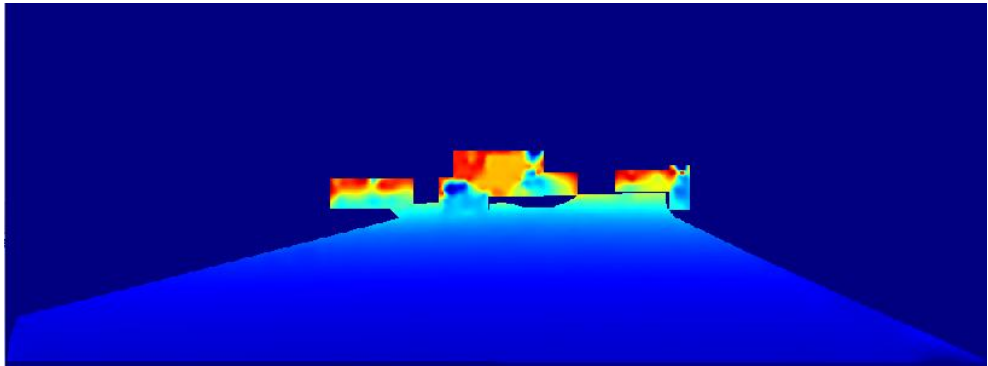
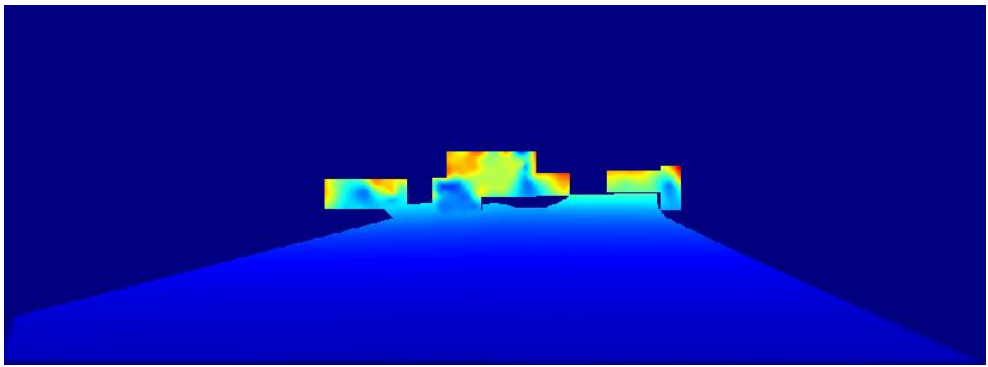
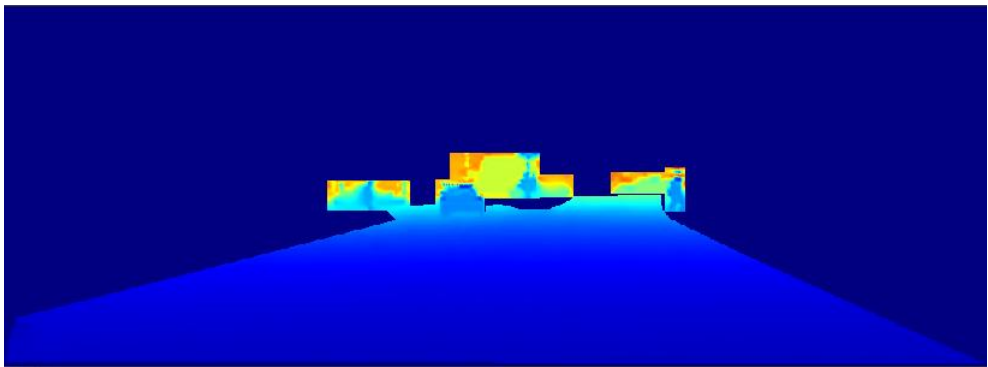
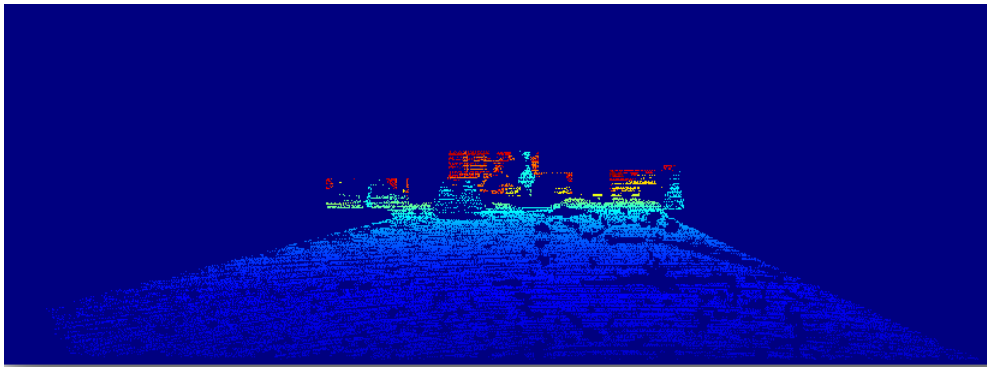


Figure 15. From top to bottom: masks of raw depth and sampling mask of random sampling, two-step sampling and proposed algorithm



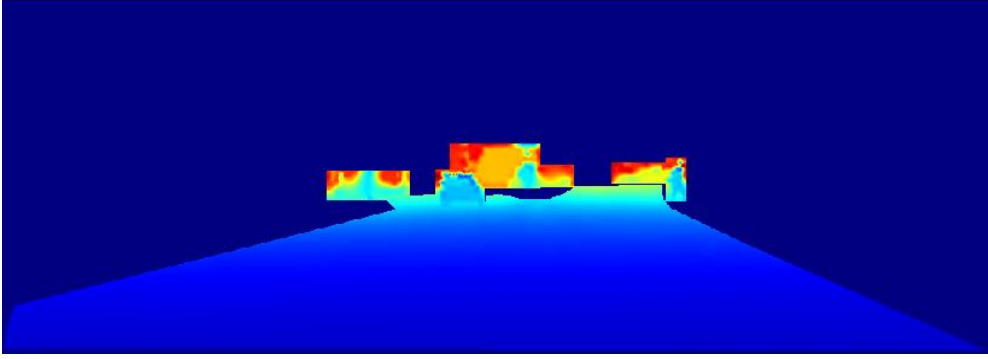


Figure 16. Depth ground truth and the reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm

Figure 15 shows sampling mask of raw depth, random sampling, two-step sampling and the proposed algorithm. There is no specific pattern for random sampling. While two-step sampling is a little denser at object area, the proposed algorithm yields a very significant difference between two areas. The final reconstructed image is shown in figure 16 and crop out of objects area is depicted in figure 17. As being demonstrated, the road area is well reconstructed in all for all three methods. Objectively, no differences between these images can be seen. However, when considering the objects area, there are significant differences. In the reconstructed image of raw depth, the shape of the car and pedestrian recognizable. In the reconstructed image of random sampling, the cars could not be recognized. They start looming in the reconstructed image of two-step sampling and in the reconstructed image of the proposed method, it becomes recognizable. Hence, the proposed algorithm has the closest output compared to upper-bound

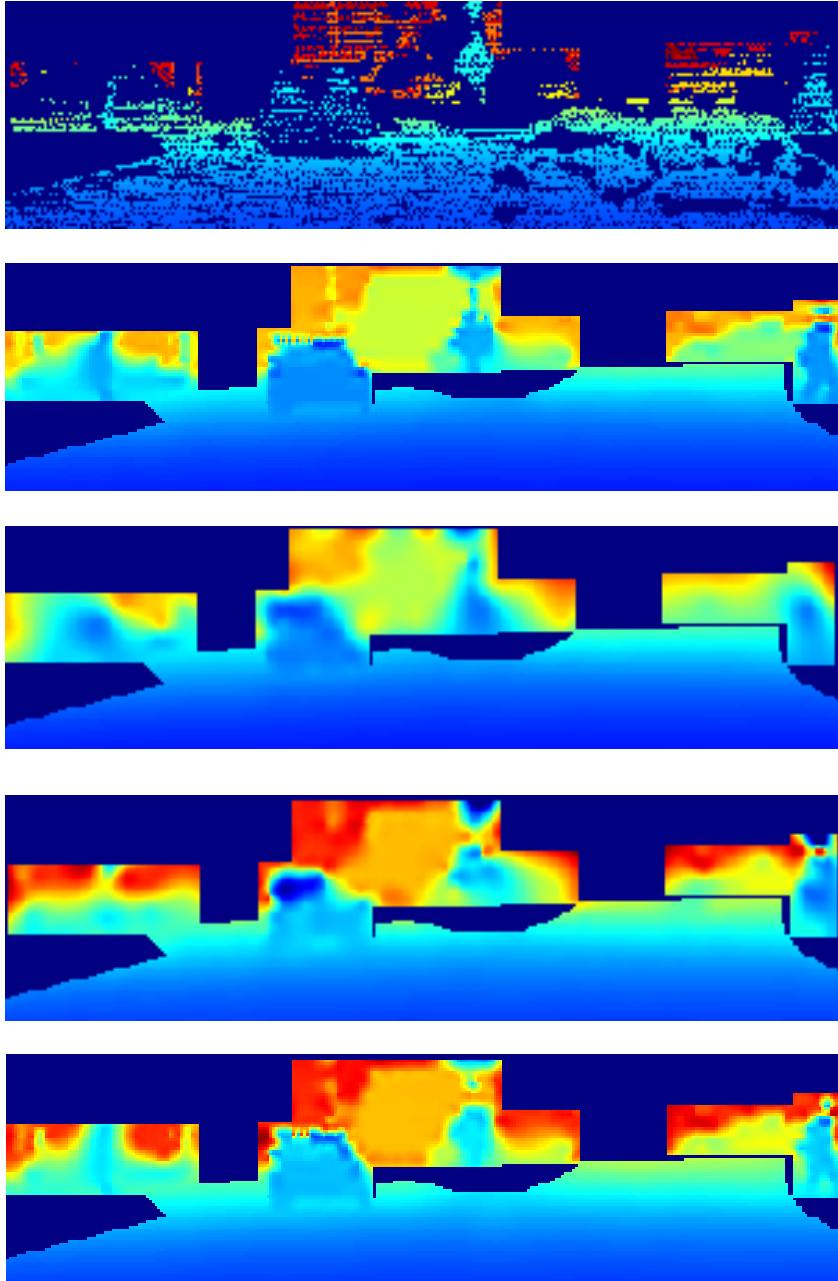


Figure 17. Zoomed out of objects area in depth ground truth and the reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm

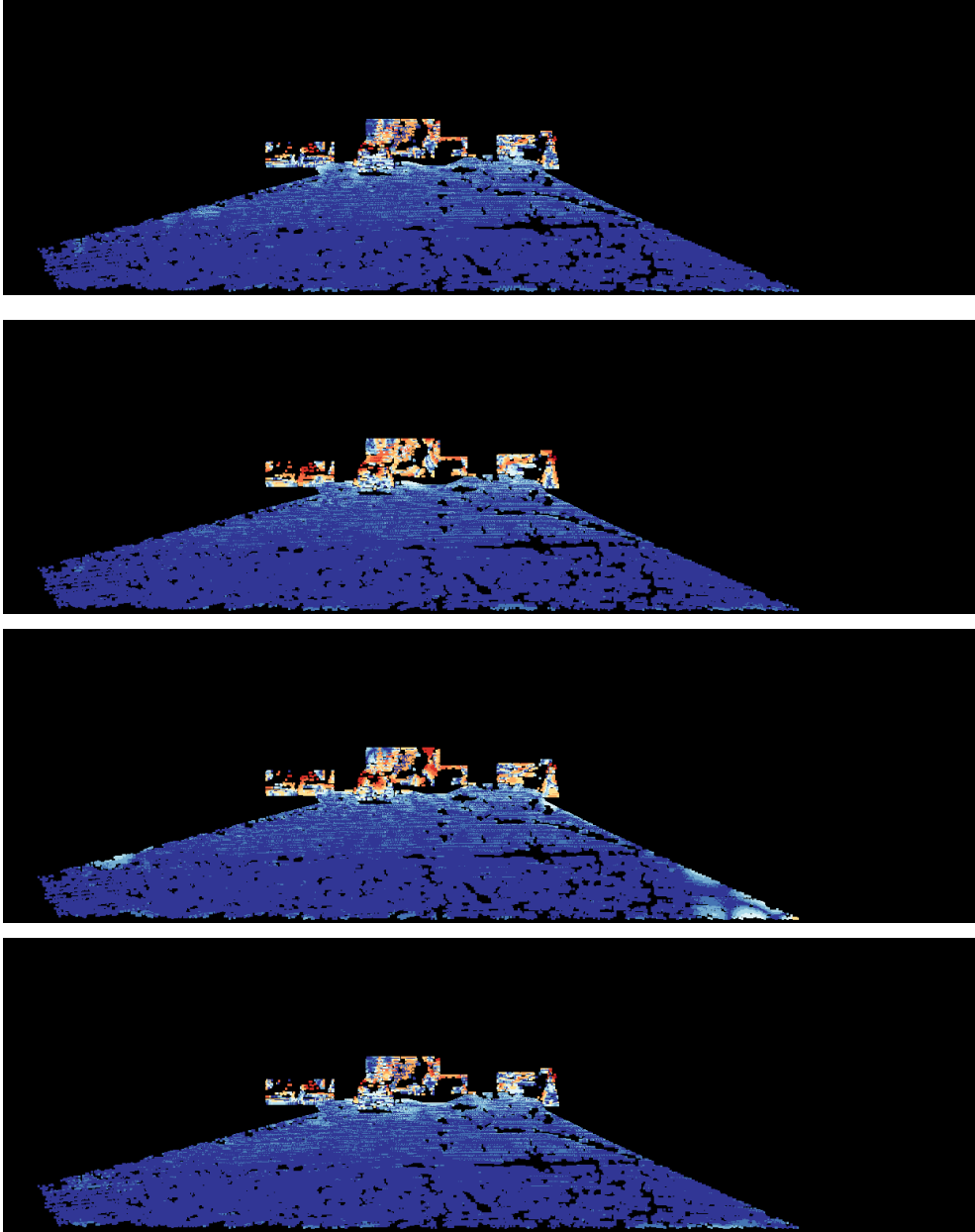


Figure 18. Error images from reconstruction result from raw depth, random sampling, two-step sampling and proposed algorithm

performance and the random sampling does worst. Above analysis shows the superiority of proposed method, the differences between these three

algorithms are significant in both sampling masks and reconstructed images. From error images, the same conclusion is drawn. Random sampling has greater errors than any other, while proposed methods are most close to raw depth in term of reconstruction result.

5.2.2 Quantitative evaluation

Table 3. Mean Absolute Error comparison between methods

N	Method	MAE (m)		
		Road	Object	Overall
Maximum		0.16696	2.1082	0.23508
1000	Random	0.16985	3.4527	0.30366
	Two-step	0.16528	2.9457	0.27984
	Proposed	0.18114	2.2286	0.26508

The quantitative result is illustrated in table 3. The same result still holds as above objective evaluations. In road area, Two-step sampling does best, following by random sampling and proposed algorithm has the largest difference compared to the upper-bound performance. However, this gap is only less than 2cm. This is in range of Velodyne LiDAR accuracy and therefore the sacrifice is acceptable. In objects area, the random sampling does worst and the proposed algorithm has a very good result with the difference in reconstructed depth from the upper-bound performance is only

about 10cm while two-step sampling yield more than 80cm difference. Objects area is semantically more important than road area, so this is an important improvement.

5.3. Evaluation in the entire frame

This section shows the evaluation of final results when both distribution of budget inside ROI and between ROI and background are applied. The sampling budget for the entire frame is set as 2000 samples per scan. To perform this evaluation, KITTI depth reconstruction benchmark as above tests and the KITTI odometry benchmark is used. The benchmark consists of 22 sequences, 11 sequences for training and 11 sequences for testing. The top performer of this benchmark is currently being the Visual-lidar odometry and mapping (V-LOAM) that uses visual data to make an estimate and then uses LiDAR data to refine the result. The purpose of this test is to evaluate the effect of different sampling algorithm to odometry result so only LiDAR data is used as input to the ICP algorithm. Because the LiDAR data is very sparse, point-to-plane metrics is used and performs significantly better than the point-to-point metrics.

To evaluate proposed algorithm, object detection and road detection results are annotated manually. There are 20 images prepared for the tests. The results of the various methods are shown in the following table:

Table 4. Odometry results of various setups

	translation drift (m)	rotation drift (degree)
raw depth	0.0385	0.004
random	0.14607	0.01421
two-step	0.09205	0.01275
Proposed ($\beta=0$)	0.71255	0.03817
Proposed ($\beta=1$)	0.10723	0.01455
Proposed ($\beta=0.65$)	0.13511	0.01752

In table 4, the first result is from raw depth with all available samples. The rest of them is taken when sampling budget is set to 2000. The second row shows the result of random sampling, the next row is from two-step sampling. The last three rows demonstrate results from the proposed algorithm with three different setups. First, the result when sampling is performed inside ROI region only is shown, the second result is archived when the same sampling rate is set inside and outside ROI area. The last result is archived when the sampling inside ROI area is denser than non-ROI area, in equation 4, the parameter α is set to 1.5 and β to 0.65.

As demonstrated in the table, the output of raw depth is very good with the translation drift only about 4 centimeters. Proposed algorithm when being performed in inside ROI region only performs worst, the error is much higher



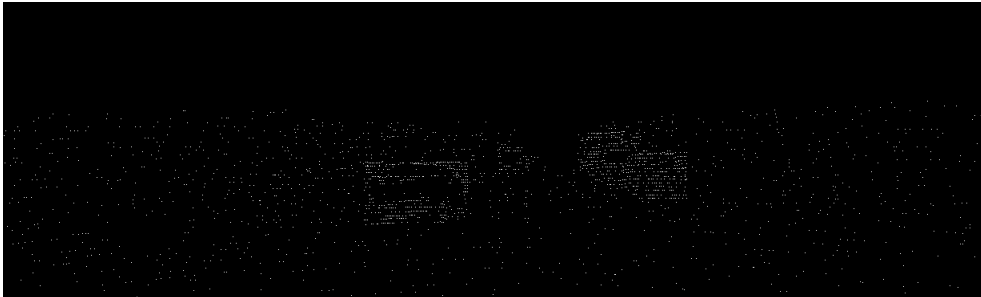
Image of the scene



Sampling mask of random sampling



Sampling mask of two-step sampling



Sampling mask of proposed algorithm

Figure 19. Sampling maps comparison of different methods

than any other methods. However, when being performed on the entire field of view, the result of the proposed method is satisfactory with the translation drift only about 10 centimeters, this is approximately the same as the accuracy of the odometry dataset. When increasing the sampling rate inside ROI area in order to achieve higher reconstruction result, the odometry result gets worse. At this setup, the proposed algorithm still gets a better result than random sampling. Two-step sampling performs best with odometry, about 1.5 centimeters better than the proposed algorithm at the best setup.

The sampling masks of different methods are shown in figure 19, where proposed method shows significantly denser sampling rate in the objects area. Considering the reconstruction result of above setups, their result is imprinted in the following table:

Table 5. Reconstruction results of various setups by MAE

	road	object	ROI	background	overall
raw depth	0.131448	0.94143	0.239427	0.46965	0.339022
random	0.142748	1.97609	0.366456	0.98562	0.636075
Two-step	0.139731	2.08233	0.378107	0.90666	0.619105
Proposed ($\beta=0$)	0.170837	1.02971	0.287913		
Proposed ($\beta=1$)	0.196628	1.26516	0.340591	0.97286	0.627206
Proposed ($\beta=0.65$)	0.197792	1.16840	0.328167	1.04832	0.675041

Comparing the reconstruction results of the ROI area, the proposed algorithm performs best. All three setups obtain better results than both random sampling and two-step sampling. The result when doing sampling inside the ROI area only obviously gives the best result, followed by the result when setting the sampling rate inside ROI higher than non-ROI area. When the sampling rate of ROI area is equal to that of non-ROI, which is the same as random and two-step sampling, the proposed algorithm also performs better. Comparing to the reconstruction result in the background area and the overall result, two-step, and random sampling are better. However, ROI is the priority so the results are reasonable.

CHAPTER 5 - CONCLUSION

This thesis introduces an idea of using results from object and road detection step in ADAS systems to better guide the LiDAR sampling process, specifically sampling on the ROI. The focus is on distributing the sampling budget between the ROI and background area and between objects and road area inside the ROI. A portion of the budget of the background area is moved to ROI are with the constraint of preserving the localization result. In the other hand, by the observation that road area is simpler and easier to construct, the proposed algorithm distributes to this area just enough points to keep a decent result, while the rest of sample budget is reserved to objects area. In comparison, two-step sampling is more general than the proposed approach and when being applied to the road environment, its gradient-based approach also results in sparser sampling in road area and denser sampling in objects area. However, the proposed algorithm is highly optimized for this typical ADAS application, therefore, it yields better results in both objective and quantitative evaluation. Consequently, the experiment shows the proposed method significantly improves reconstruction result in objects area by 44.9% and entire frame by 15.1% while just sacrificing little in road area.

REFERENCES

- [1] "The Economist Special Report," [Online]. Available: <https://www.economist.com/news/special-report/21737425-foreseen-and-unforeseen-consequences-self-driving-cars-will-profoundly-change-way>. [Accessed 11 05 2018].
- [2] U. J. J. S. S. J. Y. P. K. S. D. H. S. K. Lee, "EureCar turbo: A self-driving car that can handle adverse weather conditions," in *The International Conference of Intelligent Robots and Systems (IROS)*, 2016.
- [3] K. D. K. F. B. M. M. S. C. W. H. Bengler, "Three decades of driver assistance systems: review and future perspectives," *IEEE Intell. Trans. Syst. Mag.*, vol. 6, p. 6–22, 2014.
- [4] D. S. G. R. e. a. Redmon J, "You Only Look Once: Unified, Real-Time Object Detection," in *Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [5] "KITTI road benchmark," KITTI road benchmark, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_road.php. [Accessed 12 05 2018].
- [6] "Velodyne LiDAR," Velodyne LiDAR, [Online]. Available: <http://velodynelidar.com/hdl-64e.html>. [Accessed 10 05 2018].
- [7] D. V. L. H. K. a. H.-J. L. Xuan Truong Nguyen, "A High-Definition LIDAR System Based on Two-Mirror Deflection Scanners," *IEEE Sensors Journal*, vol. 18, no. 2, pp. 559-568, Jan. 2018.

- [8] Y. S. W. C. C. K. U. K. a. S. S. Park Y, "Calibration between color camera and 3D LIDAR instruments with a polygonal planar board," *Sensors (Basel)*, 2014.
- [9] L. D. B. a. B. B. H. Alismail, "Automatic Calibration of a Range Sensor and Camera System," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, Zurich, 2012.
- [10] J. Z. a. S. Singh, "Low-drift and real-time lidar odometry and mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401-416, 2017.
- [11] J. Z. a. S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," in *Robotics: Science and Systems Conference*, July, 2014.
- [12] M. K. a. K. D. S. Hawe, "Dense disparity maps from sparse disparity measurements," in *IEEE Int. Conf. Computer Vision (ICCV'11)*, Nov. 2011.
- [13] S. C. a. T. N. L.-K Liu, "Depth reconstruction from sparse samples: Representation, algorithm, and sampling," vol. 26, no. 6, p. 1983–1996, Jun. 2015.
- [14] Z. L. T. N. Lee-Kang Liu, "Sharp disparity reconstruction using sparse disparity measurement and color information," *IVMSP Workshop*, Sep. 2013.

- [15]A. W. a. D. A. C. S. Schwartz, "Saliency-guided compressive sensing approach to efficient laser range measurement," *J. Vis. Commun. Image R.*, vol. 24, no. 2, p. 160–170, 2013.
- [16]J. F. a. T. K. a. A. Geiger, "A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.
- [17]J. a. H. A. a. W. S. L. Ku, "In Defense of Classical Image Processing: Fast Depth Completion on the CPU," *arXiv preprint arXiv:1802.00036*, 2018.
- [18]A. G. a. P. L. a. R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [19]A. G. a. P. L. a. C. S. a. R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *International Journal of Robotics Research (IJRR)*, 2013.