The International Journal of
Robotics Research
30(13) 1543–1552

DOI: 10.1177/0278364911400640
ijr.sagepub.com
SAGE

Ford Campus vision and lidar data set

Gaurav Pandey[1], James R McBride[2] and Ryan M Eustice[1]

Abstract

In this paper we describe a data set collected by an autonomous ground vehicle testbed, based upon a modified Ford F-250 pickup truck. The vehicle is outfitted with a professional (Applanix POS-LV) and consumer (Xsens MTi-G) inertial measurement unit, a Velodyne three-dimensional lidar scanner, two push-broom forward-looking Riegl lidars, and a Point Grey Ladybug3 omnidirectional camera system. Here we present the time-registered data from these sensors mounted on the vehicle, collected while driving the vehicle around the Ford Research Campus and downtown Dearborn, MI, during November–December 2009. The vehicle path trajectory in these data sets contains several large- and small-scale loop closures, which should be useful for testing various state-of-the-art computer vision and simultaneous localization and mapping algorithms.

Keywords

Mobile and distributed robotics SLAM, sensing and perception computer vision, field and service robotics

1. Introduction

The University of Michigan and Ford Motor Company have been collaborating on autonomous ground vehicle research since the 2007 DARPA Urban Grand Challenge, and through this continued collaboration have developed an autonomous ground vehicle testbed based upon a modified Ford F-250 pickup truck (Figure 1). This vehicle was one of the finalists in the 2007 DARPA Urban Grand Challenge and showed capabilities to navigate in the mock urban environment, which included moving targets, intermittently blocked pathways, and regions of denied global positioning system (GPS) reception (McBride et al. 2008). This motivated us to collect large-scale visual and inertial data of some real-world urban environments, which might be useful in generating rich, textured, 3D maps of the environment for navigation purposes.

Here we present two data sets collected by this vehicle while driving in and around the Ford Research Campus and downtown Dearborn in Michigan. The data includes various small- and large-scale loop-closure events, ranging from feature rich downtown areas to featureless empty parking lots. The most significant aspect of the data is the precise co-registration of the 3D laser data with the omnidirectional camera imagery thereby adding visual information to the structure of the environment as obtained from the laser data. This fused vision data along with odometry information constitutes an appropriate framework for mobile robot platforms to enhance various state-of-the-art computer vision and robotics algorithms. We hope that this data set will be very useful to the robotics and vision community and will provide new research opportunities in terms of using the image and laser data together, along with the odometry information.

2. Sensors

We used a modified Ford F-250 pickup truck as our base platform. Although, the large size of the vehicle might appear like a hindrance in the urban driving environment, it has proved useful because it allows strategic placement of different sensors. Moreover, the large space at the back of the truck was sufficient enough to install four 2U quad-core processors along with a ducted cooling mechanism. The vehicle was integrated with the following perception and navigation sensors.

[1]University of Michigan, Ann Arbor, MI, USA
[2]Ford Motor Company Research, Dearborn, MI, USA

Corresponding author:
Gaurav Pandey, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48105, USA
Email: pgaurav@umich.edu

Fig. 1. The modified Ford F-250 pickup truck.

2.1. Perception sensors

1. The *Velodyne HDL-64E lidar* (Velodyne 2007) has two blocks of lasers each consisting of 32 laser diodes aligned vertically, resulting in an effective 26.8° vertical FOV. The entire unit can spin about its vertical axis at speeds up to 900 rpm (15Hz) to provide a full 360° azimuthal FOV. The maximum range of the sensor is 120m and it captures about 1 million range points per second. We captured our data set with the laser spinning at 10Hz.
2. The *Point Grey Ladybug3 omnidirectional camera* (Pointgrey 2009) is a high-resolution omnidirectional camera system. It has six 2-megapixel (1,600 × 1,200) cameras with five positioned in a horizontal ring and one positioned vertically. This enables the system to collect video from more than 80% of the full sphere. The camera can capture images at multiple resolutions and multiple frame rates, it also supports hardware jpeg compression on the head. We collected our data set at half resolution (i.e. 1,600 × 600) and 8 fps in raw format (uncompressed).
3. The *Riegl LMS-Q120 lidar* (RIEGL 2010) has an 80° field of view (FOV) with very fine (0.2° per step) resolution. Two of these sensors are installed at the front of the vehicle and the range returns and the intensity data corresponding to each range point are recorded as the laser sweeps the FOV.

2.2. Navigation Sensors

1. The *Applanix POS-LV 420 INS with Trimble GPS* (Applanix 2010) is a professional-grade, compact, fully integrated, turnkey position and orientation system combining a differential GPS, an inertial measurement unit (IMU) rated with 1° of drift per hour, and a 1,024-count wheel encoder to measure the relative position, orientation, velocity, angular rate and acceleration estimates of the vehicle. In our data set we provide the six-degree-of-freedom (6-DOF) pose estimates obtained by integrating the acceleration and velocity estimates provided by this system at a rate of 100 Hz.
2. The *Xsens MTi-G* (Xsens 2010) is a consumer-grade, miniature size and low-weight 6-DOF microelectromechanical system (MEM) IMU. The MTi-G contains accelerometers, gyroscopes, magnetometers, an integrated GPS receiver, static pressure sensor and a temperature sensor. Its internal low-power signal processor provides real time and drift-free 3D orientation as well as calibrated 3D acceleration, 3D rate of turn, and 3D velocity of the vehicle at 100 Hz. It also has an integrated GPS receiver that measures the GPS coordinates

of the vehicle. The 6-DOF pose of the vehicle at any instance can be obtained by integrating the 3D velocity and 3D rate of turn.

2.3. Data capture

In order to minimize the latency in data capture, all of the sensor load is evenly distributed across the four quad-core processors installed at the back of the truck. Time synchronization across the computer cluster is achieved by using a simple network time protocol (NTP)-like (Mills 2006) method whereby one computer is designated as a 'master' and every other computer continually estimates and slews its clock relative to the master's clock. This is done by periodically exchanging a small packet with the master and measuring the time it takes to complete the round trip. We estimate the clock skew to be upper bounded by 100 μs based upon observed round-trip packet times.

When sensor data is received by any of these synchronized host computers, it is timestamped along with the timestamp associated with the native hardware clock of the sensor device. These two timestamps are then merged according to the algorithm documented in Olson (2010) to determine a more accurate timestamp estimate. This sensor data is then packaged into a Lightweight Communication and Marshalling (LCM) (Huang et al. 2010) packet and is transmitted over the network using multicast user datagram protocol (UDP). This transmitted data is captured by a logging process, which listens for such packets from all of the sensor drivers, and stores them on the disk in a single log file. The data recorded in this log file is timestamped again, which allows for synchronous playback of the data later on.

3. Sensor calibration

All of the sensors (perceptual and navigation) are fixed to the vehicle and are related to each other by static coordinate transforms. These rigid-body transformations, which allow the re-projection of any point from one coordinate frame to the other, were calculated for each sensor. The coordinate frames of the two navigation sensors (Applanix and MTi-G) coincide and are called the body frame of the vehicle—all other coordinate frames are defined with respect to the body frame (Figure 2). Here we use the Smith et al. (1988) coordinate frame notation to represent the 6-DOF pose of a sensor coordinate frame where $X_{ab} = [x, y, z, roll, pitch, yaw]^\top$ denotes the 6-DOF pose of frame b with respect to frame a. The calibration procedure and the relative transformation between the different coordinate frames are described in the following and summarized in Table 1. We also use the concept of a local frame (Moore et al. 2009), which is a smoothly varying coordinate system with arbitrary origin. The vehicle moves in this local frame according to the best available relative motion estimate coming from the IMU.

3.1. Relative transformation between the Velodyne laser scanner and body frame (X_{bl})

A research grade coordinate measuring machine (CMM) was used to precisely obtain the position of some known reference points on the truck with respect to the body frame, which is defined to be at the center of the rear axle of the truck. The measured CMM points are denoted by X_{bp}. Typical precision of a CMM is of the order of micrometers, thus for all practical purposes we assumed that the relative position (X_{bp}) of these reference points obtained from CMM are true values without any error. We then manually measured the position of the Velodyne from one of these reference points to get X_{pl}. Since the transformation X_{pl} is obtained manually, the uncertainty in this transformation is of the order of a few centimeters, which for all practical purposes can be considered to be 2–5 cm. The relative transformation of the Velodyne with respect to the body frame is thus obtained by compounding the two transformations (Smith et al. 1988)

$$X_{bl} = X_{bp} \oplus X_{pl}. \tag{1}$$

3.2. Relative transformation between the Riegl lidars and the body frame (left Riegl = X_{bR_l}, right Riegl = X_{bR_r})

These transformations are also obtained manually with the help of the CMM as described above in Section 3.1.

3.3. Relative transformation between the Velodyne laser scanner and Ladybug3 camera head (X_{hl})

This transformation allows us to project any 3D point in the laser reference frame into the camera head's frame and thereby into the corresponding camera image. The calibration procedure requires three or more views of a checkerboard pattern to be viewed simultaneously from both the camera and lidar. The 3D points lying on the checkerboard pattern in the laser reference frame and the normal to the plane in the camera reference frame, obtained from the image using Zhang's method (Zhang 2000), constrains the rigid body transformation between the laser and the camera reference frame. We can formulate this constraint as a non-linear optimization problem to solve for the optimum rigid body transformation. The details of the procedure can be found in Pandey et al. (2010). The transformation obtained using this method is X_{hl}.

3.4. Relative transformation between the Ladybug3 camera head and body frame (X_{bh})

Once we have X_{bl} and X_{hl}, the transformation X_{bh} can be calculated using the compounding operation (Smith et al. 1988):

$$X_{bh} = X_{bl} \oplus (\ominus X_{hl}). \tag{2}$$

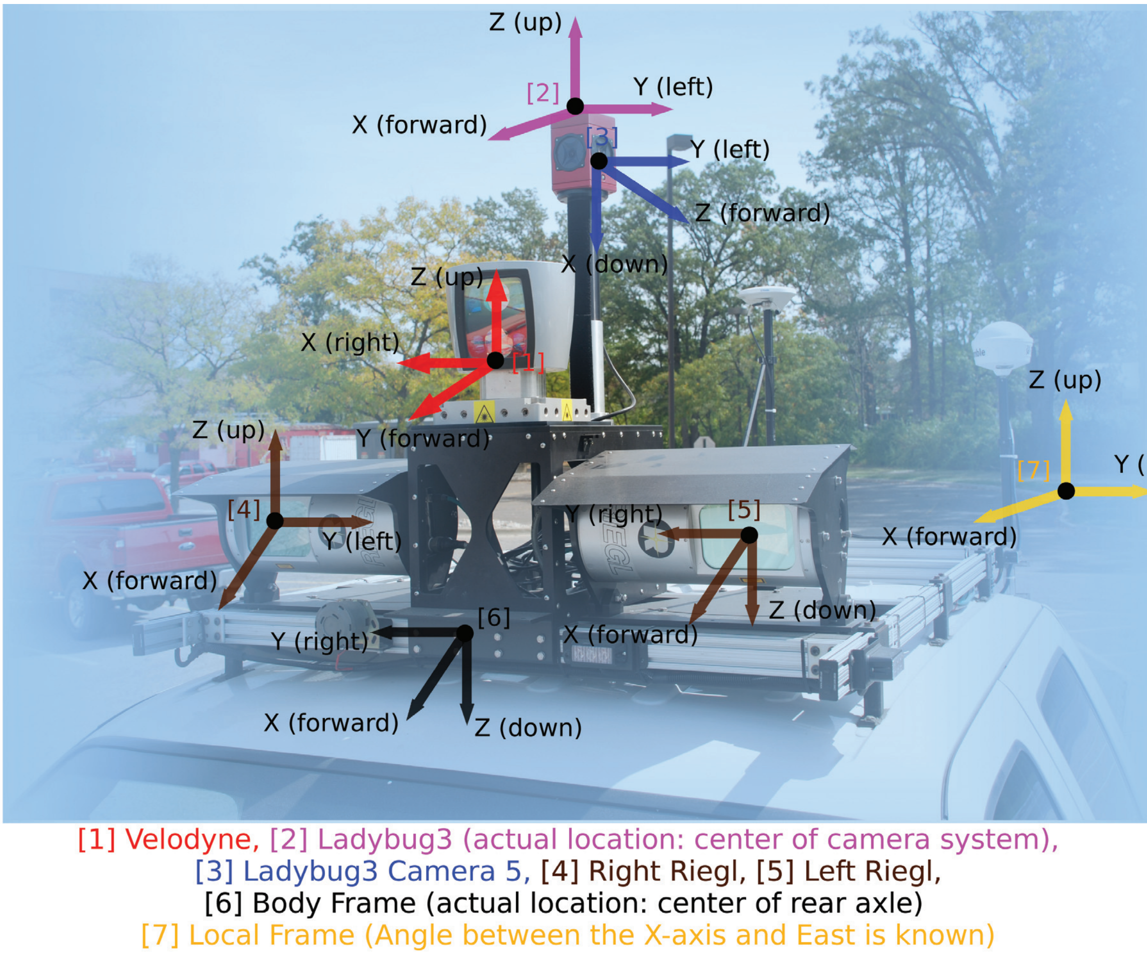


Fig. 2. Relative position of the sensors with respect to the body frame.

Table 1. Relative transformation of sensors.

Transform	Value (meters and degrees)
X_{bl}	[2.4, –0.01, –2.3, 180°, 0°, 90°]
X_{bR_l}	[2.617, –0.451, –2.2, 0°, 12°, 1.5°]
X_{bR_r}	[2.645, 0.426, –2.2, 180°, 6°, 0.5°]
X_{hl}	[.34, –0.01, –0.42, 0.02°, –0.03°, –90.25°]
X_{bh}	[2.06, 0, –2.72, –180°, –0.02°, –0.8°]

4. Data collection

The data was collected around the Ford research campus area and downtown area in Dearborn, MI, henceforth referred to as the test environment. It is an ideal data set representing an urban environment and it is our hope that it will be a useful data set to researchers working on autonomous perception and navigation in unstructured urban scenes.

The data was collected while driving the modified Ford F-250 around the test environment several times while covering different areas. We call each data collection exercise to be a trial. In every trial we collected the data keeping in mind the requirements of state of the art simultaneous localization and mapping (SLAM) algorithms, thereby covering several small and large loops in our data set. A sample trajectory of the vehicle in one of the trials around downtown Dearborn and around Ford Research Complex is shown in Figure 4. The unprocessed data consists of three main files for each trial. One file contains the raw images of the omnidirectional camera, captured at 8 fps, and the other file contains the timestamps of each image measured in microseconds since 00:00:00 1 January 1970 Coordinate Universal Time (UTC). The third file contains the data coming from remaining sensors (perceptual/navigational). This file stores the data in a LCM log file format similar to that described by Huang et al. (2010). The unprocessed data consist of raw spherically distorted images from the omnidirectional camera and the raw point cloud from the lidar (without any compensation for the vehicle motion). Here we present a set of processed data with MATLAB scripts that allow easy access of the data set to the users. The processed data is organized in folders and the directory structure is as shown in Figure 3. The main files and folders are described in the following:

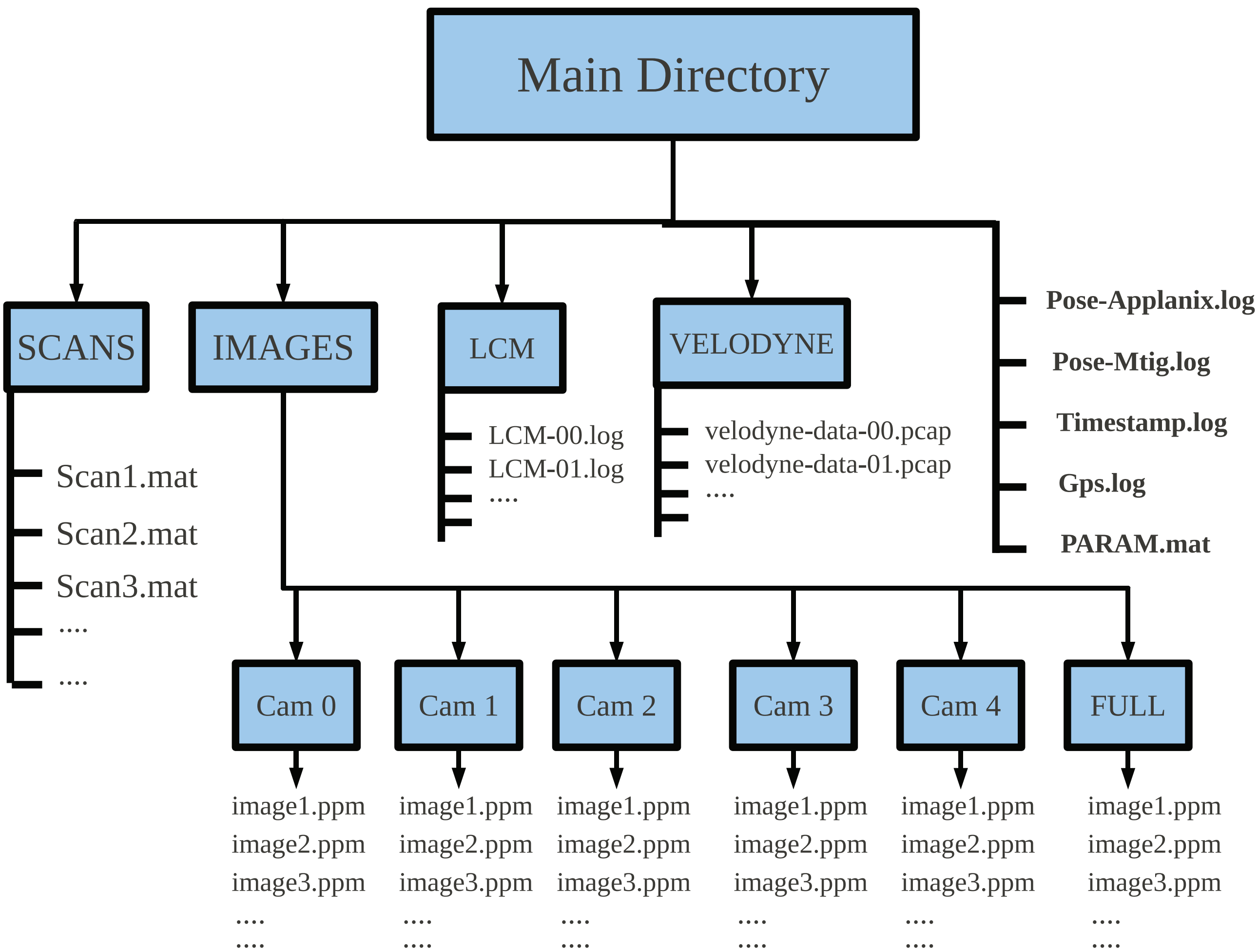
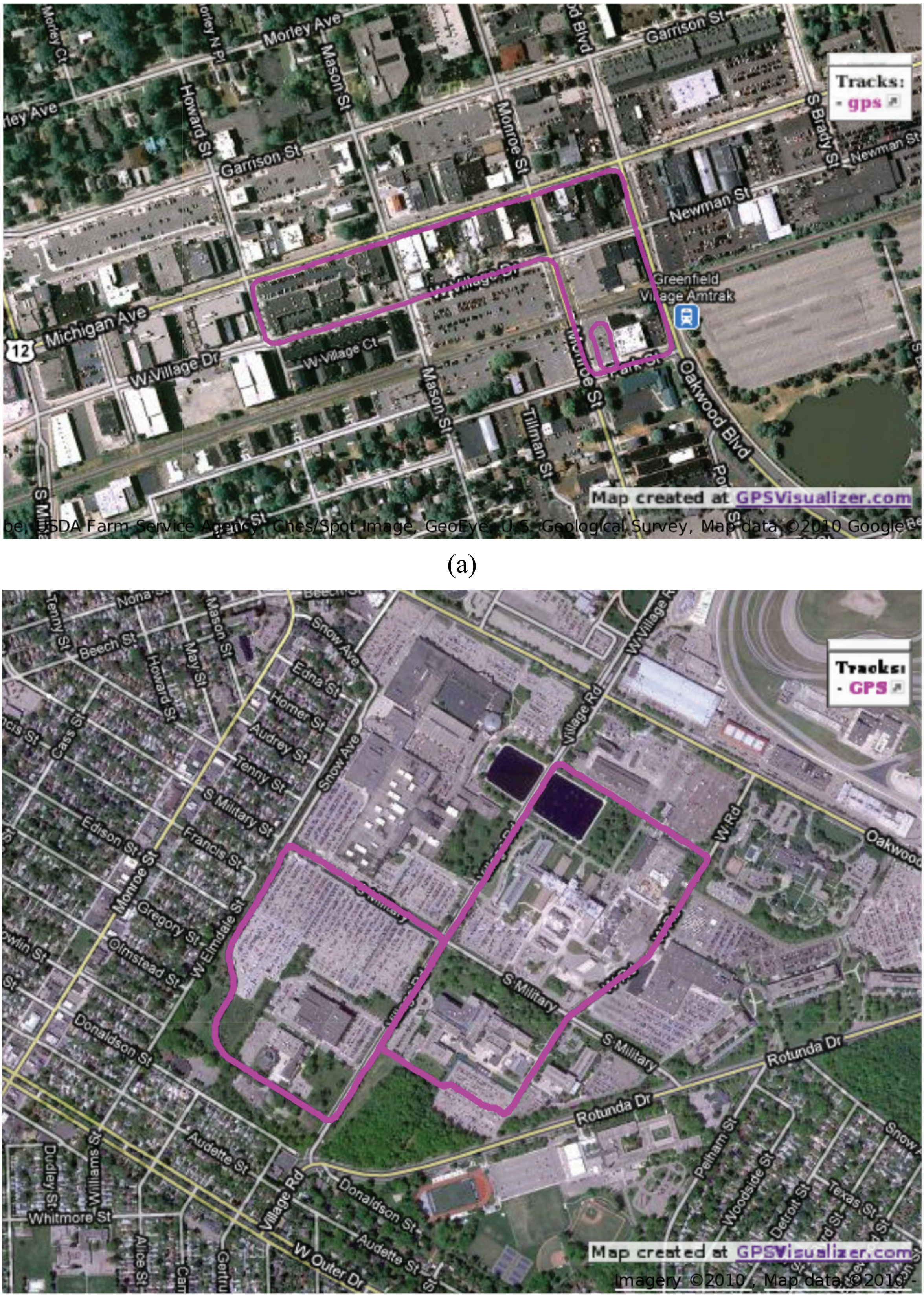


Fig. 3. The directory structure containing the data set. The rectangular blocks represent folders.

1. **LCM**: This folder contains the LCM log file corresponding to each trial. Each LCM log file contains the raw 3D point cloud from the Velodyne laser scanner, the lidar data from the two Reigl LMS-Q120s and the navigational data from the navigational sensors described in Section 2. We provide software to playback this log file and visualize the data in an interactive graphical user interface.
2. **Timestamp.log**: This file contains the Unix timestamp, measured in microseconds since 00:00:00 1 January 1970 UTC, of each image captured by the omnidirectional camera during one trial.
3. **Pose-Applanix.log**: This file contains the 6-DOF pose of the vehicle in a local coordinate frame, as described by Moore et al. (2009). The local frame is arbitrarily fixed at the location where the Applanix is initialized for the first time (Ford parking lot) and all of the vehicle poses are reported with respect to this local frame. The angle θ that the x-axis of this local frame makes with East is known and is recorded in the Gps.log (see Gps.log below). The local frame is thus an East-North-Up (ENU) coordinate frame, but rotated by an angle θ. The acceleration and rotation rates given by the IMU (Applanix POS-LV) are first transformed into the local frame and then integrated to obtain the pose of the vehicle in this local reference frame. The Pose-Applanix.log when loaded in MATLAB has the following fields for each vehicle pose:

   (a) *Pose.utime* is the Unix timestamp measured in microseconds.
   (b) *Pose.pos* is the 3-DOF position of the vehicle in the local reference frame.
   (c) *Pose.rph* is the roll, pitch and heading of the vehicle. Here the heading is with respect to the orientation of the vehicle when the Applanix is initialized. It is not the true heading calculated with respect to East. In order to get the true heading you need to subtract the angle that the local frame makes with East, which is given in the GPS.log.
   (d) *Pose.vel* is the 3-DOF velocity of the vehicle in the local reference frame.
   (e) *Pose.rotation_rate* is the angular velocity of the vehicle.
   (f) *Pose.accel* is the 3-DOF acceleration of the vehicle in local reference frame.



(a)

(b)

Fig. 4. (a) The trajectory of the vehicle in one trial around downtown Dearborn. (b) The trajectory of the vehicle in one trial around Ford Research Complex in Dearborn. Here we have plotted the GPS data coming from the Trimble overlaid atop of an aerial image from Google maps.

(g) *Pose.orientation* is the orientation of the vehicle given in quaternions.

We have not fused the IMU data with the GPS in our data set, but we provide the GPS data separately along with the uncertainties in the GPS coordinates. We have not provided the uncertainty in the pose estimates in our data set but it can be calculated from the measurement noise obtained from the Applanix POS-LV specification sheet (Applanix 2010).

4. **Pose-Mtig.log**: This file contains the navigational data: 3D rotational angles (roll, pitch and yaw), 3D accelerations and 3D velocities of the vehicle along with the timestamp provided by the Xsens MTi-G, during one trial. Here we provide the vehicle pose estimated by integrating the velocities. The raw data is available in the LCM log file and can be extracted from there. The Pose-Mtig.log when loaded in MATLAB has the following fields for each vehicle pose:

   (a) *Pose.utime* is the Unix timestamp measured in microseconds.
   (b) *Pose.pos* is the 3-DOF position of the vehicle in North-East-Down (NED) coordinate frame with origin at the position where the vehicle starts.
   (c) *Pose.rph* is the Euler roll, pitch and heading of the vehicle. Heading here is reported with respect to North.
   (d) *Pose.vel* is the 3-DOF velocity of the vehicle in the NED reference frame
   (e) *Pose.rotation_rate* is the angular velocity of the vehicle.
   (f) *Pose.accel* is the 3-DOF acceleration of the vehicle.

5. **Gps.log**: This file contains the GPS data of the vehicle along with the uncertainties provided by the Trimble GPS, obtained during one trial. This data structure has the following fields:

   (a) *Gps.utime* is the Unix timestamp measured in microseconds.
   (b) *Gps.lat_lon_el_theta* is the $[4 \times 1]$ array of GPS coordinates. The first three entries are the latitude, longitude and elevation/altitude whereas the last entry *theta* is the angle that the x-axis of the local frame makes with East (i.e. θ).
   (c) *Gps.cov*: is the $[4 \times 4]$ covariance matrix representing the uncertainty in the GPS coordinates.

6. **IMAGES**: This folder contains the undistorted images captured from the omnidirectional camera system during one trial. The folder is further divided into subfolders containing images corresponding to individual cameras from the omnidirectional camera system. This folder also contains a folder named 'FULL', which contains the undistorted images stacked together in one file as depicted in Figure 5.

7. **PARAM.mat**: This contains the intrinsic and extrinsic parameters of the omnidirectional camera system and is a (1×5) array of structures with the following fields:

   (a) *PARAM.K*: This is the (3×3) matrix of the internal parameters of the camera.
   (b) *PARAM.R, PARAM.t*: These are the (3×3) rotation matrix and (3×1) translation vector, respectively, which transforms the 3D point cloud from the laser reference system to the camera reference system. These values were obtained by compounding the following transformations:

$$X_{c_il} = X_{c_ih} \oplus X_{hl}, \tag{3}$$

   where X_{c_ih} defines the relative transformation between camera head and the ith sensor (camera) of the omnidirectional camera system. The transformation X_{c_ih} is precisely known and is provided by the manufacturers of the camera.
   (c) *PARAM.MappingMatrix*: This contains the mapping between the distorted and undistorted image pixels. This mapping is provided by the manufacturers of the Ladybug3 camera system and it corrects for the spherical distortion in the image. A pair of distorted and undistorted images from the camera is shown in Figure 6.

8. **SCANS**: This folder contains 3D scans from the Velodyne laser scanner, motion compensated by the vehicle pose provided by the Applanix POS-LV 420 IMU. Each scan file in this folder is a MATLAB file (.mat) that can be easily loaded into the MATLAB workspace. The structure of individual scans once loaded in MATLAB is shown in the following:

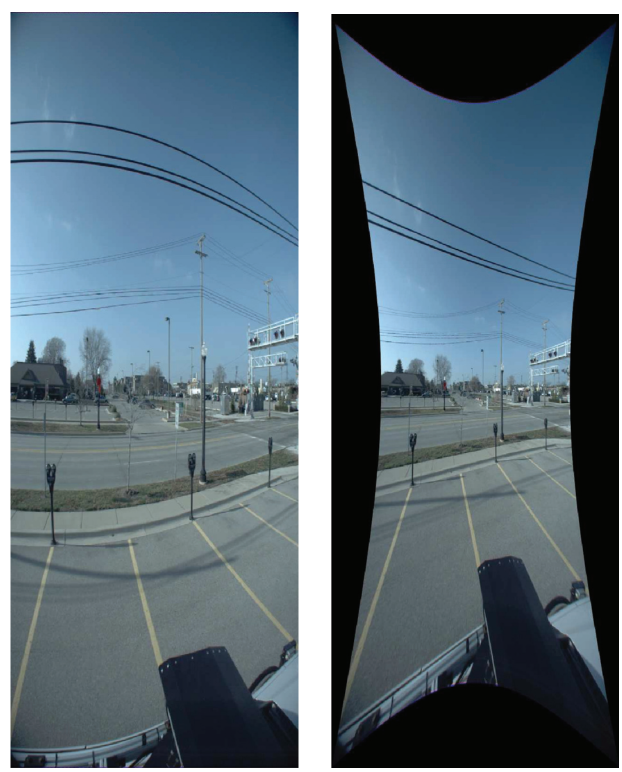
   (a) *Scan.XYZ*: is a $(3 \times N)$ array of the motion compensated 3D point cloud represented in the Velodyne's reference frame (described in Section 3). Here N is the number of points per scan, which is typically $\sim$80,000–100,000.
   (b) *Scan.timestamp_laser*: is the Unix timestamp measured in microseconds since 00:00:00 1 January 1970 UTC for the scan captured by the Velodyne laser scanner.
   (c) *Scan.timestamp_camera*: is the Unix timestamp measured in microseconds since 00:00:00 1 January 1970 UTC for the closest image (in time) captured by the omnidirectional camera.
   (d) *Scan.image_index*: is the index of the image that is closest in time to this scan.
   (e) *Scan.X_wv*: is the 6-DOF pose of the vehicle in the world reference system when the scan was captured.
   (f) *Scan.Cam*: is a (1×5) array of structures corresponding to each camera of the omnidirectional camera system. The format of this structure is given as follows:

Fig. 5. The distorted images, from the five horizontal sensors of the omnidirectional camera, stacked together.

- *Scan.Cam.points_index*: is a $(1 \times m)$ array of index of the 3D points in laser reference frame, in the field of view of the camera.
- *Scan.Cam.xyz*: is a $(3 \times m)$ array of 3D laser points $(m < N)$ as represented in the camera reference system and within the field of view of the camera.
- *Scan.Cam.pixels*: This is a $(2 \times m)$ array of pixel coordinates corresponding to the 3D points projected onto the camera.

9. **VELODYNE**: This folder has several '.pcap' files containing the raw 3D point cloud from the Velodyne laser scanner in a format that can be played back at a desired frame rate using our MATLAB playback tool. This allows the user to quickly browse through the data corresponding to a single trial.

We have provided MATLAB scripts to load and visualize the data set into the MATLAB workspace. We have tried to keep the data format simple so that it can be easily used by other researchers in their work. We have also provided some visualization scripts (written in MATLAB) with this data set that allow the re-projection of any scan onto the corresponding omnidirectional imagery as depicted in Figure 7. We have also provided a C visualization tool that uses OpenGL to render the textured point cloud as shown in Figure 7.

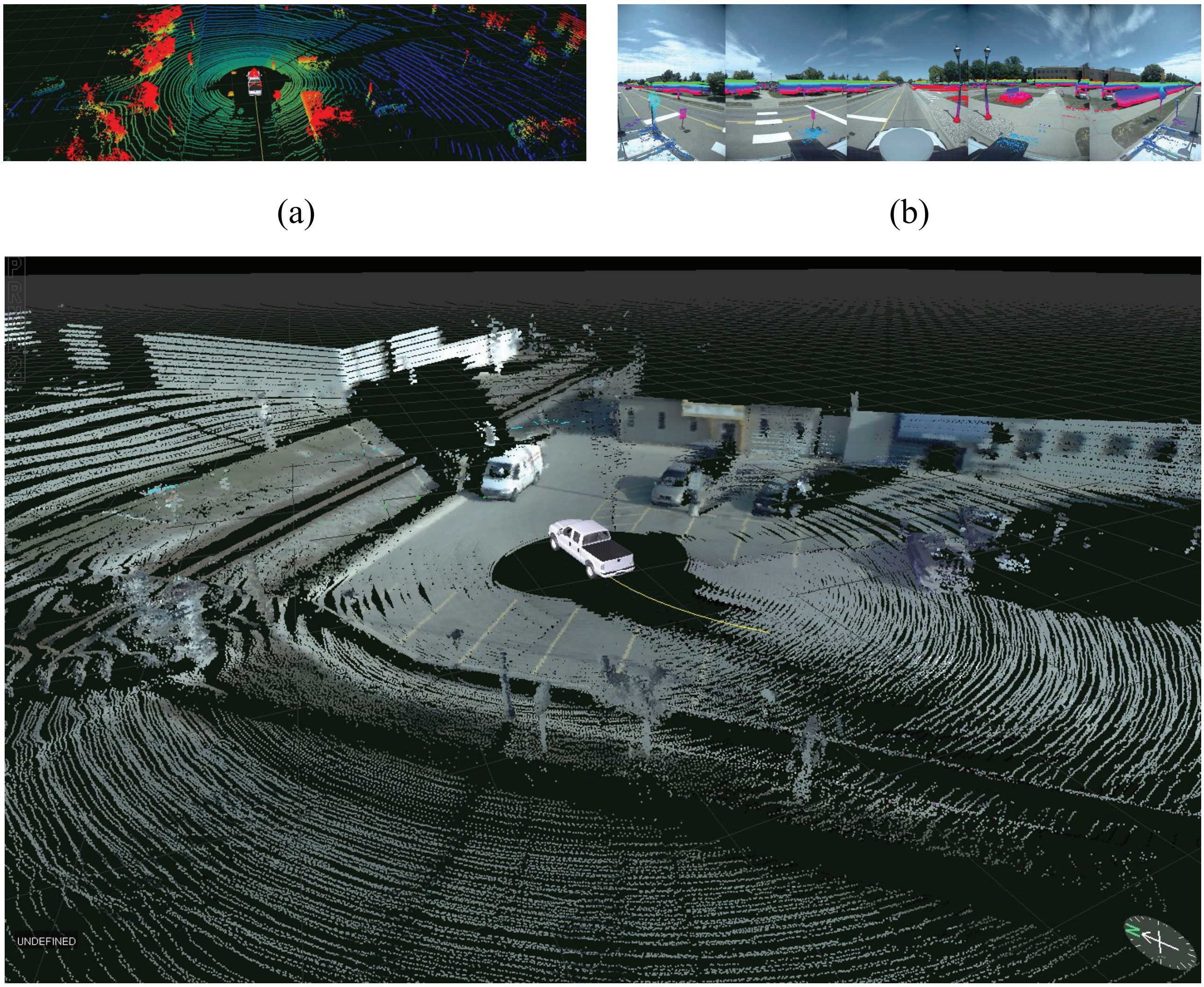


(a) (b)

Fig. 6. (a) A spherically distorted image obtained from the Ladybug3 camera. (b) The corresponding undistorted image obtained after applying the transformation provided by the manufacturer.

4.1. Notes on data

The time registration between the laser and camera data is not exact due to a transmission offset caused by the 800 Mb s^{-1} Firewire bus over which the camera data is transferred to the computer. The camera data is timestamped



(a) (b)

(c)

Fig. 7. (a) A perspective view of the Velodyne lidar range data, color-coded by height above the estimated ground plane. (b) The above-ground-plane range data projected into the corresponding image from the Ladybug3 cameras. (c) Screenshot of the viewer showing the RGB textured point cloud corresponding to a parking lot in downtown Dearborn captured during one of the trials in the month of December.

as soon as it reaches the computer, so there is a time lag between when the data was actually captured at the camera head and the computer timestamp associated with it. We calculated this approximate time lag (=size of image transferred/transfer rate) and subtracted it from the timestamp of the image to reduce the timing latency.

The *Ford Campus vision and lidar data set* is available for download from our server at http://robots.engin.umich.edu/SoftwareData/Ford. Current data sets were collected during November–December 2009, in the future we plan to host more data sets corresponding to some other times of the year.

5. Summary

We have presented a time-registered vision and navigational data set of a downtown urban environment. We believe that this data set will be highly useful to the robotics community, especially to those who are looking toward the problem of autonomous navigation of ground vehicles in an *a priori* unknown environment. The data set can be used as a benchmark for testing various state-of-the-art computer vision and robotics algorithms such as SLAM, iterative closest point (ICP), and 3D object detection and recognition.

Funding

This work was supported by the Ford Motor Company via the Ford-UofM Alliance (award number N009933).

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We are grateful to Professor Silvio Savarese of the University of Michigan for his continued feedback throughout the data collection process.

References

Applanix (2010) POS-LV: Position and orientation system for land vehicles. Specification sheets and related articles available at http://www.applanix.com/products/land/pos-lv.html.

Huang A, Olson E and Moore D (2010) LCM: Lightweight communications and marshalling. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, pp. 4057–4062.

McBride JR, Ivan JC, Rhode DS, Rupp JD, Rupp MY, Higgins JD, et al. (2008) A perspective on emerging automotive safety applications, derived from lessons learned through participation in the darpa grand challenges. *Journal of Field Robotics* 25: 808–840.

Mills D (2006) *Network Time Protocol Version 4 Reference and Implementation Guide*. Technical Report 06-06-1, University of Delaware.

Moore D, Huang A, Walter M, Olson E, Fletcher L, Leonard J, et al. (2009) Simultaneous local and global state estimation for robotic navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, pp. 3794–3799.

Olson E (2010) A passive solution to the sensor synchronization problem. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, pp. 1059–1064.

Pandey G, McBride J, Savarese S and Eustice R (2010) Extrinsic calibration of a 3D laser scanner and an omnidirectional camera. In *7th IFAC Symposium on Intelligent Autonomous Vehicles*.

Pointgrey (2009) Spherical vision products: Ladybug3. Specification sheet and documentations available at http://www.ptgrey.com/products/ladybug3/index.asp.

RIEGL (2010) LMS-Q120: 2D laser scanner. Specification sheet and documentations available at http://www.riegl.com/nc/products/mobile-scanning/produktdetail/product/scanner/14.

Smith R, Self M and Cheeseman P (1988) A stochastic map for uncertain spatial relationships. In *Proceedings of the 4th International Symposium on Robotics Research*. Cambridge, MA: MIT Press, pp. 467–474.

Velodyne (2007). Velodyne HDL-64E: A high definition LIDAR sensor for 3D applications. Available at: http://www.velodyne.com/lidar/products/white_paper.

Xsens (2010). MTi-G a GPS aided MEMS based Inertial Measurement Unit (IMU) and static pressure sensor. Specification sheet and documentations available at http://www.xsens.com/en/general/mti-g.

Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 1330–1334.