General description:

The article presents a conception and application of a high-level perception and localization system for autonomous driving. The system's hardware consists of an inertial navigation system (dual-antenna GNSS receiver; inertial unit measuring acceleration and angular velocity; attitude unit measuring roll, pitch and yaw), two global shutter cameras with 90 degrees lenses, 128 channels rotating lidar, GPS time server, two ethernet switches and an embedded computer based on Nvidia Jetson AGX with expanded data storage.

First reported application of the system was to create a dataset of 10 thousands annotated lidar frames synchronized with measurements taken from other calibrated sensors including shutter cameras. The dataset was created for numerous drives of the vehicle carried out under different weather conditions, different traffic and population densities. It was stated by the authors, that among others the annotated multi-modal datasets including data from 128 channels lidar are rare in the literature.

Second presented application of the system was to implement and test selected perception and localization algorithms which were based on measurements taken mainly from the lidar. The first two perception algorithms were selected as clusterring analysis by RBNN (radially bounded nearest neighbor) and object detection by DCNN (deep convolutional neural network). It was reported that the RBNN reached 98.76% of true positive rate for objects located within a radius of 10 m and 80% fro objects located between 40 and 50 meters from the lidar. The processing time was less than 20 ms with a 10 cm sub-sampling grid.

Additional advantage of the second DCNN is detection of vehicle class. Here, it was reported that the algorithm reached 95% of true positive rate for cars and medium vehicles between 2-20 meters while the detection performance decreases beyond 20 meters from the lidar. In conclusion, it was stated that obtained results of average precision are similar to those reported in the literature. A total prediction time was achieved by the network in time less than 30 ms (with an inference time of about 10-15 ms).

The third algorith of simultaneous localization and mapping was divided into phases of map building essentially based on lidar and inertial measurement data as well as final localization phase made by correlating actual lidar scans to the map. Additionally, drift of the localization algorithm was avoided using GNSS data and loop closures inside a pose graph optimization. The processing time of the localization algorithm was about 15 ms as well as a few centimeters accuracy was reached.

The article consists of 6 sections. In the first section a state-of-the-art was presented in the field of automated driving systems, perception and localization systems and algorithms as well as datasets for algorithm training and validation available in the literature. The second section presents description of the applied platform including perception and localization system installed on an experimental vehicle as well as sensor synchronization and calibration methods. In the third section a creation of the dataset was reffered. The fourth section described implemented algorithms, i.e. RBNN, DCNN and lidar based localization. In the fifth section a discussion was reported while conceptions presented in the article was concluded in the sixth section.

General comments:

In the reviewer's opinion the article fits into the scope of Sensors Journal. It is interesting with respect to the presented experimental setup, sets of analyzed data and types of implemented algorithms. However, several key elements of the article should be improved according to the list of specific comments listed below. They are related among others to the insufficient descritpion of novelty of the article, not complete description of presented algorithms and reported results. As a consequence, the reviewer recommends the article for major revision.

Specific comments:

General: Novelty of the article should be more clearly emphasized in the first section.
Page 6, Line 191: Please refer to the statement that „sequences (..) at 10 Hz were selected" assuming that data from IMU was taken at 50 Hz (Table 1).
Page 8, Line 229: Phrase „intersection air of the convex hull" is not clear.
Page 8, Line 239, Section 4.2.: Application of deep convolutional neural network in the algorithm of object detection should be described in more details including e.g. the structure of the implemented neural network, procedure of training, organization of training, validation and test sets.
Page 9, Line 268: Exact definition of FDR (false discovery rate) was not given.
Page 9, Line 271: TFD quality index was not defined.
Page 10, Line 290: In the reviewer's opinion author should also recall exact values from other studies and not be limited in the reviewed article to the general statement „similar".
Page 10, Line 303, Figure 11b: More details and mathematical description should be given about the implemented Localization algorithm, e.g. presented in the block diagram in Figure 11b.

Sample parts of the article but not limited which are recommended for correction:

Page 2, Line 55: outputs sensors
Page 2, Line 67: It consists in
Page 2, Line 68: 3 camera
Page 3, Line 89: consisted on
Page 4, Line 141: driving scene information's,
Page 6, Line 194: It worth noting
Page 7, Line 206: a simultaneous and localization and mapping algorithms
Page 7, Line 209: Analyze a scene by clustering
Page 10, Line 296: the common used technology
Page 10, Line 307: consist in
Page 11, Line 329: news sensors
Page 11, Line 331: consisted in