UNIVERSAL OBJECT SEGMENTATION IN FUSED RANGE-COLOR DATA

Jeff Finley and Chris Lewis

Electrical and Computer Engineering
Kansas State University
Manhattan, KS 66502
jfinley@ksu.edu, clewis@ksu.edu

KEY WORDS: Expectation Maximization, Data Fusion, SICK LMS, CCD camera, Segmentation

ABSTRACT:

This paper presents a method for universal object segmentation from fused laser range and color image data collected by a mobile robot. By combining both range and color, objects can more easily be distinguished than from either data type alone. The proposed method utilizes the Expectation Maximization, (EM), algorithm to both determine the number of objects in the scene and to segment the data into objects modeled by a universal six-dimensional Gaussian distributions. This is achieved by an iterative split and merge process. Objects segmented from the data by the EM algorithm are then further subdivided, again by EM until the desired resolution is reached. The resulting objects are then recombined if they are found to have similar traits that indicate the data belongs to the same object. The universal object model performs well in environments consisting of both man-made (walls, furniture, pavement) and natural objects (trees, bushes, grass). This makes it ideal for use in both indoor and outdoor environments. Our algorithm does not require the number of objects to be known prior to calculation nor does it require a training set of data. The ultimate goal of segmentation is to facilitate automatic classification of objects and to provide landmark recognition for robotic mapping and navigation.

1 INTRODUCTION

As the robotics evolves, the amount and types of data collected by those robots increases in complexity. The field of data fusion combines different types of sensor data into a unified data set. The challenge is to automatically process, and make sense of, these complex, fused data sets. This paper describes a method to segment colorized range data into universal objects suitable for both indoor and outdoor environments. No training set is required and the number of objects segmented is automatically determined. This method utilizes the Expectation Maximization (EM) algorithm to segment data into six-degree Gaussian mixture models.

Methods have been developed for feature detection from range-only data such as that from laser scanners. Pu and Vosselman scanned the outside facade of buildings and used the data to detect features like doors, windows, and walls (**?**). Another work by Vosselman et al. used voxel space and a 3D Hough transform to extract geometric shapes like cylinders, spheres, and boxes from range data in order to produce computer models of industrial environments and city landscapes. Biber processed range-only information before mapping images onto his models(**?**). Instead of parameterizing geometric shapes, Biswas et al. (**?**) kept isosurfaces of a variety of previously scanned objects. The isosurfaces were created using radial basis functions about 500 pivot points. Subsequent scans were then compared, using spherical decomposition, against a database of previously created isosurfaces to locate similar shapes.

The fusion of range and camera data is attracting a growing body of researchers. A large portion of this research is devoted to finding better fusion techniques. Teams like Abmayr et al.(**?**), Sequeira et al.(**?**), Nyland (**?**), and Dias et al.(**?**) fall into this category. All use high resolution laser scanners along with megapixel cameras. The goal of these authors is to create accurate 3D models of real-world environments for such subjects as historical preservation, city planning, and high-resolution map building. Processing such high resolution data on a mobile robot is computationally impractical.

The paper "A Real-Time Expectation-Maximization Algorithm for Acquiring Multiplanar Maps of Indoor Environments With Mobile Robots" (**?**) explained a method that utilized the Expectation Maximization (EM) algorithm to locate objects in three-space. The authors achieved this by maximizing the likelihood of the range data using planar models. In their implementation of EM, the expectation of each data point was calculated in the E-Step and the maximum likelihood estimate of planar model parameters were calculated in each M-Step. Thrun et al. also developed algorithms to evaluate how "good" a plane fit was. Planes that were considered "weak" were removed while planes that were similar were merged. This removal/merging code was run in between iterations of the EM algorithm. For indoor environments, a planar model performs very well. Man-made objects are commonly planar and most indoor environments consist mainly of walls. For example, in one data set, Thrun et al. report that planes contained almost 95% of all the measurement points. For data that is not accurately defined by a plane, their algorithm leaves the data as a fine-grained polygonal model. Like Biber, Thrun et al. applied color information to their models but did not utilize it in the processing.

A work by Ueda, Nakano, Gharamani, and Hinton presented an Expectation Maximization algorithm along with split and merge criteria to locate two-dimensional Gaussian densities (**?**). The authors remark that their solution helps solve problems when too many Gaussian densities occupy one region while other regions contain relatively few Gaussian densities. Ueda et al. described split values and merge values that are used to represent how accurately points are defined by a density. First, each density receives a split value and a pair of densities receive a merge value. Next, the density with the lowest split value is split while the density pair with the highest merge value is merged. These two calculations are made after every M-Step. A partial E-Step is run once the mixtures have been reorganized. If the splitting and merging has yielded a better likelihood, the new mixtures are kept. If they did not yield a better likelihood, the new mixtures are discarded and EM iterates. It may be noted, that in their work, every time a split occurs, a merge occurs. This kept the number of mixtures constant and implied that the number of mixtures was known a priori.

2 EQUIPMENT

Image and range data was gathered from different locations using the mobile robot in (1). Range data was collected using a SICK LMS 200 laser range scanner. The range scanner returns range points in 1° increments in a 180° field-of-view. Images are captured with a Watec camera using a Computar lens. During data collection, the scanner begins at a tilt of -30° and increments up to 60°, collecting one laser scan and taking one image every 1°.

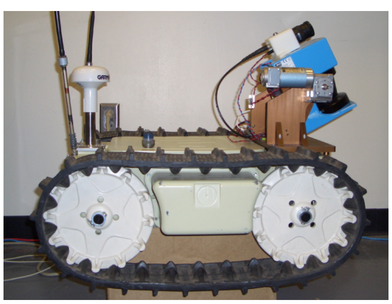
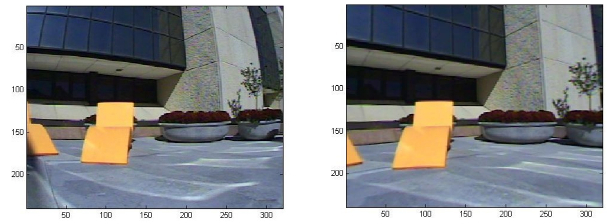


Figure 1: A mobile robot outfitted with a SICK LMS 200 laser scanner and Watec WAT250D color camera mounted on a tilt mechanism. Fused laser and range data is collected and processed for mapping and navigation

2.0.1 Image Calibration

As with most cameras, the images collected are distorted. This distortion originates from both the lens and the CCD array. For an accurate merging of range and color, this distortion must be modeled. Strobl et al. provides an inter-active MATLAB program to calibrate a camera image for distortion called `calib_gui.m` (**?**). This program requires several shots of a well-defined checkerboard to produce a model for the camera. Using this model, images taken with the camera can be rectified. Rectified images are used for the range-color merging.



(a) (b)

Figure 2: A picture taken with a tilt angle of 0° before (a) and after (b) correcting for distortion.

3 DATA FUSION

Once the range information and image are in the same frame, they may be fused into a colorized range image. However, the laser scanner has lower spacial resolution than the camera. If each range point was back-projected onto the image plane to retrieve a color, the higher resolution of the image would be lost. To maintain resolution, each pixel is projected out onto the range data. This could have been accomplished by simply determining which range point's vector is closest to each image pixel's vector using dot products. Because of the higher image resolution, multiple pixels would be associated with a single range point. However, this results in colorized range images with anomalies. To reduce these anomalies the laser range data is segmented into a fine grained object model. With this model, any given range point considers its neighbors to only be those adjacent scan points who have been segmented as belonging to the same object. To create a colorized range image, pixels are projected out from the camera frame and receive a range where they intersect the plane defined by the nearest range point and its neighbors. With this method, pixels that fall on the boundary between two objects do not project to a range where no object resides.

3.1 EM Pre-segmentation

To fuse the image and range data, we must first estimate which range points belong to the same object. The Expectation Maximization algorithm, using three-dimensional (x, y, z) Gaussian densities, has been shown to segment range data into classes where all members of a class are from the same real object. Here, EM is used to pre-segment the range data into a somewhat arbitrary number of classes (64). The number of classes should be greater than the actual number of objects in an environment. Over-segmentation is preferable to under-segmentation because with over-segmentation, classes do not span more than one real object. Problems occur in the under-segmented case when a class spans multiple objects, as the neighboring points used to compute the plane of projection come from different objects. This causes anomalous pixel projections that adversely affect the later segmentation process on the fused data.

3.2 Norm Vector Calculation

Once each range point has been segmented into a class, a polygon for each class can be determined. Planes can be described by the

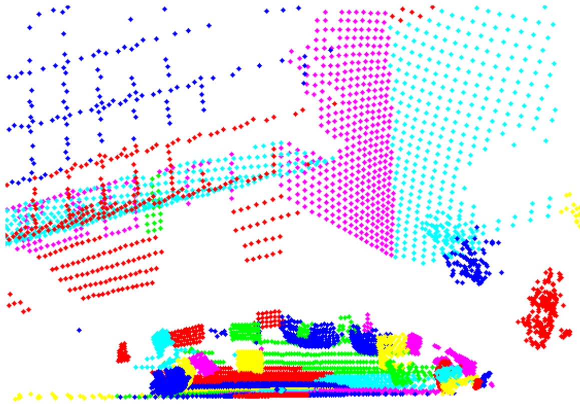


Figure 3: Color-coded range data. Each color represents a class after the EM pre-segmentation step.

equation $0 = ax + by + cz + d$, where (a, b, c) represents the plane's normal vector and d is defined as $-ax - by - cz$. To calculate this normal, we use a least-squares plane fit.

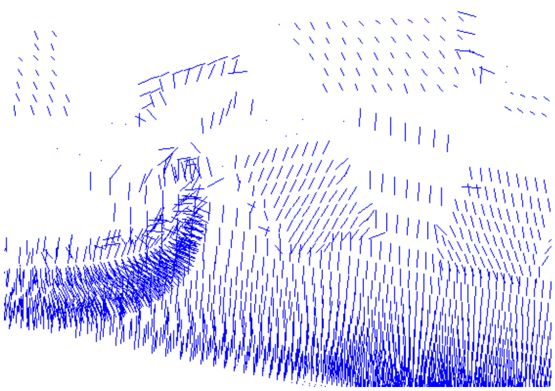


Figure 4: A planar fit considering only adjacent scan points segmented within the same class provides a normal vector for each range point.

3.3 Data Fusion

At this step in the process, the rectified image pixels are ready to be projected out onto the pre-segmented range data. Consider $\overline{p_0} = (x_0, y_0, z_0)$ as the normalized pixel vector of interest. Multiplying $\overline{p_0}$ by some constant K extends the pixel vector out onto the range data. This means that $K \cdot \overline{p_0}$ resides on one of the planes whose normal was computed in the previous section. With this, a range estimate is made for each pixel resulting in a colorized range image which is a fusion of laser and image data.

4 OBJECT SEGMENTATION

With the colorized range image in hand, the next goal is to segment the data into "objects" using information of both color and range. This work assumes no prior knowledge of the environment. Also, the exact number and type of objects present in the

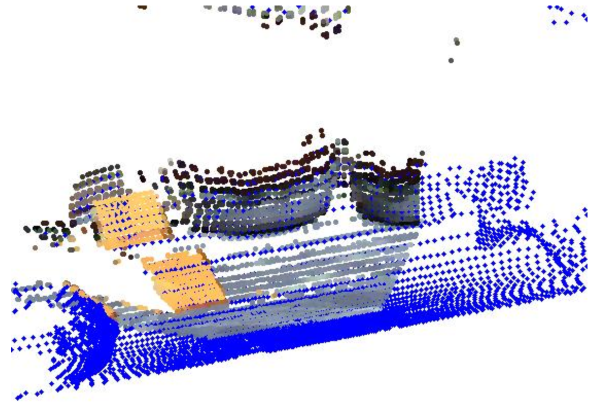


Figure 5: Data from the camera and the laser scanner are fused by projecting each pixel into the scene. The original range data is shown in blue. Note the field of view for the camera is significantly less than the field of view for the laser range scanner.

scene are both unknown. Existing methods for extracting objects from data such as this are often tailored to one specific environment like inside buildings or around city blocks. The goal of this research has been to develop a method suitable for generic objects both indoors and outdoors. The object model must apply to rigid geometric bodies as well as more loosely defined objects. To accomplish this, a six-dimensional Gaussian density is used as a universal object model. The parameters of the object model are exactly the mean and covariances of the distribution for the x,y and z coordinates and the r,g and b color information.

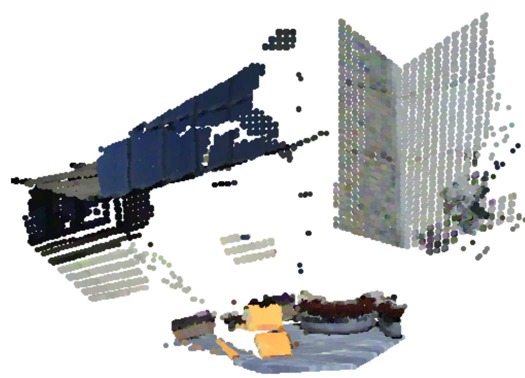


Figure 6: A colorized range scan of an outdoor scene containing both planar objects such as the walls and floor and non-planar objects such as the orange chairs, the tree, and the potted plants.

4.1 Repetitive EM

The proposed algorithm for segmentation starts with a splitting phase that successively processes subsets of the colorized range image data with the EM algorithm, dividing each subset into two classes. Beginning with a single class containing all the data, the first splitting step returns two classes. Each of these two classes is then split again and again until a fine grained segmentation is achieved. Similar to the earlier range only segmentation phase, the goal is to sufficiently subdivide the data into classes where no one class contains data from more than one physical object. Each class is parameterized by its density function.

The resulting fine grained segmentation is crude. Data from a single object often falls into multiple classes. Therefore, the entire data set is reprocessed. This time all at once, with an EM implementation considering all the object models calculated in the splitting phase. Unlike the splitting phase, the reclassification phase does not increase the number of classes, but allows the data from an object to be reclassified into a single model. Objects that were artificially split may be joined. Some classes grow while some classes shrink. The reclassification phase also modifies the distribution models. After this stage, classes that contain too few points are pruned and their data reclassified. As can be observed in the results in Fig. 7, although some physical objects are represented with multiple classes, no class contains multiple objects. The next section will further refine this segmentation by fitting simple geometric models to each class and merging classes with similar models.

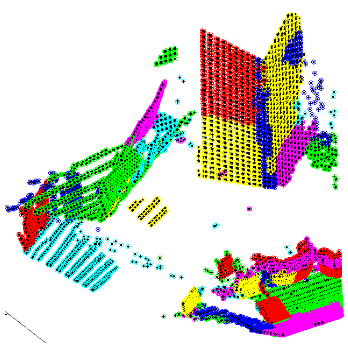


Figure 7: Groups of points are segmented into color-coded classes using several iterations of the Expectation Maximization algorithm. Note that while an object may have several classes, no single class spans multiple objects.

4.2 Planar Extraction

Many scenes contain objects that are man-made. Typically, these objects are characterized well by planar objects. Locating and segmenting planes from a data set would identify a variety of man-made objects. The planar extraction step attempts to locate planes in the fused data using the segmented classes. To achieve this, a least-squares plane fit is performed for each class. If the variance of the residuals is small, that class is deemed a plane. In addition, similar planes are merged, reducing the total number of classes. Figure 8 displays all the planar classes found in the scene. Note that the purple wall is now a single plane.

4.3 Discussion of Results

The algorithm presented in this paper was run on three additional scenes. These scenes were chosen because of the variety of physical objects located in each. Also, the scenes demonstrate the universal nature of this work. Preliminary results, like those in Figure 9, look promising. Highly planar objects, like the wall and floor in Figure 9f, are easily extracted and clustered from the Gaussian densities. The tree trunk in Figure 9e resides in its own class, as do the three bushes in Figure 9e. However, the evergreens in the background of Figure 9e each contain two classes. The same algorithm segmented two partially viewable desks on the sides of Figure 9f. While harder to view in these figures, pixel mis-projections did occur in every scene. However, often these mis-projected pixels segment into their own class which could potentially be culled.

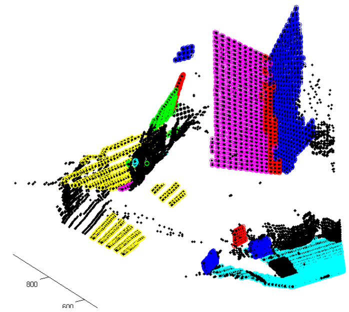


Figure 8: Color-coded planar regions after planar merging. Points in black were determined not to belong to any planar class.

5 CONCLUSION AND FUTURE WORK

This paper presented a method to fuse laser range data and color information gathered from a camera. The resulting data set contained r,g,b and x,y,z information. These data points were then segmented using an Expectation Maximization algorithm and a universal, six-degree Gaussian density object model. The resulting segmentation produces favorable results on a variety of scenes containing a variety of objects both man-made and natural. Furthermore, these different objects were all segmented using a single, universal model. It was only after segmentation that planer objects were more specifically defined from the generic object model. Subsequent efforts could search for other simple geometric models such as cylindrical or spherical objects. Alternatively, these new models could be included in a likelihood function, creating a heterogeneous mixture model. Thrun also used a heterogeneous mixture model by including a “noise” mixture along with planar mixtures (**?**). Objects that are not accurately described by any simple geometric model may be left modeled by the generic Gaussian densities.

REFERENCES

Abmayr, T., Hrtl, F., Mettenleiter, M., Heinz, I., Heinz, A., Neumann, B. and Frhlich, C., n.d. Realistic 3d reconstruction - combining laserscan data with rgb color information.

Biber, P., Andreasson, H., Duckett, T. and Schilling, A., 2004. 3d modeling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In: IEEE/RSJ International Conference on Intelligent Robots and Systems.

Biswas, S., Aggarwal, G. and Chellappa, R., 2006. Invariant geometric representation of 3d point clouds for registration and matching. In: 2006 IEEE International Coference on Image Processing.

Dias, P., Sequeira, V., Vaz, F. and ao G.M. Gonçalves, J., 2003. Registration and fusion of intensity and range data for 3d modelling of real world scenes. In: Proceedings of the 4th IEEE International Conference on Recent Advances in Digital Imaginand and Modeling.

Nyland, L. S., 1998. Capturing dense environmental range information with a panning, scanning laser rangefinder.

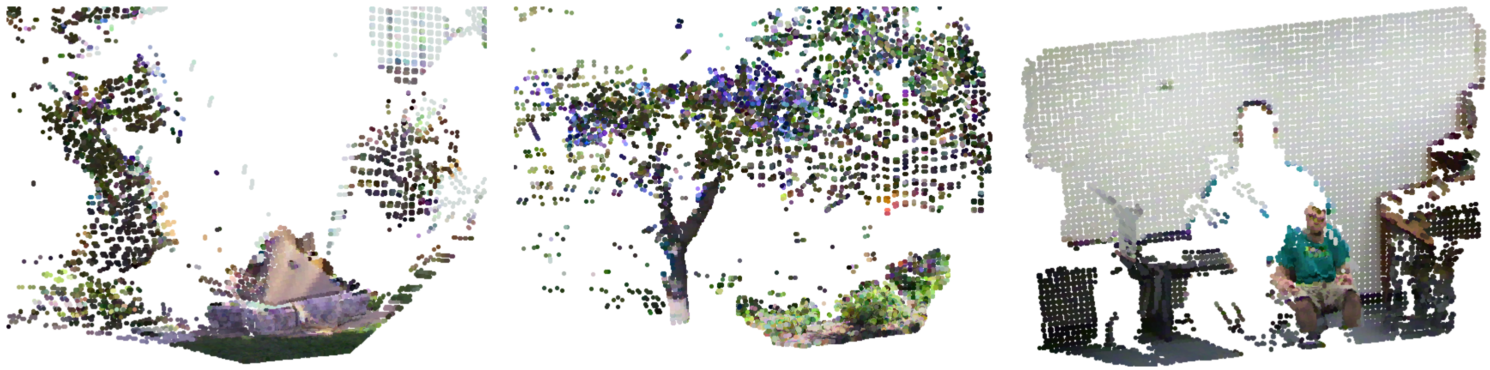
Pu, S. and Vosselman, G., n.d. Automatic extraction of building features from terrestrial laser scanning.

Sequeira, V. and Gonçalves, J. G., 2002. 3d reality modelling: Photo-realistic 3d models of real world scenes. In: Proceedings of the IEEE First Internation Symposium on 3D Data Processing Visualization and Transmission.

Strobl, K., Sepp, W., Fuchs, S., Paredes, C. and Arbter, K., n.d.

Thrun, S., Martin, C., Liu, Y., Hhnel, D., Emery-Motemerlo, R., Chakrabarti, D. and Burgard, W., 2004. A real-time expectation-maximization algorithm for acquiring multiplanar maps of indoor environments with mobile robots. IEEE Transactions on Robotics and Automation.

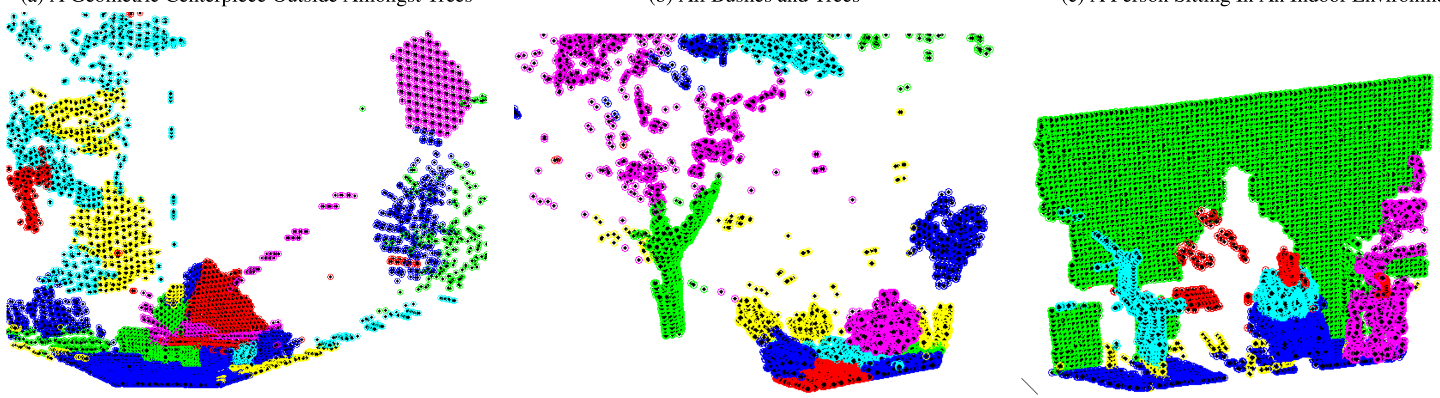
Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G. E., 1998. Split and merge em algorithm for improving gaussian mixture density estimates. In: Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop.



(a) A Geometric Centerpiece Outside Amongst Trees

(b) All Bushes and Trees

(c) A Person Sitting In An Indoor Environment



(d) A Geometric Centerpiece Outside Amongst Trees

(e) All Bushes and Trees

(f) A Person Sitting In An Indoor Environment

Figure 9: Range-color data for three distinct scene and their resulting color-coded final segmentation.