Probabilistic Perception System for Object Classification Based on Camera -LiDAR Sensor Fusion

Johan Samir Obando Ceron, Victor Romero Cano, Nicolás Llanos Neuta, Walter Mayor Toro

Probabilistic Perception System for Object Classification Based on Camera - LiDAR Sensor Fusion

Johan S. Obando-Ceron, Víctor Romero Cano, Nicolás Llanos Neuta, Walter Mayor Toro
Automation and electronics department, Universidad Autónoma de Occidente, Colombia

1. Research Problem and Motivation

One of the most basic needs to guide the definition of urban, agro-industrial and territorial management policies is to have a digital topographic representation or map of cities, crops and forests. These maps should ideally be created from multiple sensors whose responses are complementary (color information, for example, complements the returns of a LiDAR sensor in the presence of rain or low reflective objects). Once a topographic representation has been constructed, it can be used to produce and geo-localize higher-level estimates (e.g., location and classification of different trees and plants, crop density, location, and types of pests).

Data can be collected using terrestrial unmanned vehicles equipped with hyper-spectral cameras, stereo cameras and LiDAR (Light Detection and Ranging) sensors. The processing of the acquired data can be used to generate a digital forest model (DFM). DFM will support forest planners in making multi-criteria decisions (MCDA) when planning harvesting operations. However creating a DFM or the map of a city, require a highly accurate and dense point cloud of the environment at hand.

Motivated for building 3D reconstructions from which representations of different vegetation features of an environment can be obtained with high quality and precision. A robust perception system is proposed for densely predicting depth, since it is an essential component in understanding the 3D geometry of a scene. It is known that cameras provide near instantaneous capture of the workspace's appearance such as texture and color, but from a single view, little geometrical information. On the other hand, laser readings may be so sparse that significant information about the surface is missing. The considerations above motivate the formulation of this work's research question: How to develop a perception system for fusing a laser scan with a RGB image in order to produce a higher-resolution range?

2. Technical Contributions

This project apply probabilistic graphical models to the problem of fusing low-resolution depth images with high-resolution camera images to enhance the resolution and accuracy of the depth image. Specifically, a Conditional Random Field (CRF)[1] method is proposed for integrating both data sources. Different from the previous efforts, the perception system developed have the ability of utilizing a unary potential, and different pair wise potentials on super pixels, without relying on any geometric priors nor any extra information, other than range measurement coming from a LiDAR sensor. Furthermore, we proposed re-training the last layers of a convolutional neural network already trained with a new dataset to predict depth on super pixels that does not have depth information associated.

Developing a technique for low-level data fusion between laser and monocular camera to perceive the environment can ensure that the estimated information about the environment is both accurate and dense in spite of the environmental conditions. The perception system developed takes into consideration also the most important data imperfections, such as incomplete data, a complex background with different colors and textures, some obscure objects, different illumination conditions, presence of dust and the blurring effect.

This framework is flexible. It works on pixels or superpixels, and superpixel size of the superpixels determines the resolution of the algorithm's output. We formulated the energy function as a typical combination of unary potentials, pairwise potentials over the nodes and

[1] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proc. Int. Conf. Mach. Learn., 2001.

edges of the image. In our case, the unary term aims to regress the depth value from a single superpixel. The pairwise term encourages neighboring superpixels with similar appearances to take similar depths, and the CNN initializes superpixels without depth information according to their training dataset. Figure 1 shows the model proposed[2].

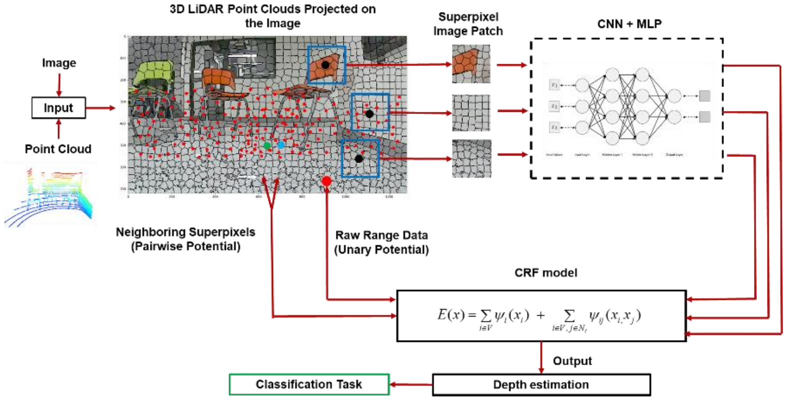

Figure 1. Sensor fusion schematic for depth estimation and classification

We also developed a system based on machine learning that allows a mobile robot to recognize and identify objects found in internal environments. This system was tested using a dense point cloud obtained by our densification system from a Camera image and a LiDAR point cloud. Figures 2 and 3 show the sparse point cloud and the densified one superposed on the color image, respectively.

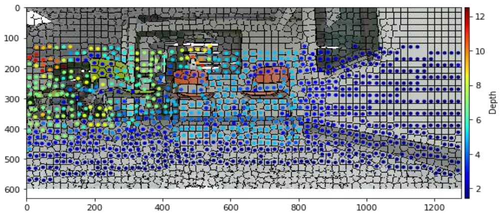

Figure 2. Sparse depth projected on the image

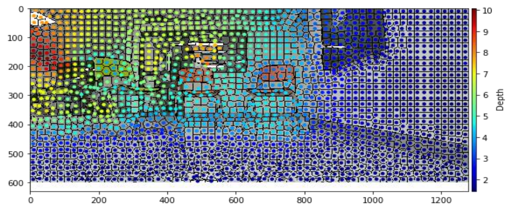

Figure 3.Predition of depth from a single RGB image and a set of sparse depth samples.

Thanks to different methods of point cloud processing and the depth estimation of our algorithm (Figure 5), it was possible to segment and identify different objects in a scene and then implement machine learning techniques like support vector machines and multilayer perceptron neural networks which allowed the object recognition. The final result was outstanding, our method contributed to the performance of the implemented classification methods (Figure 6) in a 12% in contrast to using the raw data given by the LiDAR sensor (Figure 4).

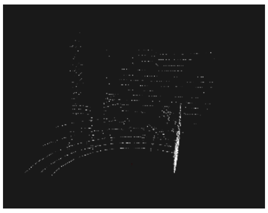

Figure 4. Sparse point cloud

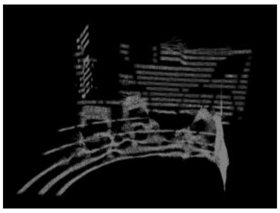

Figure 5. Dense point cloud Output of our perception system.

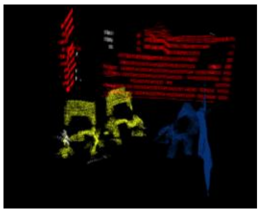

Figure 6. Classification result using SVM with point cloud obtained with our acquisition system

A crucial factor that affects the classification of point clouds is their density. Therefore, we used our perception system to provide a better point cloud to improve the performance and accuracy of our systems. When validating both classification methods implemented; SVM and MLP, we concluded that, regardless of the number of object classes, their accuracy rates were improved when using the densified point cloud generated by our CRF-based method.

---

[2] The code for this project can be found in: https://github.com/JohanSamir/PCL-Perception-Project/tree/master/PointCloud_Object-Classification

3. References

Llanos, Nicolas. Aponte, Sebastian. Velandia, Natali. Rodriguez, Otoniel. Cano, Victor. “Low-cost recognition and classification system based on LIDAR sensors”, presented in the 2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA), Nov. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8588139

K. Lai, L. Bo and D. Fox, ”Unsupervised Feature Learning for 3D Scene Labeling,” presented in the IEEE International Conference on Robotics and Automation (ICRA), May 2014. [Online]. Available: https: //goo.gl/nZk9Y6.

Ron Alterovitz, Sven Koenig, Maxim Likhachev, Robot Planning in the Real World: Research Challenges and Opportunities. [Online]. AI Magazine [Accessed in 11 of November of 2018]. Available in: https://www.aaai.org/ojs/index.php/aimagazine/article/view/2651.

Markus Melin, Aurélie C. Shapiro, and Paul Glover-Kapfer. LIDAR for ecology and conservation. [Online]. WWF-UK. [Accessed in 11 of November of 2018]. Available in: https://www.wwf.org.uk/conservationtechnology/documents/Lidar-WWF-guidelines.pdf.

J. Diebel and S. Thrun. An application of markov random fields to range sensing. In NIPS, 2005. 1, 4, 5, 7.

Trung-Dung Vu, O. Aycard, and F. Tango. Object perception for intelligent vehicle applications: A multi-sensor fusion approach. In Intelligent Vehicles Symposium Proceedings, 2014 IEEE, pages 774–780, June 2014.

J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in Proc. Int. Conf. Mach. Learn., 2001.

T. Qin, T.-Y. Liu, X.-D. Zhang, D.-S. Wang, and H. Li, “Global ranking using continuous conditional random fields,” in Proc. Adv. Neural Inf. Process. Syst., 2008.

F. Liu, C. Shen, and G. Lin, “Deep convolutional neural fields for depth estimation from a single image,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 5162–5170.