SAE INTERNATIONAL

Multi-Modal Image Segmentation for Obstacle Detection and Masking

2014-01-0164

Published 04/01/2014

Cheng-Lung Lee
Univ. of Detroit Mercy

Hongyi Zhang, Hong Nguyen, Yu-Ting Wu, Christopher Smalley, Utayba Mohammad, and Mark J. Paulik
Univ. of Detroit Mercy

CITATION: Lee, C., Zhang, H., Nguyen, H., Wu, Y. et al., "Multi-Modal Image Segmentation for Obstacle Detection and Masking," SAE Technical Paper 2014-01-0164, 2014, doi:10.4271/2014-01-0164.

Abstract

A novel multi-modal scene segmentation algorithm for obstacle identification and masking is presented in this work. A co-registered data set is generated from monocular camera and light detection and ranging (LIDAR) sensors. This calibrated data enables 3D scene information to be mapped to time-synchronized 2D camera images, where discontinuities in the ranging data indicate the increased likelihood of obstacle edges. Applications include Advanced Driver Assistance Systems (ADAS) which address lane-departure, pedestrian protection and collision avoidance and require both high-quality image segmentation and computational efficiency. Simulated and experimental results that demonstrate system performance are presented.

Introduction

Obstacle identification and classification is crucial for an autonomous vehicle to navigate an environment safely. Image processing is a powerful tool for recognizing objects in an environment, but it can be computationally expensive. High quality Image segmentation is a particularly costly image processing technique. One way to improve quality and speed of image segmentation is by combining the information from another sensor such as a light detection and ranging (LIDAR) sensor. For the scope of this work the LMS 111 LIDAR was used. This sensor has a 270 degree field of view (FOV) with quarter degree resolution. This LIDAR provides accurate information about three dimensional obstacles at a rate of either 50 frames per seconds (fps) or 25 fps. In this work the LIDAR was configured to run at the slower frequency which was adequate for the application. This information can be used to identify and classify objects near an autonomous vehicle. For this application the area of interest of the LIDAR is within the FOV of the camera. The camera was used with this work and was mounted at a 60 degree downward angle in front of the robot. Objects recognized by the LIDAR within the view of the camera can be transferred to the camera image and provide fast and accurate obstacle information in the image frame.

The transformation of obstacles from the LIDAR coordinate frame to the camera requires careful camera calibration. This work takes advantage of camera calibration techniques that have already been developed in the Matlab and ROS camera calibration [1].

Calibrating the camera and finding the transformation between the LIDAR and Camera coordinate frames allows for the fusion of these sensors' data. The method of sensor fusion proposed in this work first uses the LIDAR to detect and classify three dimensional objects. The classification of objects in three dimensions allows for accurate and precise scene segmentation in the camera image. Finally the camera calibration results are used to change the perspective of the camera's images into rectified frame for the vehicle. This perspective correction allows the pixels in an image frame to have a direct correlation with a physical location relative to the robot which allows the camera's images to be in mapping or other vehicle perspective oriented applications.

The algorithmic details and results of this work are organized as follows: Section 2 introduces the details of the camera calibration method applied; Section 3 presents the scene segmentation applications; Section 4 includes experimental results; and section 5 presents conclusions and suggested future work.

Camera Calibration

Camera calibration, in the context of robotic computer vision, is the process of determining the geometric and optical characteristics of a camera (intrinsic parameters) and the position and orientation of the camera frame relative to the robot and world frame of reference (extrinsic parameters). The most important application of camera calibration is interpreting 3D information from the 2D view of a camera. The geometric information of objects in space can be used to gain valuable information about 3D objects. [2]

In this application, Robot Operating System (ROS) camera calibration was used to assist with obtaining the intrinsic parameters. Once the intrinsic parameters are found, the next step is to generate the extrinsic parameters. The Matlab camera calibration toolbox assisted in the generation of the extrinsic parameters [1].

In order to derive the intrinsic parameters for the AVT Stingray, the pinhole camera model was assumed. This model uses similar triangles to form a relationship between the physical world and pixels in a camera Charged Coupled Device (CCD). The geometry of this relationship can be generalized into the matrix seen in Equation 1 [1]. This equation encapsulates the information from the camera Focal length and Principal point in a way that can directly measure the scale of image pixels. The parameters f_x and f_y refer to the field of view of the camera in the horizontal and vertical directions respectively. They are generated using the geometry between the focal length and the physical size of the CCD. The skew coefficient γ is the angle offset between the camera and the CCD and is assumed to be zero. The Principal point is represented by the parameters c_x and c_y. This point is located in the center of the CCD.

$$\begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Intrinsic Parameter Matrix

(1)

The extrinsic parameters translate the camera coordinates to the vehicle and world coordinate frame and are obtained based on the geometry of Figure 1. This geometry results in the transformation matrix seen in Equation 2 [1].

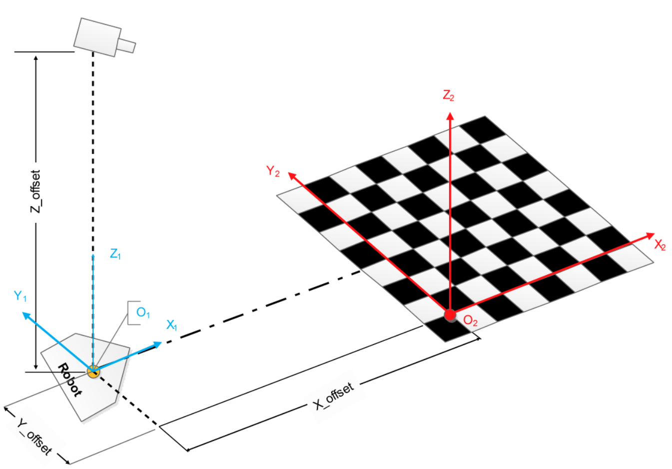

Figure 1. Camera calibration extrinsic parameters diagram

$$\begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r23 & r33 & t3 \end{bmatrix}$$

Extrinsic Parameter Matrix

(2)

The parameters t1, t2 and t3 correspond to Y_offset, X_offset and Z_offset respectively based on how the coordinates are defined in Figure 1. The equations to describe the rotation matrices Rx,Ry and Rz are seen in Equations 3, 4, and 5. These parameters rotate between the camera frame of reference in the x, y, and z directions respectively to match with the vehicle reference frame. The 3×3 sub matrix of Equation 2 from r11 to r33 denote the rotation matrix, which is generated by the straight multiplication of Equations 3, 4, and 5. [1]

$$R_x(th\_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(th\_x) & -\sin(th\_x) \\ 0 & \sin(th\_x) & \cos(th\_x) \end{bmatrix}$$

(3)

$$R_y(th\_y) = \begin{bmatrix} \cos(th\_y) & 0 & \sin(th\_y) \\ 0 & 1 & 0 \\ -\sin(th\_y) & 0 & \cos(th\_y) \end{bmatrix}$$

(4)

$$R_z(th\_z) = \begin{bmatrix} \cos(th\_z) & -\sin(th\_z) & 0 \\ \sin(th\_z) & \cos(th\_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(5)

Obstacle Detection and Recognition Using LIDAR

Obstacle detection is an essential component in any mobile robotic application. Classification of objects can provide an even higher level of cognition in robotics and can allow for tremendous processing improvements. Image segmentation can be vastly improved with the use of 3D information provided by a highly reliable LIDAR sensor. Detection and segmentation of barrels in an image frame was done in this application, but the concept can be extended to include a wider range of objects including humans or other dynamic obstacles.

Barrel Detection Using LIDAR

A LIDAR uses the time-of-flight ranging principle to measure the distance from the sensor to obstacles. LIDAR has several advantageous features including: fast detection speed, highly accurate sensing capability, and high precision. Figure 2 shows an image from a vehicle's perspective and the associated LIDAR distance data. The LIDAR reports its data every time it completes a scan of its entire field of view. This data contains the distances of every physical object from the sensor at every angle. The data is received as a vector where the first element is the leftmost point in the field of view and the last element is the rightmost point.

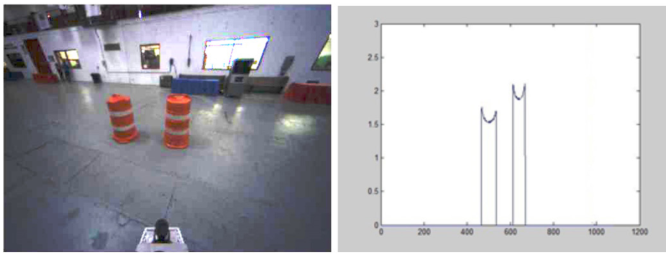

Figure 2. Vehicle Perspective Image and Corresponding LIDAR Data Diagram

An obstacle can be located relative to a vehicle based on its edges. In order to recognize the edges, the differential of the received data vector is computed. The result of this operation can be seen in Figure 3. The positive spikes indicate the left edge of an obstacle and the right edge is indicated by the large negative spikes. If the objects in a known field are limited to one particular shape then the edge points are all that is needed for image segmentation of the obstacle. When further classification is required the LIDAR data between the edge points can be further analyzed. One analysis technique is to perform a polynomial line fit to the LIDAR points between the previously recognized edge points. The shape and concavity of this line fit can be used to tell if an object is flat, round, or some other shape. For example the round barrels seen in Figure 2 have a concavity that corroborates with its shape. Higher level heuristics could also be applied to adjacent sets of edge points to expand the number of possible classifications, but this was not done in the scope of this project.

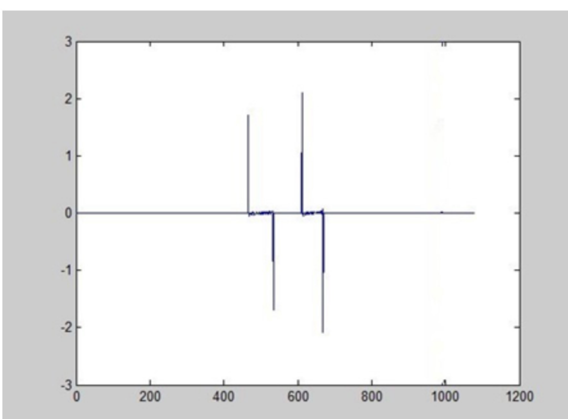

Figure 3. Result of Differential on LIDAR Data

Once an obstacle is recognized to have a round shape the width of the obstacle is calculated using Equation 6. The geometry used to generate this equation can be seen in Figure 4. As seen in this figure, Equation 6 uses d1 and d2 to denote the left and right distance from LIDAR to the obstacle. The parameter α is the angle between d1 and d2 and r is the radius of the obstacle. Finally the result is w, the width of barrel.

$$w = 2 * r = 2 * \left[\frac{d_1 + d_2}{2}\right] * \tan\left(\frac{\alpha}{2}\right) \tag{6}$$

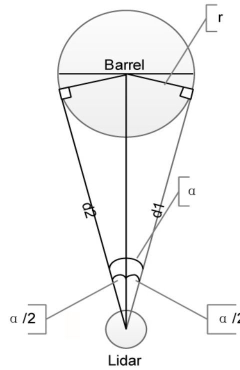

Figure 4. Geometry of Barrel Width Calculation

Image Segmentation Generation

Once the size, shape, and location of obstacles are recognized or assumed using the data of a sensor such as a LIDAR, an accurate and precise segmentation of an image can be done. The LIDAR provides distance and location information about objects with respect to the position of the sensor. If the sensor's position is known and the position, intrinsic parameters, and extrinsic parameters of the camera are known, then the position of an object in the real world can be accurately and precisely placed in a two dimensional image. This principle is the basis of the hybrid sensor image segmentation technique used to determine where barrels are located in an image frame. The result of an original and segmented image can be seen in Figure 5.

Figure 5. Camera Image and Associated Segmented Image

Perspective Correction (Rectified View)

Vehicle mounted cameras tend to be pointed at a downward facing angle as seen in Figure 3. This is done to capture the scene in front of the robot while avoiding potential bright light above the horizon. The perspective of this view is not much use to the vehicle if the pixels of the image are not related to points in the robot's frame of reference. Another problem with the image captured by the camera is the vanishing point of parallel lines as they approach the horizon. It is important for vehicle navigation that parallel lines are actually parallel in an image so that lane lines and roadways can be properly identified [3]. These problems can both be solved with the use of homography to perform a perspective correction. In this case the camera is stationary and the ground is the target image perspective plane. Both of these planes are assumed to be constant while the vehicle is driving so that the correction applied to every image frame is the same. Making this assumption allows the transformation matrix to be used to create a lookup table that maps every pixel in the original image to a perspective corrected image [4], An example of an image with this correction can be seen in Figure 6.

Figure 6. Camera Image and Associated Perspective Corrected Image

The same transformation performed in Figure 6 can be applied to every image to provide a correlation between image pixels and a physical location relative to the robot. In Figure 6 every pixel in the corrected image represents 12mm2 on the ground plane. Correcting the image to have this relationship also gives the image the illusion of an aerial perspective.

In order to perform the perspective correction quickly while the vehicle is moving a lookup table needs to be generated. For this to happen, every pixel value must correspond to a location relative to the robot. This goal is achieved by selecting a 6 by 10 meter region of interest in the camera's view. The intrinsic and extrinsic parameters of the camera that have been generated using the pinhole camera model can then be used to find a new location of the pixel value in the image that has a direct correspondence to a physical location relative to the robot. [4]

Performing this operation consists of extensive computations as every pixel in the image has the transformation applied to it. Once this has been done once however, the relationship between the original image and the perspective corrected image can be saved as the lookup table.

As previously mentioned the image is mapped to appear flat at ground level. This means that three dimensional objects will be mapped as if they were flat. Identification and segmentation of 3D objects with a sensor such as a LIDAR provide a means of recognizing and disregarding an incorrect remap of a specific 3D object to a 2D rectified view.

Experimental Results

ROS Stage was initially used to verify the results of each image segmentation step. The Stage simulator provides camera and LIDAR data in the same way the actual sensors do. It also allows the sensors to provide data from offset positions. The camera calibration and perspective correction can be done on the simulated images. The intrinsic and extrinsic parameters are simply given by the simulation environment. Once everything was verified in simulation the same processes were tested on an actual robot with an LMS 111 SICK LIDAR and an AVT Stingray camera.

Camera Calibration Results

The camera calibration was done using built in tools in ROS and Matlab. The intrinsic and extrinsic parameters that have been generated by these techniques need to be verified to guarantee accurate image segmentation results. The Matlab camera calibration toolbox is capable of providing this verification. Four corner points need to be selected on a grid with known size as seen in Figure 7. The coordinates of these corner points are used by the intrinsic and extrinsic parameters that have been generated to create the grid or red markings seen in Figure 7. The generated grid points are seen to match with the checkerboard corner points which means the camera calibration is able to correctly identify perspective lines in the image and is able to correctly correspond location on a level ground plane with locations in an image frame.

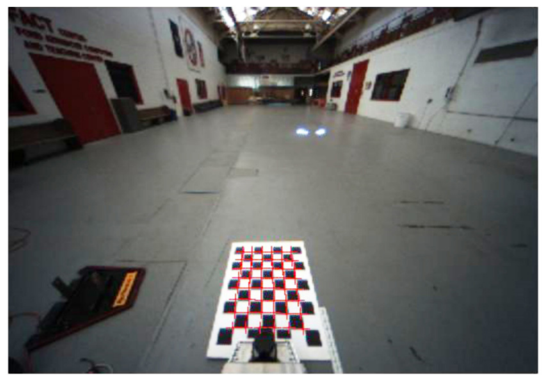

Figure 7. Camera Calibration Using Checkerboard

Barrel Detection Results

Obstacle detection and classification in simulation is identical to detection and classification with a real LIDAR on a real vehicle. This is due to the LIDAR's high reliability and the fact that the frequency in simulation can be matched with the physical sensor. Figure 8 illustrates the areas of sight for a vehicle with a 270° LIDAR. Edge detection using the differential technique provided the edge points of each obstacle seen in Figure 9. The edge points are paired and used to identify the location and general shape of each obstacle. Once the information in Figure 9 is confirmed to match the location of obstacles in Figure 8 the image frame can then be segmented with the classified obstacle.

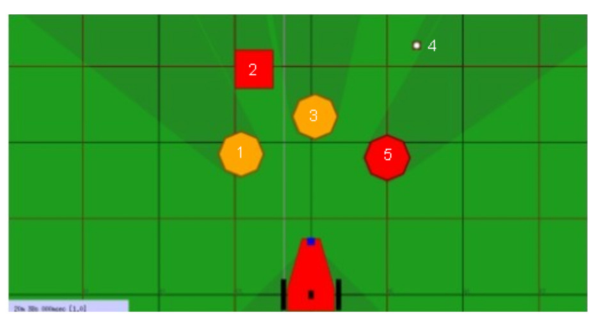

Figure 8. LIDAR barrel detection image

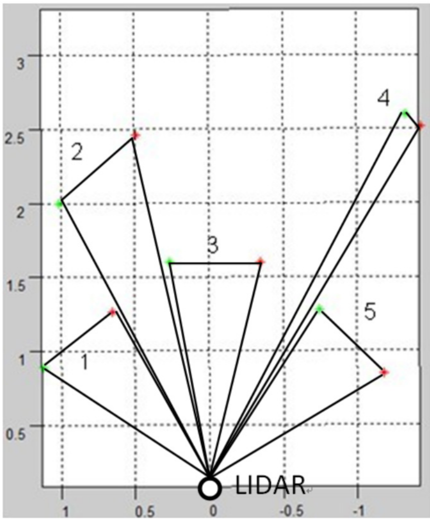

Figure 9. LIDAR edge plot picture

Scene Segmentation Results

Once the obstacles have been identified the intrinsic and extrinsic parameters can be used to translate the three dimensional object from the LIDAR's frame to the image frame. These parameters are given in Stage simulator which means the classification results can be verified in this testing. Results for an image with barrels and the segmentation of this image can be seen in Figure 10.

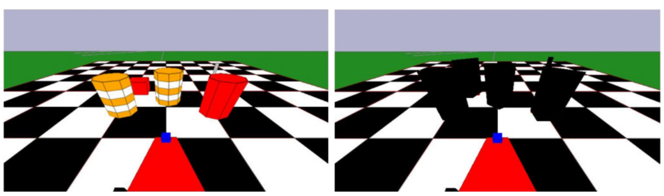

Figure 10. Simulation Scene Segmentation Results

The obstacles found by the LIDAR have clearly been translated to the image correctly as the three dimensional objects have been removed in this case from the image frame. Depending on the application the three dimensional obstacles may want to be removed from an image or everything except the three dimensional objects may want to be removed. Generally if the goal is to identify two dimensional obstacles such as lane lines then the three dimensional objects should be removed to prevent confusion or misclassification. If the three dimensional objects need to be further analyzed then they can easily be removed and further processed as a sub image. Figure 11 shows results of image segmentation of barrels where lane lines need to be detected. The barrels are successfully removed from the image leaving the white lanes line to be further processed. This result also shows the barrel location and shape is correctly identified in the image frame.

Figure 11. Barrel mask real robot result

The Result of Perspective Correction

Perspective correction is another aspect of the camera that can be tested in simulation, but still needs to be verified with an actual vehicle. A lookup table was generated with the given intrinsic and extrinsic parameters of the Stage simulator. The results of the perspective correction in simulation can be seen in Figure 12. The parallel lines of the checkerboard are successfully reassembled and the rectified view has the illusion of an aerial view. Each pixel in the simulation image also has a direct correlation with a location in the simulation environment.

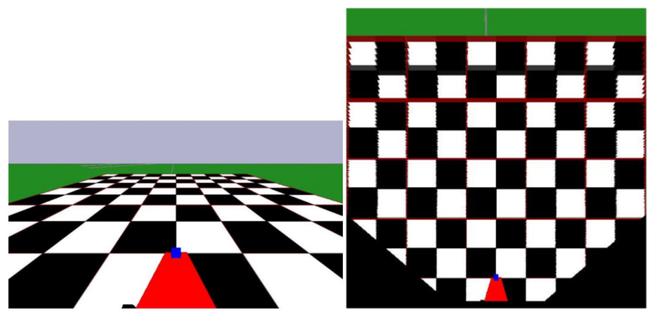

Figure 12. Stage® Perspective view and Corrected Image

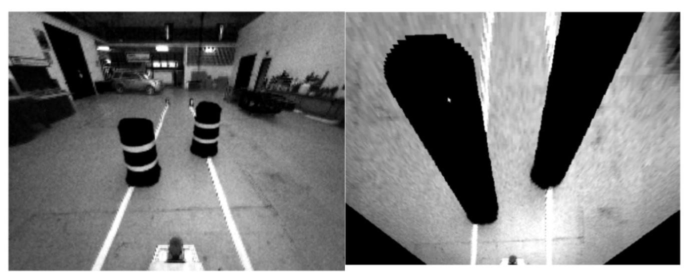

Figure 13. Original image and Perspective Correction

Figure 13 shows results of perspective correction performed on an image from the actual robot. This image also has three dimensional objects in the image frame. As seen in this figure the lines are successfully corrected and made parallel. The barrels however are mapped as though they were flat objects. It is clear that locating three dimensional obstacles in an image is important when perspective correction is performed. Image processing algorithms which look for two dimensional objects will be more able to avoid misclassifications if all of the three dimensional objects are known.

Conclusions

This paper presents a successful framework for camera calibration and the application of sophisticated data fusion between camera and LIDAR sensors. Image segmentation can be dramatically improved with the use of LIDAR sensor data which has many applications in autonomous vehicle navigation. A simple but effective obstacle identification technique is described here to perform scene segmentation with round barrels. A greater variety of obstacles can also be detected and classified with the use of LIDAR and other sensors. Combining this with image processing will allow for computationally efficient segmentation of any complex scene with three dimensional objects.

References

1. Heikkila J. and Silven O., “A four-step camera calibration procedure with implicit image correction,” presented at the Computer Vision and Pattern Recognition, pp. 1106-1112, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997.
2. Hartley Richard and Zisserman Andrew, “Multiple View Geometry in Computer Vision. Cambridge University Press,” pp. 155-157. ISBN 0-521-54051-8, 2003
3. He Qiang, Chu Chee-Hung Henry, “Lane Detection and Tracking through Affine Rectification,” Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2007), Tokyo Japan, May 16-18, 2007.
4. Liebowitz D. and Zisserman A., “Metric rectification for perspective images of planes,” presented at the Computer Vision and Pattern Recognition, pp. 482-488, 1998. Proceedings. 1998 IEEE Computer Society Conference on, 1998

The Engineering Meetings Board has approved this paper for publication. It has successfully completed SAE’s peer review process under the supervision of the session organizer. The process requires a minimum of three (3) reviews by industry experts.

ISSN 0148-7191

http://papers.sae.org/2014-01-0164