1. Appendix A. Appearance & trajectory

1.1. The appearance of a Mine from Bird Eyes' View

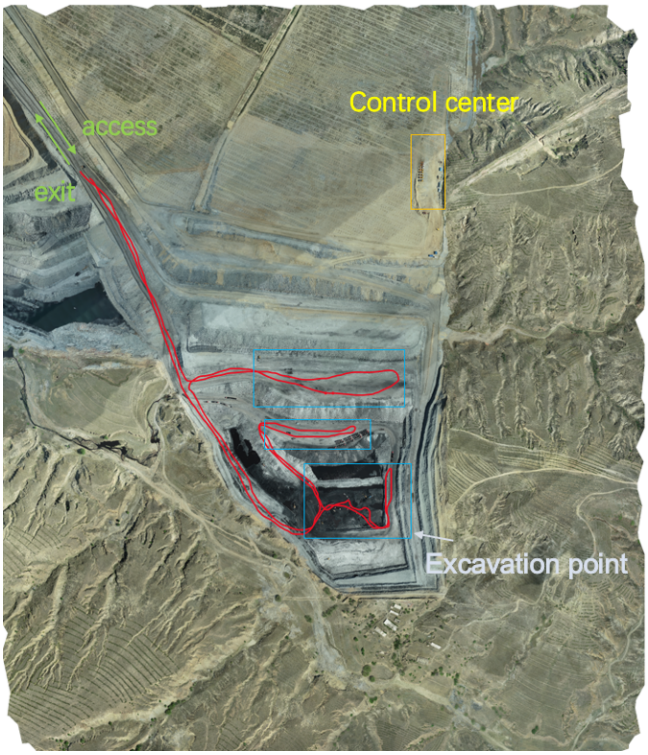


Figure 1. The appearance of a open-pit mine from BEV and the trajectory of the collection platform

As shown in Fig. 1, this is a typical appearance of one open-pit mine from an unmanned aerial vehicle. The center of the mining area is the excavation point. Excavators continue to dig down though the mining process and trucks as well as wide-body trucks transport the soil and ore. The green arrows in the figure indicate the direction of the vehicle entering and exiting, and the red is the trajectory of the collection vehicle. The orange box on the right top of picture is the monitoring and control center, and the blue areas are loops formed by the vehicle's running route. In addition, scholars can observe the unstructured roads macroscopically through the BEV.

1.2. The trajectory of the collection platform

Fig. 2 is one of the trajectories of the wide-body truck collection platform. We can observe the a few loops and the altitude difference. During the movement of this vehicle, the lowest point is 1348m and the highest altitude point is 1363m, with a drop of nearly 20m in this route.

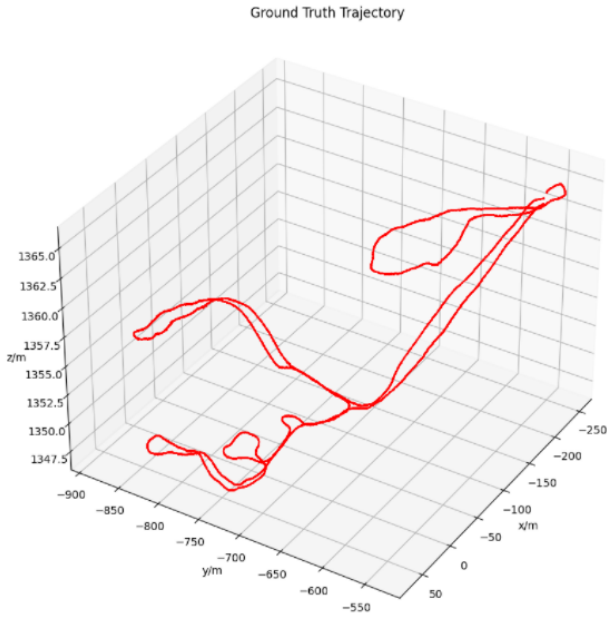


Figure 2. The trajectory of the collection vehicle in 3D perspective

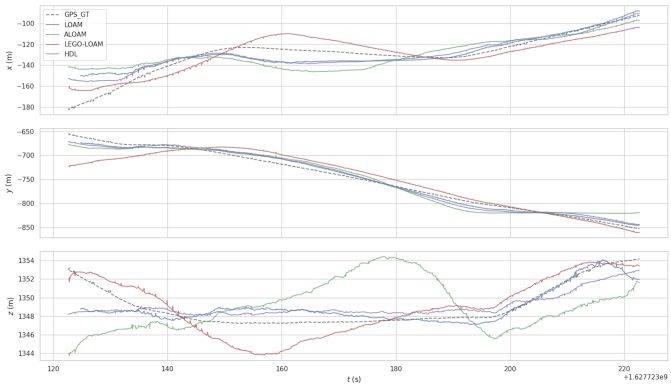


Figure 3. The errors of four outputs from lidar localization methods in x, y, z dimension

2. Appendix B. Performance of lidar and visual localization

Fig. 3 illustrates the deviation of outputs from the four models and the true value from the GPS in the x, y, and z dimension. The model has a large positioning deviation in the z-dimension. As the analysis in the paper, this is related to the fact that the mining area contains various uphills as well as downhills and the characteristics on unstructured roads are sparse. Fig. 7 shows the sparse point cloud on unstructured roads. Therefore, we conclude that the effective features on the mines are much lower than those of the urban scenarios.

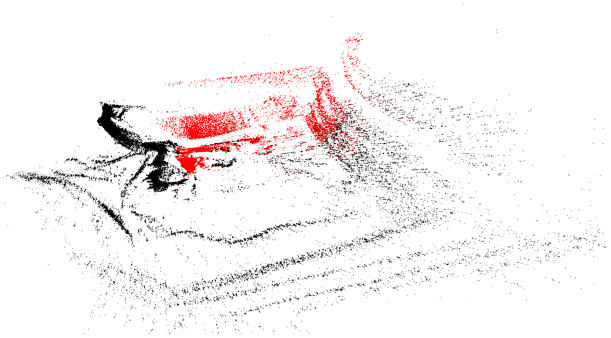


Figure 4. A keypoint map from monocular localization

Fig. 4 is the feature point map in a mine, which is constructed by the key points from the monocular visual positioning algorithm. We notice there is a small amount of key points from the outline of the mining area. Similar to the failure of the lidar localization algorithms, the monotony of the mining scene makes it difficult to capture a sufficient number of key points through images.

3. Appendix C. Sensor installation method and details

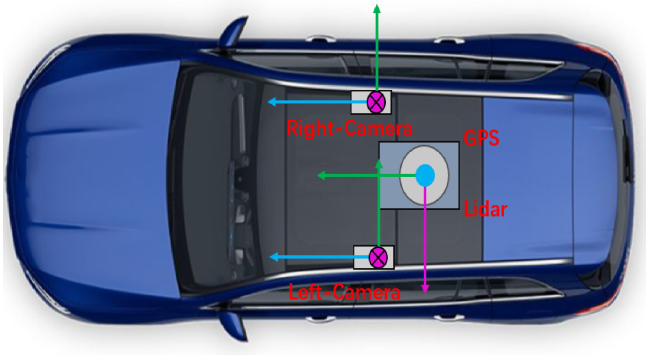
Researchers can find out our sensor placement on the SUV and the mining truck in Fig. 5. See Fig. 5(a) for the SUV installment strategy including two industrial cameras, one Ouster-64 lidar and an inertial navigation system. Fig. 5(b) and Fig. 5(c) are the left and top view of the mining truck. More than one edge lidars are installed on that platform. We list the detailed sensor general parameters in Tab. 1.

4. Appendix D. Calibration

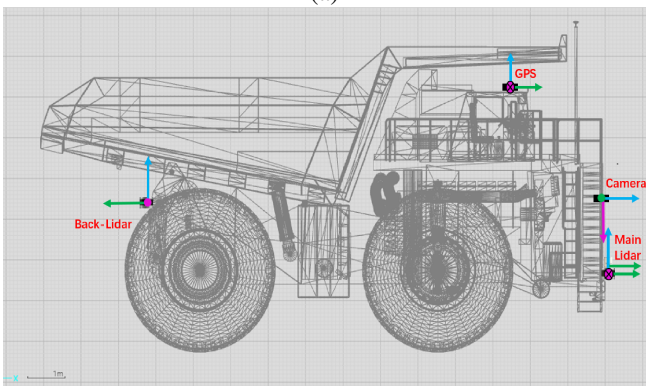
We obtain camera-to-camera, laser-to-laser, and laser-to-camera positional relationships by means of sensor-to-sensor calibration methods. The joint calibration of the left and right cameras requires that the two camera images contain an overlapping area, and in this area, the calibrator moves the calibration plate to obtain the space relationship of the two camera coordinates by matching the same key points.

The joint calibration of the left camera and the main lidar requires the existence of common feature points existing in the picture and in the point cloud. we select two large calibration plates as feature plates, picking the key point cloud, and selecting the pixels of the corresponding point in the image as in Fig. 6. Then, the calibrator changes the position and repeats the steps mentioned above. In the end, a total of eight feature points are selected to calculate the matrix change and we get the joint calibration results.

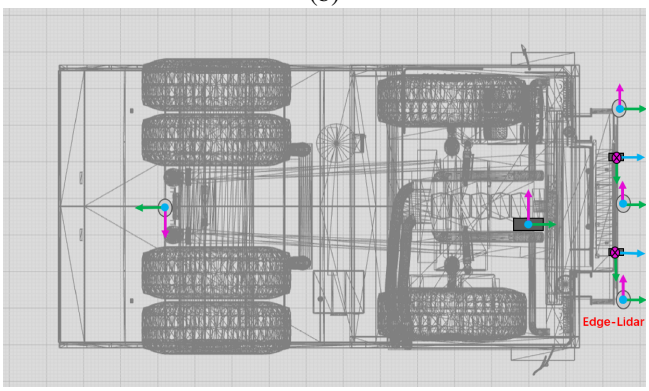
The laser-to-laser calibration method starts by manually



(a)



(b)



(c)

Figure 5. Sensor setup for the SUV and the mining truck

selecting the same key points in both point cloud maps, including the corner points of the control center boardroom, other large trucks and the large calibration plate. The initial transformation matrix is iterated continuously by the geometric consistency assumption. We can obtain the result calibration matrix when the total error is less than a defined threshold or after fixed iterations.

Fig. 7 shows the effect of mapping the point cloud to the picture after joint calibration of the main-lidar and the left camera. The acquisition platform is the wide-body truck, and the joint calibration process for a large truck is much

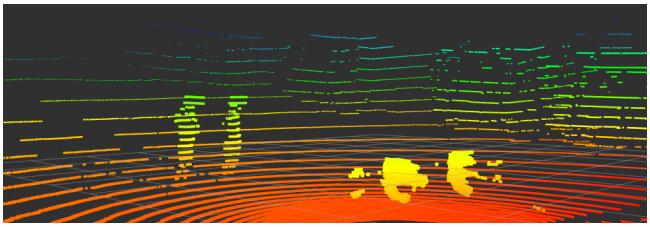


Figure 6. The errors of four outputs from lidar localization methods in x, y, z dimension

more complex than for a passenger vehicle. Finding a common field of view on a large platform is pretty difficult, and there are fewer objects with rich characteristics can be utilized in a mine.

5. Appendix E. Different climatic conditions and unstructured roads

Through the investigation, we notice that intelligent vehicles in mining areas face tough weather and temperature challenges. Fig. 7(a) and Fig. 7(b) are the same scenarios at different weather.Fig. 7(a) is the sandstorm and scholars can observe the dust raised in the lower left corner of the picture interferes the lidar. In addition, the high latitude mining areas will face a low temperature with -30 degree centigrade, which will lead to the unstable output from sensors.

Fig. 7(c) and Fig. 7(d) are two typical unstructured roads in mining areas. Fig. 7(c) is a narrow road for single vehicle traffic only, with low earth slopes on both sides of the route and a relatively rugged surface. Fig. 7(d) is a spacious road for two direction traveling, with the rocky soil on one side. Both kinds of ground are lack of targets and obvious features, on the roads, which is one of the reasons for the visual and lidar positioning errors. In addition, it can be seen that the mining area faces the tricky problem of high light exposure.

6. Appendix F. Parameters set for detection and localization

In this section, we provide training details for localization and 3D perception tasks.

6.1. Implementation for localization

ORB-SLAM2 We set nFeatures as 10000, which is the number of key points. The scaleFactor and nLevels are 1.10 and 20. That means we decide to get more feature pairs from low contrast images.

ORB-SLAM3 We use a high value for nFeatures, 9000. The scaleFactor is 1.15 and the nLevels is 16 in our experiments.

DSO We set the PYR-levels as 6 and allow the extractor obtain a deeper feature map for semantic information because of the low contrast of images in mining areas.

Sensor	Details
Camera	RGB channels, 55Hz capture frequency, 1/1.8" CMOS, 2048×1536 resolution, 70FOV, JPG/PNG compressed
Lidar-Ouster	64 beams, 20Hz capture frequency, 360°horizontal FOV, -7.9° ∼ 7.9°vertical FOV, 150m range, ±3cm accuracy, near 1.3M points/s
Lidar-32	32 beams, 20Hz capture frequency, 360° (180° available) horizontal FOV, -30° ∼ 10° vertical FOV, 70m range, ±2cm accuracy, near 1.4M points/s
Lidar-16	16 beams, 20Hz capture frequency, 360°(270° available) horizontal FOV, -15° ∼ 15°vertical FOV, 70m range, ±3cm accuracy, near 300K points/s
GPS&IMU	0.09°heading, 0.03°roll/pitch(RMS), 20mm position accuracy RTK, 10Hz update
Livox	0.28°(vertical) × 0.03°(horizontal) beam divergence, 38.4°circular FOV, 260m range, ±2cm accuracy, near 100K points/s
Blinding Lidar	20Hz capture frequency, 10cm ∼ 30m range, ±3cm accuracy, 360°horizontal FOV, 90°vertical FOV, near 600K points/s
Radar	0.4m measure resolution, ±0.1m accuracy, -9° ∼ 9°(close)/-45° ∼ 45°(far) horizontal FOV, 18°vertical FOV

Table 1. Sensor general parameters

6.2. Implementation for 3D perception

During the experiments, we set 0.2, 0.25, 0.25, 0.2, 0.1 as the parameter α_{di} in $[0, 10]$, $[10, 20]$, $[20, 35]$, $[35, 60]$, $[60, \inf]$. The choice of these parameters depends on the percentage of the number of targets in this interval.

7. Appendix G. Visualization of mining elements

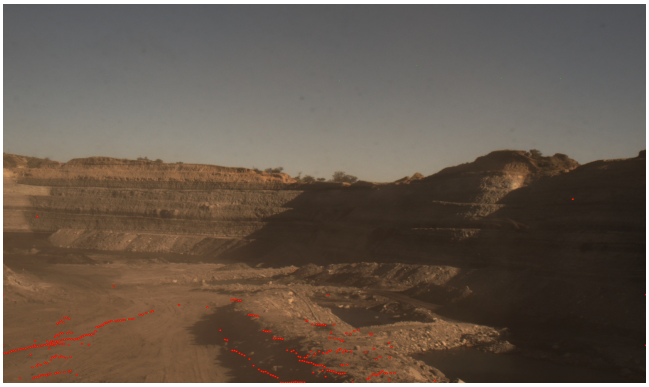
In this section, we show the annotated 3D boxes in the point cloud and corresponding 2D boxes in the image. Our dataset AutoMine includes three crucial characteristics as mentioned in the paper, the unstructured roads, large dimension difference objects with 9 degrees of freedom in extreme climatic conditions and the multi-platform acquisition strategies. Fig. 7, Fig. 8 and Fig. 9 exhibit the features of unstructured roads, which contain few visible road mark

or key pointrugged and rough roads as well as huge elevation Difference, making feature based localization strategies particularly difficult.

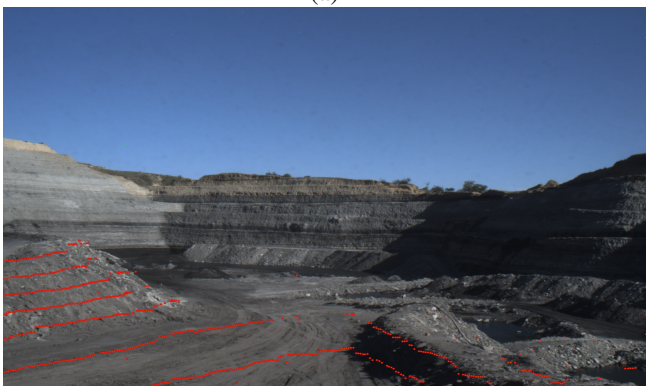
As can be seen in Fig. 8, the data in these images were acquired by the SUV, which has a 180 degrees horizontal FOV of the lidar because it is mounted on the top of the vehicle. In Fig. 8(a), there are wide-body transport trucks, which a massive weight advantage. Fig. 8(b) is a typical night scene in the mining area, where the sparse lighting and strong illumination from oncoming head lights cause difficulty in distinguishing the targets in images. Fig. 8(c) shows single rigid-body trucks, which often can be seen in mines and highways. In addition, the cooperation operation between the excavator and trucks is being carried out in this scene. Fig. 8(d) demonstrates the double bodies truck, which can be divided into two parts: the head and the trailer, that means it can settle the angular singularity.

Fig. 9(a) contains several cooperation scenarios, while another double bodies truck can be seen including a head with load capacity. It can also be observed that the target has a roll and pitch angle. For instance, the target in the figure provides a 15 to 20 degrees deviation of the pitch angle in Fig. 9(b), which is difficult to capture on urban roads. Fig. 9(c) and Fig. 9(d) show multiple trucks parked at the unloading area in details.

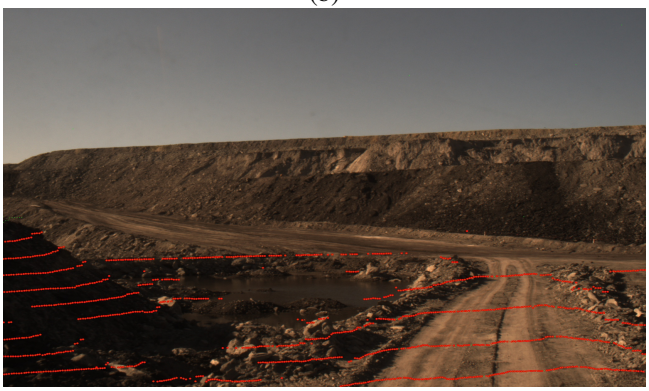
Large mining trucks are found in mining areas with high throughput. Fig. 10 show the data collected by the mining truck, and researchers can notice that the horizontal FOV is only 180 degrees and there is less targets at large mines. Fig. 10(c) are the point cloud from the livox and blinding lidars installed on the front, left and right edge of the wide-body truck (acquisition platform).



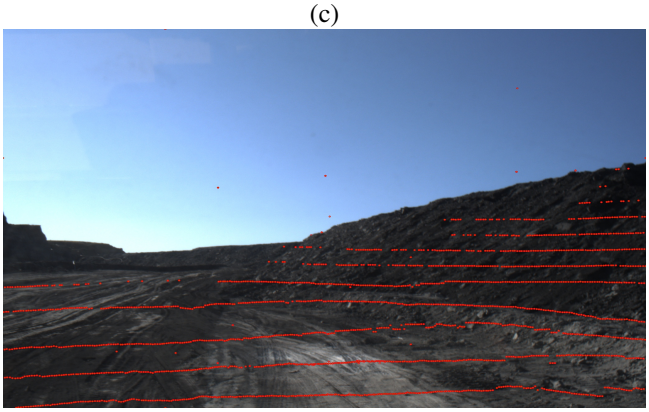
(a)



(b)

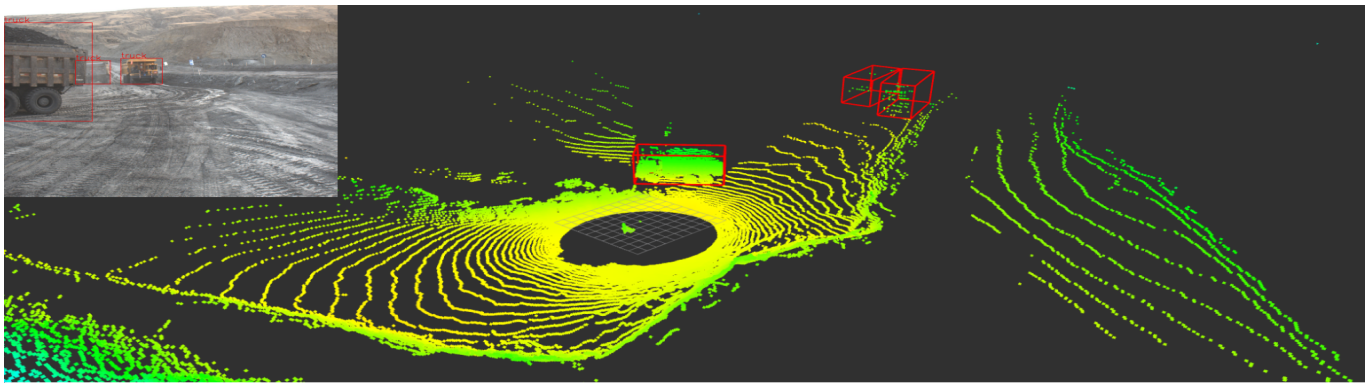


(c)

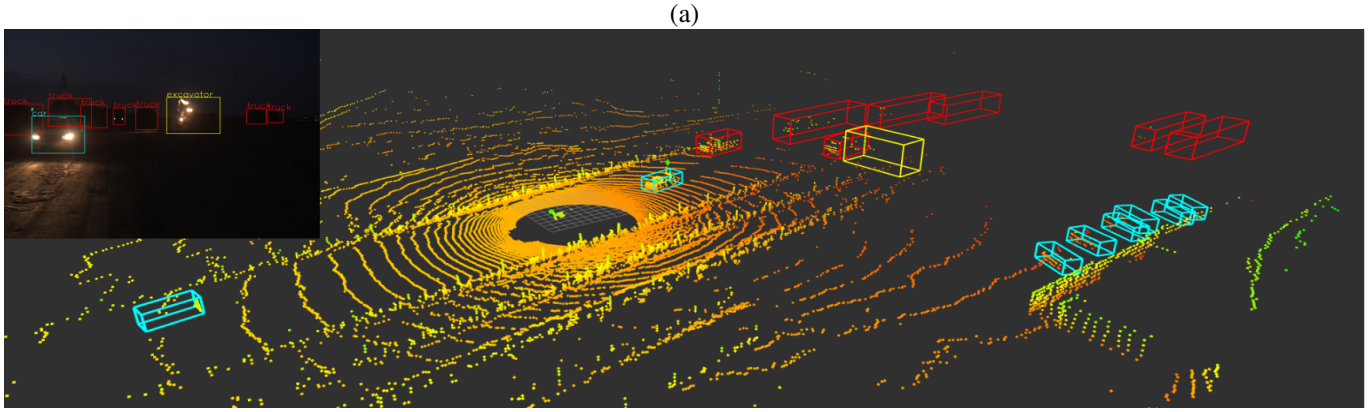


(d)

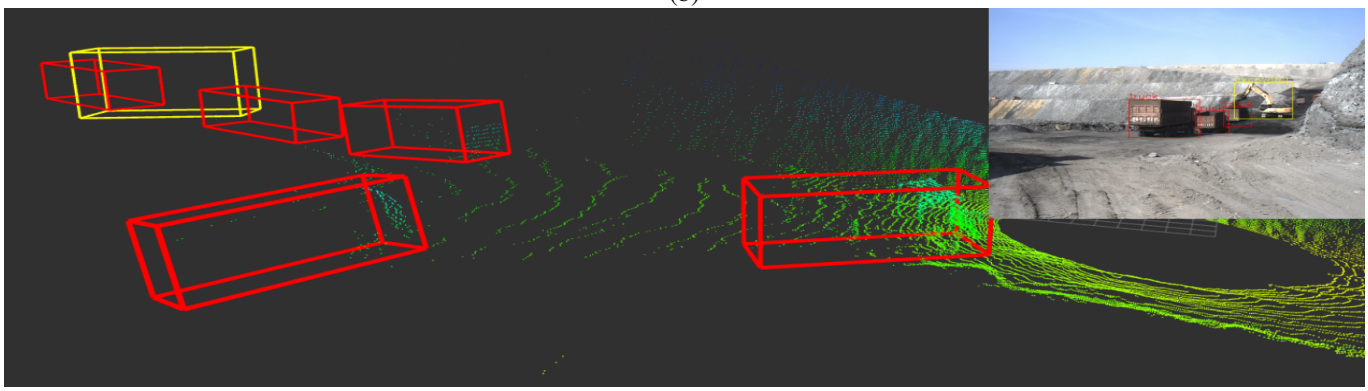
Figure 7. The unstructured roads and association calibration result in mining areas



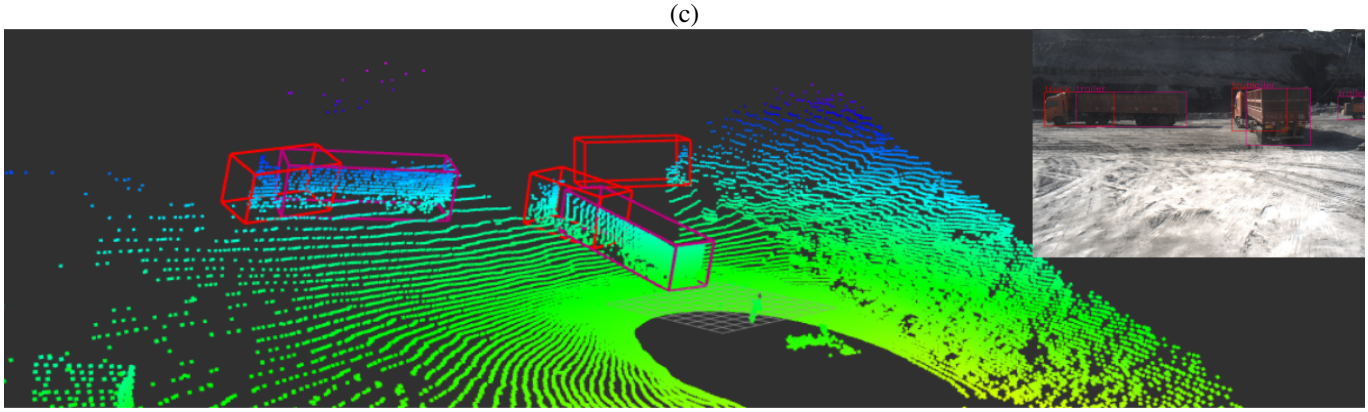
(a)



(b)

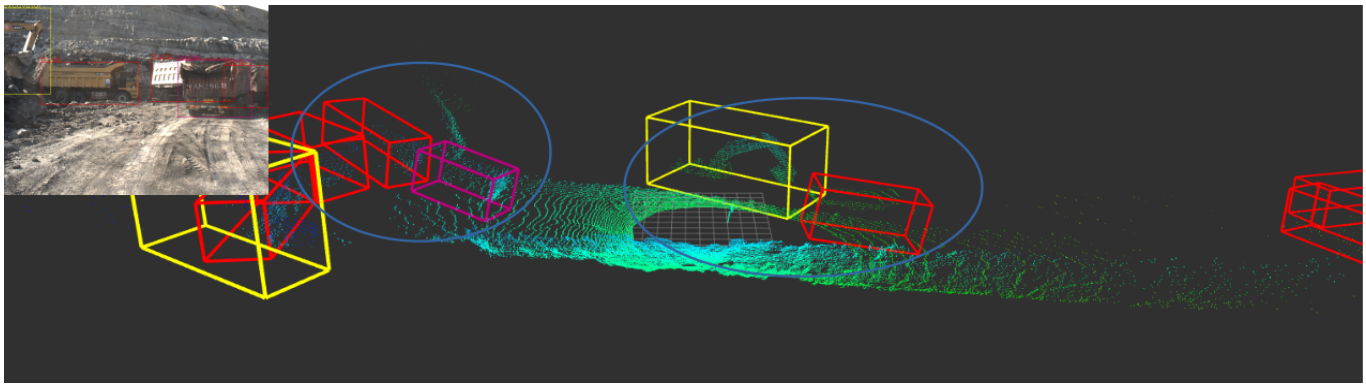


(c)

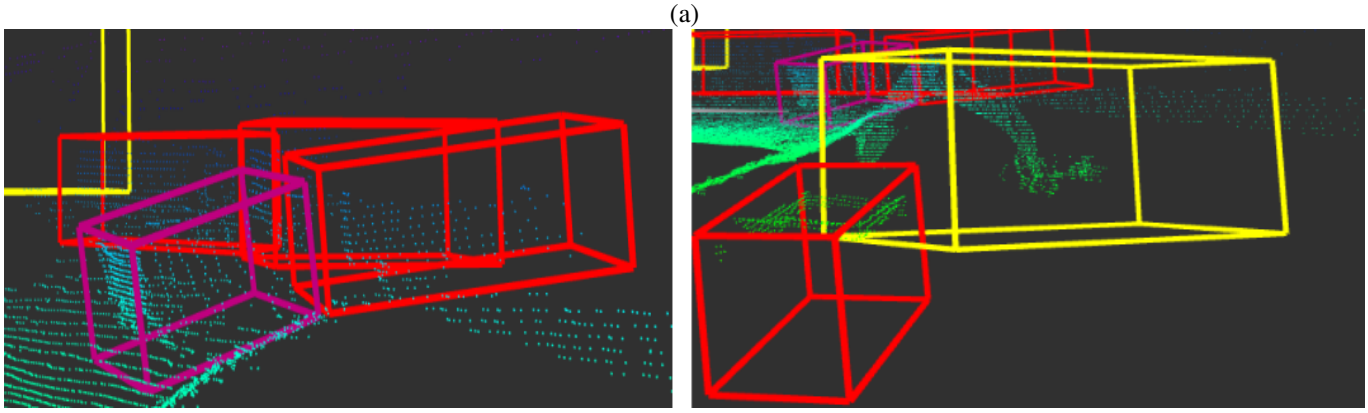


(d)

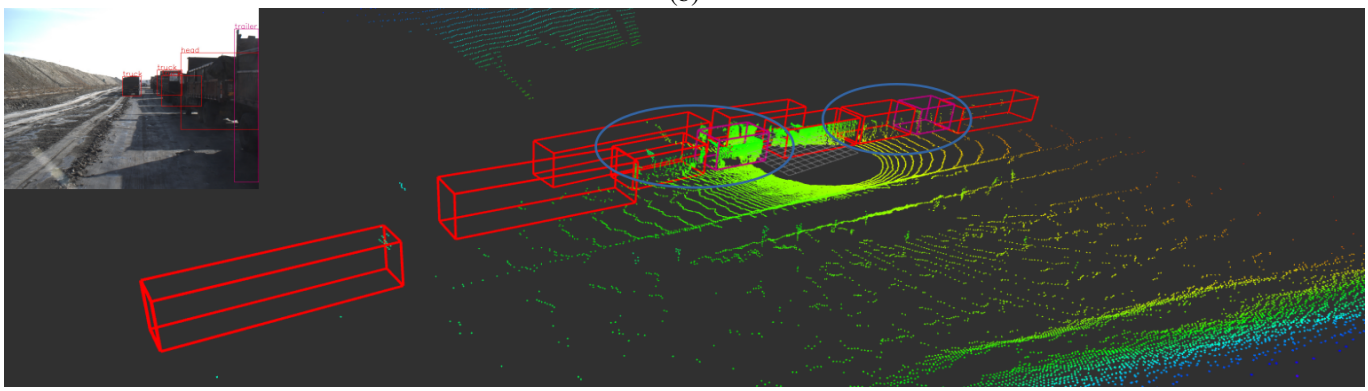
Figure 8. The captured data and labels from the SUV perspective



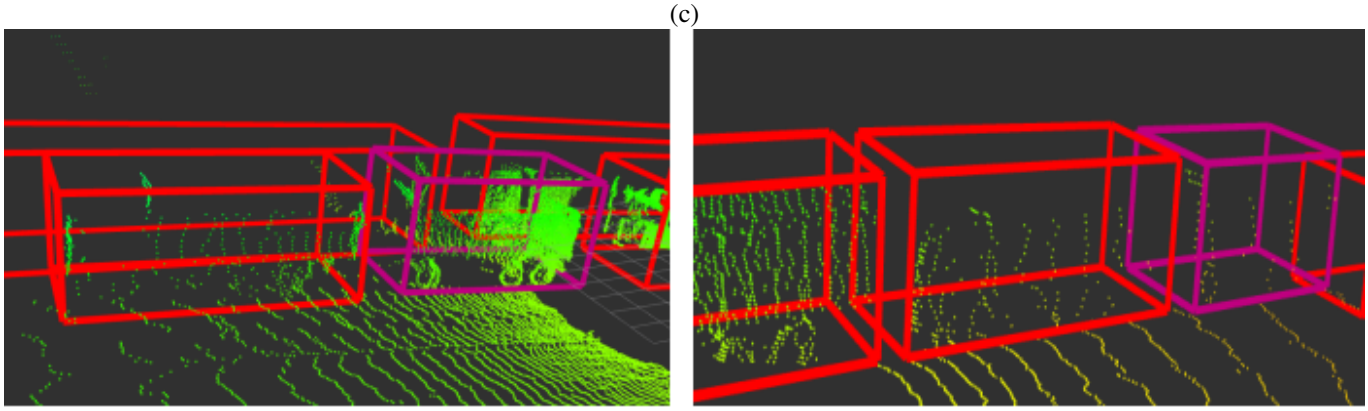
(a)



(b)

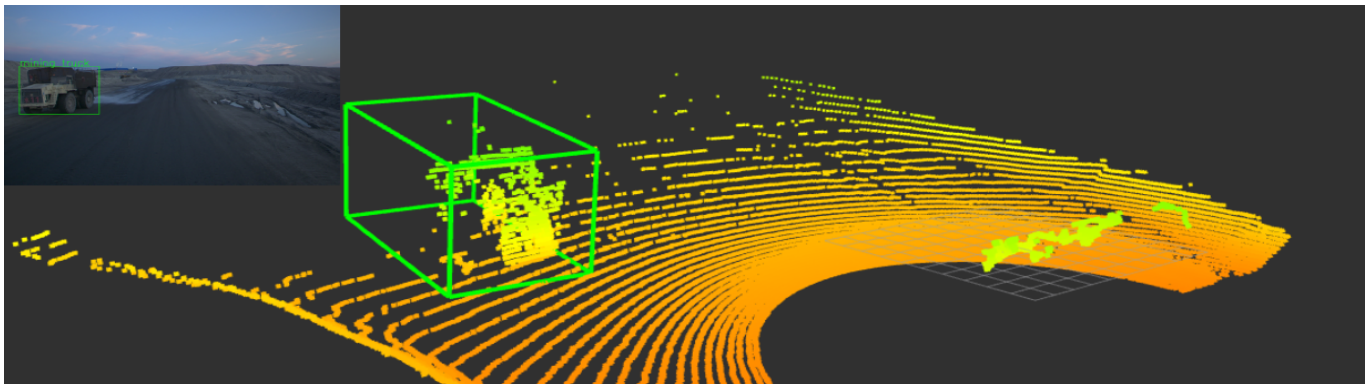


(c)

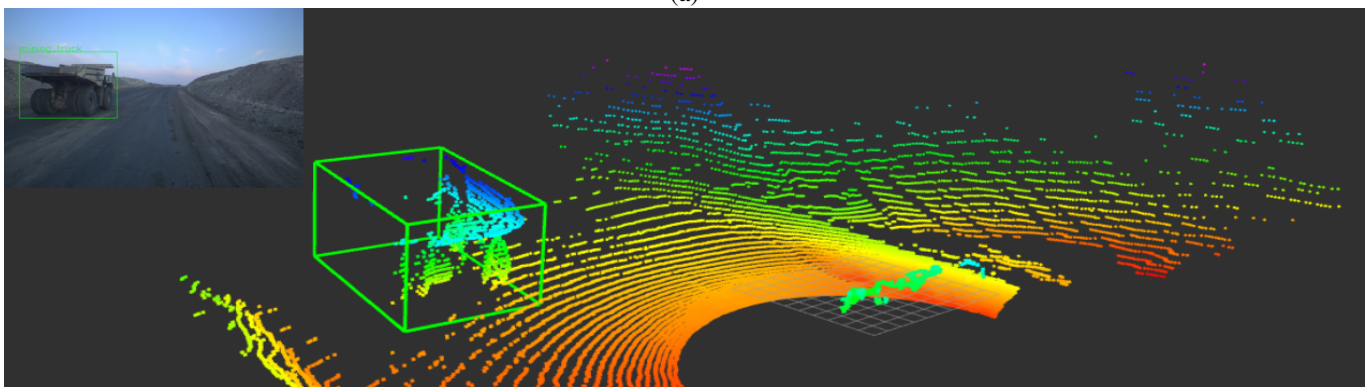


(d)

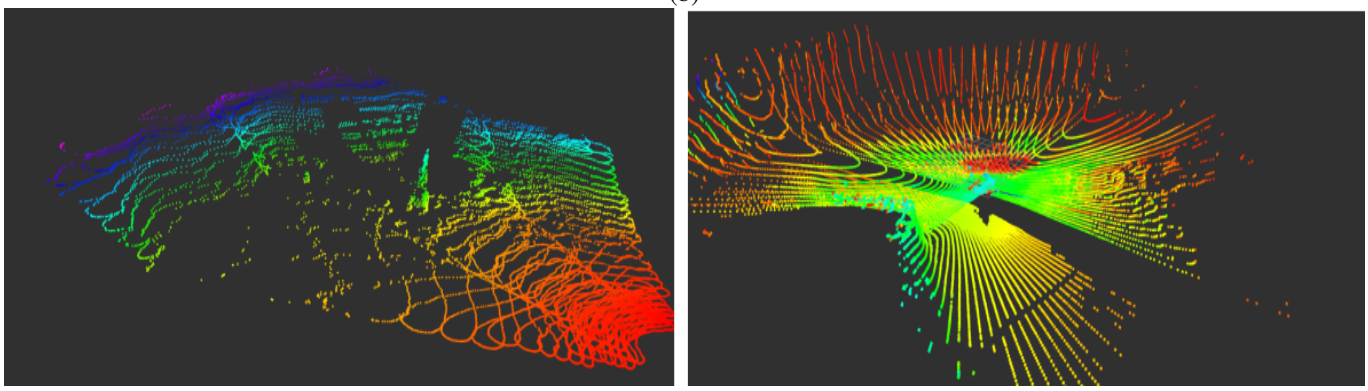
Figure 9. Cooperation operation and various trucks with 9 DoF



(a)



(b)



(c)

Figure 10. The captured data and labels from the truck